Studying the Dynamic Properties of Intrinsically Disordered Proteins in Complex Environments by Nuclear Magnetic Resonance and Molecular Modeling

Vincent Schnapka

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : PHYS - Physique
Spécialité : Physique pour les Sciences du Vivant
Unité de recherche : Institut de Biologie Structurale

Étude des Propriétés Dynamiques des Protéines Intrinsèquement Désordonnées en Milieu Complexe par Résonance Magnétique Nucléaire et Modélisation Moléculaire

Studying the Dynamic Properties of Intrinsically Disordered Proteins in Complex Environments by Nuclear Magnetic Resonance and Molecular Modeling

Présentée par :

Vincent SCHNAPKA

Direction de thèse :

Martin BLACKLEDGE DIRECTEUR DE RECHERCHE, CEA CENTRE DE GRENOBLE	Directeur de thèse

Rapporteurs :

MIKAEL AKKE
PROFESSEUR, LUNDS UNIVERSITET
STEPHAN GRZESIEK
PROFESSEUR, UNIVERSITÄT BASEL

Thèse soutenue publiquement le **20 décembre 2023**, devant le jury composé de :

MIKAEL AKKE PROFESSEUR, LUNDS UNIVERSITET	Rapporteur
STEPHAN GRZESIEK PROFESSEUR, UNIVERSITÄT BASEL	Rapporteur
MASSIMILIANO BONOMI CHARGE DE RECHERCHE HDR, CNRS ILE-DE-FRANCE MEUDON	Examinateur
GUILLAUME BOUVIGNIES CHARGE DE RECHERCHE, CNRS DELEGATION PARIS CENTRE	Examinateur
MARC JAMIN PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Président

Acknowledgement

Jury members

First of all, I would like to thank the jury members for their time, expertise and thoughtful feedback during the reviewing process and the defense of this thesis.

Fabien and Loic

I also want to thank Loic and Fabien for being part of my Thesis advisory committee. I greatly appreciated the relevant and supportive feedback provided during our meetings. Additionally, I want to thank Fabien more specifically for guiding me through my choice of Ph.D. thesis during my master's studies.

Martin

Of course, I am grateful to my thesis director, Martin Blackledge, whose guidance, encouragement and precious expertise have been decisive in the success of this thesis. I am particularly grateful for the amount of intellectual freedom that I could enjoy during my stay. I could satisfy my curiosity with audacious experiments, and with that I should pay tribute to the room temperature probe of the old 600(599) MHz Oxford-Bruker spectrometer that served me so well during these years. I will always enjoy our various critical discussions about the last article that came out, or sometimes on much older papers often overlooked by modern scientists. Although I cannot do rock-climbing, I share my love for mountains and I will always enjoy a hike in Chartreuse.

Nicola

I cannot continue this acknowledgment section without mentioning Nicola. Although I had him during only one year and a half, he had a significant impact on me and my scientific skills. He teached me everything, from running an NMR experiment to running and analyzing MD simulations. And I will not forget to mention the very valuable skill of installing NMRpipe. You also were an example to me as a scientist.

The group

Besides my supervisors, I also want to express my gratitude to all the members of the FDP group, who all had an impact on the success of my thesis.

First, I wanna thank Malene for her help and valuable insight in NMR exchange techniques in addition to being an model for me. In addition to the science, she is an incredible human being, and I wanna thank her for supporting me during the rare difficult moments I could face.

Thank you to Damien, who teached me the proper way to produce proteins and who was always available when something went awry in the wet lab.

I would also like to express my gratitude to Anas, the colleague turned friend with whom I likely spent the most time. Our intense cultural and scientific discussions were always enjoyable, and I believe it is reasonable to say that our fruitful exchanges played a role in enhancing my scientific creativity during this adventure. I must also mention that it is thanks to him and our gym sessions that I could maintain reasonable physical health during this Ph.D. journey.

I extend my thanks to Sam, who, despite leaving Grenoble to join Sigrid in Berlin after a year and a half, had a significant impact on my well-being in this laboratory. I will always cherish our lunchtimes and our evenings filled with music, games, and creativity.

I must also thank Thibault, who was as valuable as a supervisor to me in the wet lab regarding protein purification. His valuable insight greatly improved my life in the laboratory, not to mention his excellent taste in music, which I enjoyed while purifying my proteins. We also enjoyed numerous concerts together and I am looking forward to attending the next one.

Another important lab mate was Maiia, who, with her good mood always made the office an enjoyable place to be. I must also mention her impeccable taste in sweets and beverages, which undoubtedly helped me in some way during the preparation of this thesis.

Among all the members of the group, I also want to express my gratitude to Lenette, whose calm demeanor perfectly complemented the office environment. I regret not being able to join her in more hiking sessions.

I also want to thank Sima, who was also helpful in the start of my thesis and who teached me a lot in the wet lab for the phase-separation-oriented experiments.

Finally, I want to thank all the other members who, in one way or another, made my life more enjoyable in the laboratory including Emmi, Aldo, Lefan, Maud, Selin, Luiza, Laura, Elise, Robert, Torsten, Shivangi, Jacqueline etc.

The NMR platform

Of course I am grateful to all the members of the NMR platform, including Alicia, Adrien and Bernard, who were very helpful when something was wrong with a spectrometer.

Other friends and family

Finally, I want to express my gratitude to all my external friends and family members who supported me during these years. Your constant encouragement and understanding have been in-

valuable throughout this journey.

Résumé en français

Les Protéines Intrinsèquement Désordonnées représentent un tiers du protéome humain et sont impliquées dans un grand nombre de mécanismes biologiques comme la signalisation cellulaire et la formation de compartiments sans membrane vitaux pour la cellule en plus d'être impliquées dans de nombreuses maladies et infections virales. Étant donné qu'elles n'ont pas de structure stable, leurs propriétés dynamiques jouent un rôle important dans leur fonction biologique. Malgré le fait que la caractérisation in vitro de ce type de protéines est relativement bien établie, leur comportement dans des conditions physiologiques comme l'environnement cellulaire incluant les compartiments sans membrane générés par séparation de phase liquide-liquide, hautement concentrés en protéines et macromolécules, est encore mal compris. Dans cette thèse, la partie C-terminale de la Nucléoprotéine du virus de la rougeole est utilisée en tant que modèle pour étudier les propriétés dynamiques de cette prototypique protéine désordonnée en phase condensée générée par séparation de phase liquide-liquide ainsi qu'en milieu encombré par Résonance Magnétique Nucléaire et simulation de Dynamique Moléculaire.

Dans un premier temps, il est montré que la séparation de phase liquide-liquide augmente les temps de corrélation rotationnels associés à la dynamique du squelette peptidique de la protéine et que la contribution de chaque mode dynamique est significativement redistribuée. Les simulations de protéines concentrées ont montré que cette redistribution observée en phase condensée est corrélée avec l'augmentation de la proximité non spécifique des chaines protéiques les unes avec les autres, conduisant à des mouvements plus restreints dans le squelette peptidique. La relaxation RMN de spin a aussi montré que les ralentissements de la dynamique au niveau de certains segments peptidiques sont corrélés avec un taux élevé de résidus chargés et aromatiques, ce qui suggère une forte importance de ces résidus et de leurs interactions dans la stabilisation de cette phase condensée en accords avec d'autres études.

Une construction plus courte de la partie C-terminale de la Nucléoprotéine du virus de la rougeole a été utilisée pour étudier les effets de l'encombrement avec du PEG10000, un polymère à longe chaine, à des concentrations qui approchent celles de certains milieux cellulaires et des phases condensées. La relaxation de spin a montré que ce très haut niveau d'encombrement avait pour effet de redistribuer les modes dynamiques de façon semblable a ce qui a été observé en phase condensée, suggérant que des milieux suffisamment encombrés pourraient modifier les propriétés dynamiques de certaines protéines intrinsèquement désordonnées. Enfin, une interaction protéine-protéine entre cette construction et son partenaire, la partie C-terminale de la phosphoprotéine du virus de la rougeole, a été étudiée en milieu encombrée. Un ralentissement important des propriétés cinétiques de cette interaction est observée et une étude plus poussée de cette interaction devrait contribuer à une meilleure compréhension à l'échelle atomique des effets de l'encombrement sur les interactions entre protéines, cruciaux pour les processus biologiques.

Cette étude apporte un aperçu de l'effet de l'encombrement sur la dynamique des protéines in-

trinsèquement désordonnées et est une étape supplémentaire vers une meilleure compréhension du comportement de cette classe importante de protéines dans les environnements physiologiques incluant les compartiments sans membranes, omniprésents chez les cellules eukariotes et les machineries virales.

Abstract

Intrinsically Disordered Proteins represent a third of the human proteome and are involved in numerous biological mechanisms from cell signalling to the formation of vital membraneless organelles in the cells as well as playing a role in numerous diseases and viruses. Since they lack a stable 3D structure, their dynamic properties play an important role in their biological functions. While characterizing this class of proteins in vitro is well established, their behavior in more physiological conditions from the crowded cellular milieu to the highly concentrated liquid-liquid phase separated membraneless organelles is poorly understood. In this thesis, we use the C-terminal domain of the Nucleoprotein of Measles Virus as a model system to investigate the dynamic properties of this prototypical disordered protein in its liquid-liquid phase separated condensed phase as well as in crowded environments using Nuclear Magnetic Resonance and Molecular Dynamics Simulations.

We first show that liquid-liquid phase separation increases the rotational correlation times associated with the backbone dynamics of the protein and that the contribution of local and long-range motional modes is significantly redistributed. MD simulations of concentrated proteins suggest that this redistribution is correlated with the increase of non-specific intermolecular proximity or entanglement within the concentrated phase, leading to more restricted motions within the protein's backbone. NMR spin relaxation also showed that the distribution of slower segmental motions is correlated with the position of charged and aromatic residues in some regions, suggesting that the phase separation of our protein is stabilized by electrostatic and cation-pi interactions, in agreement with previous proposals.

A shorter construct of the Measles C-terminal domain of the Nucleoprotein was further studied upon crowding with a high molecular weight PEG10000 polymer up to concentrations that correspond to those found in biomolecular condensates. NMR spin relaxation measurements showed that high levels of macromolecular proximity in super-crowded conditions redistribute the dynamic modes of the protein backbone in a way that is not observed at lower levels of crowding, suggesting that significantly crowded conditions can modify the dynamic properties of certain IDPs. Finally, a protein-protein interaction between this construct and its partner in the C-terminal domain of the Measles Virus Phosphoprotein was studied upon crowding and exhibited a significant slow down of the kinetic properties of the interaction. Further studies of this interactions by NMR should contribute to a better understanding of the atomic-resolution effects of crowding on protein-protein interactions, crucial for biological processes.

This study provides insight into the effect of crowded environments on the backbone dynamics of Intrinsically Disordered Proteins and is a step towards a better understanding of how this crucial class of proteins behave in more physiological environments including the currently actively studied membrane-less organelles, ubiquitous in eukariotic cells and viral machineries.

Contents

Introduction

"If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms." - Richard Feynman

As Richard Feynman liked to point out, all biological processes can be explained in the nanoscale by the jiggling and wiggling of numerous macromolecules. All conventional living systems work as follows: DNA encodes the biological information for a given species. This DNA is then transcribed into RNA, which is going to carry this information towards a big molecular machine called a ribosome, to translate this information into what we call proteins. In fact, the process of transduction and translation already involves numerous protein machineries such as the ribosomes (which also contains RNA) and the DNA polymerase, which may make us wonder what came first between the chicken and the egg. Proteins are the most important macromolecules in living organisms. They are responsible for nearly all the biological processes that maintain life. Among famous types of proteins we can cite enzymes that trigger and facilitate biochemical reactions, antibodies that are involved in the immune system, transport proteins, receptor proteins, signaling proteins etc.

A protein is defined as a hetero-polymer composed of different amino-acids, forming a long polymer chain. The atoms that are common to all amino-acids form the backbone of the protein in which the relatively rigid covalent bonds between amino-acids are responsible for a planar topology between the backbone atoms defining so-called peptide-planes. The amino-acid-specific atoms on the other hand form the lateral side chains and determine all the sequence-specific features of the protein. In living organisms, proteins are made of 20 different amino-acids that all have different properties. This high degree of diversity allows the obtention of very different species depending on the amino-acid composition. In fact for a protein with N residues, there is 20^N different possible combinations of amino-acids resulting in proteins with different structural and physical properties.

It took mankind several centuries to understand what drives biological processes from a single cell to a whole multi-cellular species such as *homo sapiens*, although most of the major progress have been done within the second half of the 20^{th} century. The first step towards the "protein revolution" was provided in 1789 by a French scientist, Antoine-François Fourcroy, with the description of "albumins", a group of proteins including albumin, fibrin and others species [1]. This study was followed by a more complete description of proteins by Mulder in 1838 [2] and the proposition of the name "*protéine*" by Berzelius in a letter written in French to Mulder in the same year [3]. At that time, very little was known about proteins, and most studies were focused on the characterization of the atomic composition and molecular weights of these molecules [4]. Until the rise of X-ray crystallography, the description of proteins remained relatively simple. In the early 20^{th} century, proteins were found to be polypeptides [5], although it was only demonstrated in 1949 with the first protein sequencing [6].

Early studies already proposed that proteins have a defined 3D structure [7, 8]. In 1958, Kendrew and coworkers published the first 3D structure of a protein, the Myoglobin, by X-ray crystallography [9]. In the following years, more and more protein structures were determined by X-ray crystallography, Nuclear Magnetic Resonance (NMR) and later on cryo-Electron Microscopy (cryo-EM). This era of structural biology saw the domination of the structure-function paradigm, where protein function was believed to be essentially determined by the protein structure.

We can describe four levels of structure in a protein:

1. The primary structure that simply corresponds to the amino-acid sequence

2. The secondary structure that corresponds to local arrangements of amino-acid residues in the chain. The main secondary structures one can encounter are α-helices, described by a

helical arrangement of the poly-peptide chain stabilized by hydrogen bonds, and β-sheets, described by strands that are laterally connected by several hydrogen bonds in the backbone.

3. The tertiary structure that corresponds to the overall 3D arrangement of the protein chain

4. The quaternary structure that is the assembly of several protein chains to form a larger structure.

The highly dynamic proteins or protein regions that didn't form any structure were ignored for decades in structural biology. The importance of these proteins, long believed to be irrelevant for biology, was reassessed only in the beginning of the 21^{th} century when it was found that they represent a third of the proteome in eukaryotic cells [10] and that they are in fact involved in biological processes as well as diseases [11–13]. These so-called Intrinsically Disordered Proteins (IDPs) are described by a rather flat energy landscape, which allow them to sample a significant amount of the conformational space, unlike folded proteins that are described with one or few structures in native conditions. In fact, the amount of structure in a protein should be viewed as a spectrum between order and disorder, rather than a discrete categorization, since some proteins are only partially disordered, or exchange between different conformations with different levels of structure. Intrinsic disorder was found to be involved in numerous biological mechanisms including signal transduction, transcription and regulation [14, 15], the circadian cycle [16], cellular protection against desiccation [17] as well as numerous diseases and infections including neurodegenerative diseases [18–20], signaling disorders [21] and viral machineries [22–25].

Because of the lack of 3D structure, the structure-function paradigm alone is not sufficient to explain all the biological functions of Intrinsically Disordered Proteins. A tremendous amount of effort has been made and is still being made to describe conformational ensembles of IDPs, with the hope to find a clue towards function, interaction patterns and drugable sites. It was found that two key characteristics of IDPs define their function. First, their amino-acid composition define their secondary structure propensity and the type of interaction they can make. Second, their dynamic properties define the rate of conformational rearrangement at multiple timescales. A relatively high amount of dynamics in IDPs can play an important role in numerous interactions. Therefore, characterizing the dynamic properties of IDPs is crucial to understand their biological functions.

Nuclear Magnetic Resonance is the most suited tool to study IDPs at atomic resolution. Unlike crystallography and cryo-EM that look at fixed structures, NMR is powerful at obtaining ensemble averaged information in highly dynamic systems at atomic resolution. In fact, the more dynamic the system is, the more suited the system is for NMR spectroscopy as we will see in chapter 1, which make IDPs ideal systems. NMR provides both structural and dynamical information at multiple timescales as illustrated in figure 2. For example, very fast chain dynamics can be probed by NMR spin relaxation while slower processes including conformational changes and protein interactions can be probed with chemical exchange techniques.

Complementary to NMR, computational techniques including classical Molecular Dynamics simulation allow a direct visualization of biological processes, and combination with experimental data allow both validation of the simulations and further analysis of complex mechanisms that are difficult to access by experiments. Just like NMR, MD simulation and computational methods in general can allow the study of protein dynamics at multiple timescales from the study of

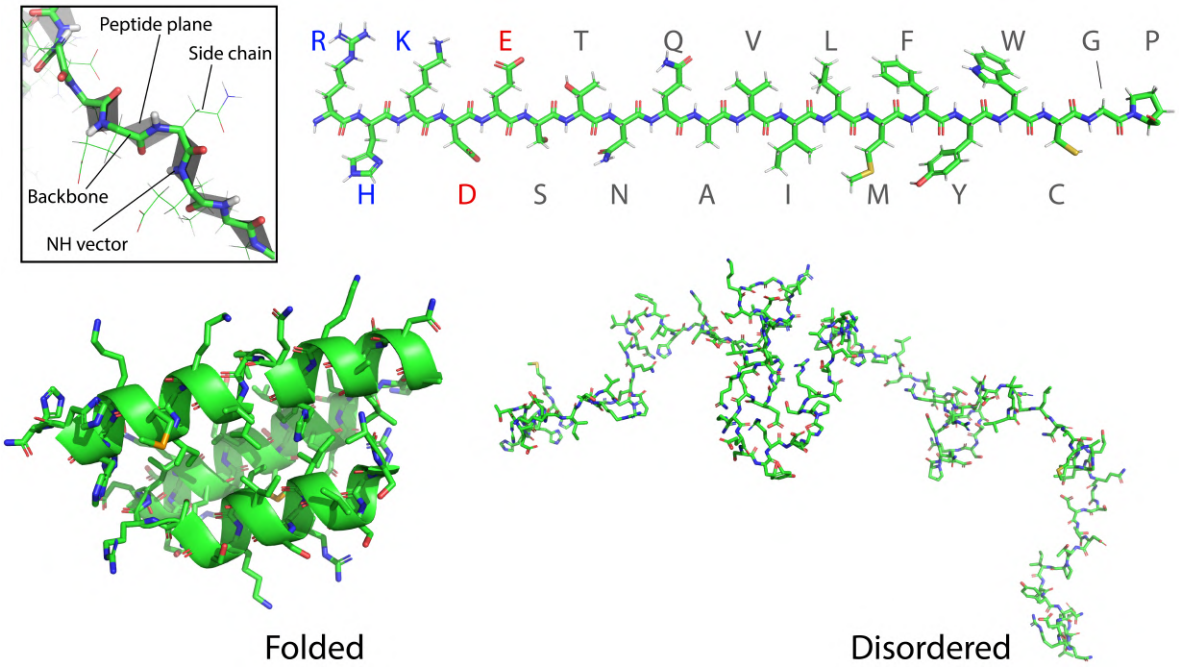


Figure 1: Overview of the structure of a protein. Panel A: Illustration of a peptide chain with a highlighted backbone. The carbon, nitrogen, oxygen, hydrogen and sulfur atoms are colored in green, blue, red, white and gold respectively. Panel B: The 20 amino-acids with their one letter code. In blue and red are the positively charged and negatively charged residues respectively. Panel C: Example of a folded protein (left: Measles Phosphoprotein XD domain) and an Intrinsically Disordered Protein (right, Measles Nucleoprotein C-terminal domain: Ntail).

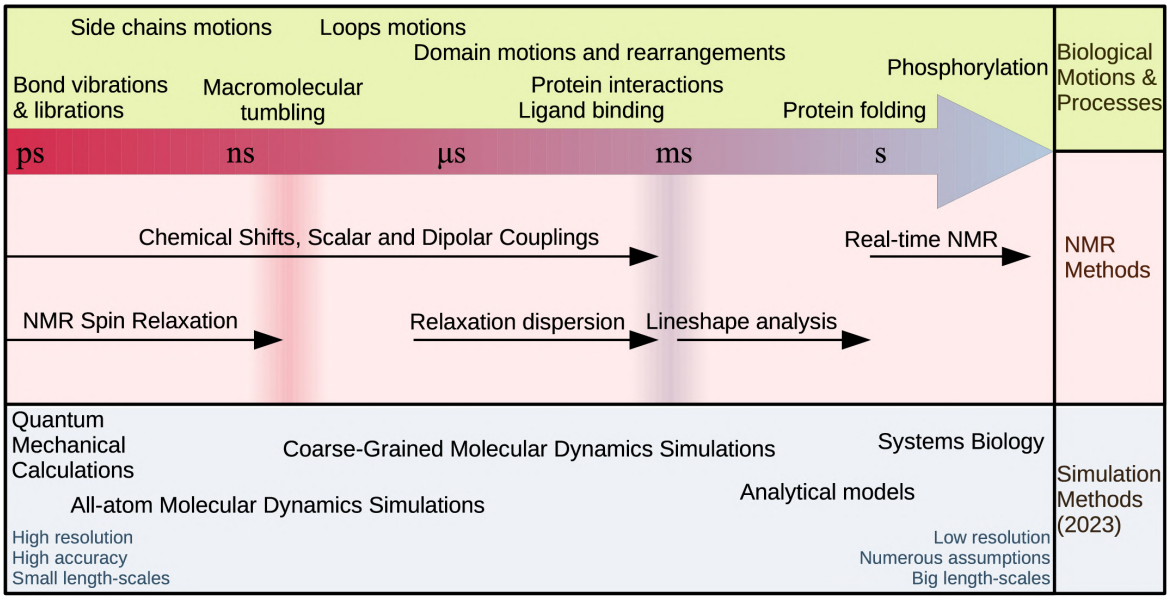


Figure 2: Timescale axis illustrated some example of Biological mechanisms associated with protein dynamics (Top, green), NMR techniques (Center, red) and the main MD simulation methodologies (Bottom, blue) to study these processes

chemical reactions and bond vibrations by Quantum Mechanical calculations to the study of large mesoscopic systems with Coarse Grained simulations and analytical methods.

In the last decades, the scientific community provided numerous methods to study IDPs in the test tube, including NMR and computational methods but also fluorescence methods, SAXS and other techniques that provide complementary information. A lot of effort and progress has been made in understanding their function in biology. Nevertheless, the question of whether what is studied in the test tube is perfectly transferable in the real physiological environment remains only partially answered. Additionally, recent progress in cell biology unveiled the importance of membrane-less organelles in eukariotic cells and viral machineries, which adds a new complexity layer to the already complex cellular environment. These liquid-liquid phase separated compartments, found to be composed of mostly IDPs and sometimes RNAs, are defined by a highly dense environments with a composition that differs from the rest of the cell.

This thesis aims to study the dynamic properties of IDPs in so-called complex environments, that include liquid-liquid phase separated system as well as crowded systems that aim at approaching the conditions in the cell. It will be organized as follow:

1. The first chapter focuses on the theoretical aspects of Nuclear Magnetic Resonance, the main technique of this thesis. We will start by describing NMR before explaining the phenomenon of NMR spin relaxation and how it can be used to characterize the dynamic properties of a protein's backbone. Then, slower processes will be discussed by introducing the theory of chemical exchange in NMR and the main techniques that take advantage of this process.

2. The second chapter focuses on the current experimental and computational challenges regarding the study of IDPs in complex environments. We first briefly introduce the experimental observables allowing to study IDPs as well as the computational techniques that allow their characterization, in the light of the current state of the art. Then, we discuss the recent advances in studying IDPs in crowded and cellular environments to finish with IDPs in liquid-liquid phase separated systems.

3. The third chapter addresses the study of the dynamic properties of MeV Ntail upon liquid-liquid phase separation by combined NMR spin relaxation and MD simulation. We show how LLPS can modulate the different relaxation-active dynamic modes of IDPs.

4. The fourth chapter further addresses the sequence-dependence of the dynamic behavior of MeV Ntail in the condensed phase. An attempt to explain the effect of inter-molecular interactions on the sequence-dependent dynamic modulation is made using NMR spin relaxation. Inter-molecular contacts and exchange processes are also further investigated by NMR.

5. The fifth chapter deals with the effect of the concentration of polymer crowders on the dynamic properties of MeV Ntail 465-525, a short construct that doesn't phase separate under the studied conditions. We show that a concentrated regime of polymer crowding changed the dynamic properties of MeV Ntail 465-525.

6. The sixth chapter is interested in the study of protein-protein interactions upon crowding. MeV Ntail 465-525 is studied with its partner MeV PXD as a model system to study the effect of polymer crowding on this important interaction involved in the Measles Virus replication machinery.

7. The seventh chapter aims to tackle several experimental challenges in studying complex environments by NMR. A first part will discuss the measurement of cross-correlated relaxation rates in biomolecular condensates. In a second part, a method for measuring NMR spin relaxation rates in highly concentrated mixtures is proposed to tackle the problem of natural abundance contamination.

Chapter 1

Protein Dynamics by NMR Spectroscopy

"I have not yet lost a feeling of wonder, and of delight, that this delicate motion should reside in all the ordinary things around us, revealing itself only to him who looks for it. I remember, in the winter of our first experiments, just seven years ago, looking on snow with new eyes. There the snow lay around my doorstep—great heaps of protons quietly precessing in the earth's magnetic field. To see the world for a moment as something rich and strange is the private reward of many a discovery" - Edward Mills Purcell

Demonstrated in 1938 by Rabi *et al.*[26] and simultaneously discovered in bulk materials in 1945 by Purcell and Bloch [27, 28], the phenomenon of nuclear magnetic resonance has now numerous applications in chemistry, biology and medicine including chemical compound characterization, protein and RNA structure determination, magnetic resonance imaging, ligand screening for drug discovery and biomolecular dynamics and interactions. The aim of this chapter is to introduce the important liquid-state Biomolecular NMR concepts and techniques used in this thesis. We will first briefly see how NMR works by exploring the magnetic properties of atomic nuclei. Then, we will see how NMR spin relaxation phenomena arise and how it can be used to obtain insightful information on fast molecular dynamics, from simple rigid molecules to our highly dynamic IDPs. Finally, we will see that it is possible to exploit so-called chemical-exchange phenomena to probe slower motion processes that report on molecular interactions, large molecular domain motions and conformational rearrangements, which are of high importance to understand biology at the nanoscale. Since NMR Spin relaxation is a fundamental concept in this thesis, it will be treated with a relatively important amount of details, but the reader can still find more exhaustive information in many excellent NMR textbooks [29–32].

1.1 Principles of NMR Spectroscopy

1.1.1 Nuclear Magnetism

NMR spectroscopy deals with the resonance of the magnetization of an ensemble of nuclear spins under a magnetic field. This magnetization arises from the properties of nuclear spin angular momentum, characterized by the nuclear spin quantum number *I*. This quantum number is defined by the relative number of protons and neutrons in a given nucleus. Nuclei with a non-zero spin quantum number possess a nuclear magnetic moment and have a number of associated energy levels equal to $2I + 1$. We can associate these energy levels with the magnetic quantum number $m \in \{-I, -I + 1..., I\}$. All these energy levels are degenerate except in the presence of a magnetic field, where they are separated by a value proportional to the intensity of the magnetic field. This phenomenon, called Zeeman splitting, is due to the slight preference for a spin magnetization to point towards a given orientation depending on the magnetic field. By convention, we point the applied static magnetic field along the z-axis of a fixed Cartesian frame that we call the laboratory frame.

We may express the magnitude of the nuclear magnetic moment as $\mu = \gamma I$ with γ being the gyromagnetic ratio of the given nucleus. This constant is different for each nucleus and determines the sensitivity of a nucleus to yield separated energy levels and, as a result, to be sensitive to NMR. By definition, the equilibrium populations of the different energy states in an ensemble of spins are given by the Boltzmann distribution:

$$p_m = \frac{N_m}{N} = \frac{exp\left(\frac{-E_m}{k_B T}\right)}{\sum_{i=-I}^{I} exp\left(\frac{-E_i}{k_B T}\right)}$$

With p_m and N_m being the proportion and number of nuclei in the *m*th state respectively, *N* the total number of nuclei, k_B the Boltzmann constant and T the temperature. The sum is over all energy levels from $-I$ to I, and E_i is the energy associated to the *i*th state under a given magnetic field B_0 with:

$$E_i = -i\hbar\gamma B_0$$

At the temperatures we are used to work with, the magnitude of the degeneracy is very small and $E_m/k_B T$ is much smaller than 1. Therefore, we can obtain an approximate expression for p_m using a first order Taylor expansion:

$$p_m \simeq \frac{1}{2I+1}\left(1 + \frac{m\hbar\gamma B_0}{k_B T}\right)$$

This expression highlight the parameters that affect the different spin populations, namely γ and B_0. The overall magnetic moment *M* of a sample comprising an ensemble of spins can be expressed as the sum of all the individual magnetic moments:

$$M = \gamma\hbar \sum_{m=-I}^{I} mN_m = \gamma\hbar N \sum_{m=-I}^{I} mp_m$$

By introducing $\Delta p_m = p_m - p_{-m}$, we got:

$$M = \gamma\hbar N \sum_{m>0}^{m\in\{-I,\dots,I\}} m\Delta p_m$$

Assuming that we work at temperatures around 298K, we can express Δp_m as:

$$\Delta p_m \simeq \frac{2m\hbar\gamma B_0}{(2I+1)\,k_B T}$$

From these equations, we can express the overall magnetic moment as:

$$M \simeq \gamma\hbar N \sum_{m>0}^{m\in\{-I,\dots,I\}} \frac{2m^2\hbar\gamma B_0}{(2I+1)\,k_B T} = \frac{N\hbar^2\gamma^2 B_0}{(2I+1)\,k_B T} \sum_{m=-I}^{I} m^2$$

With this expression, we directly see that the overall magnetization of a sample is proportional to the magnitude of the applied magnetic field, the total number of sensitive spins in the sample and the square of the gyromagnetic ratio of the associated nucleus. The most important nuclei in biomolecular NMR spectroscopy are the proton ^1H, the nitrogen-15 ^{15}N, the carbon-13 ^{13}C, all with a spin quantum number of $1/2$ and thus two energy levels, and the deuterium ^2H with a spin quantum number of 1. Unlike the proton, most of these nuclei cannot be found naturally in high abundance. One must usually isotopically enrich the studied biomolecule with the nuclei of interest in order to have enough overall magnetization during the experiment. In this thesis, and in biomolecular NMR in general, we deal principally with systems comprising of nuclei with spin quantum numbers of $1/2$. Deuterated water is usually added to the samples as a reference compound to lock and correct over time the slowly decaying magnetic field of the spectrometer.

1.1.2 Nuclear Magnetic Resonance

This section provides a phenomenological description of NMR. Given the fact that NMR deals with quantum mechanical systems, this section is not able to describe all the phenomena arising in NMR, but it gives a good picture of what is happening to the overall magnetization of a simple spin ensemble. When no magnetic field is present in a sample, the orientation distribution of the spin magnetic moments is isotropic. When an external magnetic field is applied, the response of a spin magnetization is to rotate with a rotation axis colinear to the direction of the magnetic field. This precession behavior arises because in addition to a magnetic moment, the nuclei possess angular momentum. Comparable behaviors can be observed on rotating objects such as a spinning gyroscopes. It can be shown that the precession angular frequency of a spin magnetic

moment ω_0, called Larmor frequency, is proportional to the gyromagnetic ratio and the applied magnetic field:

$$\omega_0 = -\gamma B_0$$

If the system is comprised of purely isolated spins with no interaction with the environment, the spin magnetic moments precess forever with the same angle with respect to the magnetic field, and as the initial condition was an isotropic distribution of magnetic moments, upon application of a magnetic field, the overall magnetization remains 0. In reality, a spin system often lies in a very dynamic environment. for example, a tumbling molecule in a liquid is surrounded by ions, water and other molecules that constantly collide and rotate at multiple timescales. Despite the relatively high degree of isolation of an atomic nucleus, all this motion induces small variations in the local magnetic field surrounding our nucleus. These magnetic field variations induce variations in the Larmor frequency as well as in the precession axis. At times longer than the precession period, these small fluctuations break the isotropy of the magnetic moment orientations and a thermal equilibrium is progressively reached yielding the overall magnetization described in the previous section. The build-up of longitudinal magnetization can often be expressed with a single exponential. The rate at which the equilibrium is reached is called the longitudinal or spin-lattice relaxation rate.

NMR signals are detected in the plane perpendicular to the magnetic field. At equilibrium under a magnetic field, there is no transverse magnetization. In order to get magnetization in the transverse plane, an oscillating magnetic field called radio-frequency pulse, with a frequency close to the Larmor frequency (resonance phenomenon) is applied to trigger a rotation of the spin magnetic momenta around the magnetic field induced by the pulse. If the duration is well chosen, it can trigger a 90 degree rotation that puts the overall magnetization in the transverse plane. After the pulse, once the overall magnetization is in the transverse plane, all the spins precess around the static magnetic field, therefore yielding a rotation of the overall magnetization at the Larmor frequency in the transverse plane. This oscillating magnetization yields a signal that can be detected. In the case where the spins don't interact with the environment, the precession will go on forever. But since the local magnetic field is constantly fluctuating, the Larmor frequency of each spins is fluctuating as well. As a result, each spin will not rotate at the same speed at a given time, which induces a progressive dephasing of the spin orientation in the transverse plane. In the long run, the transverse magnetization is decaying until reaching 0 at thermal equilibrium. The rate at which the transverse signal vanishes is called transverse relaxation rate.

Phenomenologically, two processes are observed in an ensemble of single-spin systems: The precession of the overall magnetization around the transverse plane, and relaxation, the return to equilibrium of the magnetization. We identified two relaxation rates so far: longitudinal relaxation and transverse relaxation. These phenomena can be modeled for a single spin system by the Bloch equations. Let's consider our overall magnetization vector in a Cartesian coordinate system $\overrightarrow{M} = \begin{pmatrix} M_x & M_y & M_z \end{pmatrix}^T$. The Bloch equations can be expressed as:

$$\frac{d}{dt}\begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} -R_2 & \gamma B_z & -\gamma B_y \\ -\gamma B_z & -R_2 & \gamma B_x \\ \gamma B_y & -\gamma B_x & -R_1 \end{pmatrix}\begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ R_1 M_0 \end{pmatrix}$$

With M_i and B_i ($i \in \{x, y, z\}$) being the *i* component of the overall magnetization and the magnetic field respectively, in the Cartesian frame, R_1 and R_2 the longitudinal and transverse relaxation rates respectively, M_0 the equilibrium magnetization and γ the gyromagnetic ratio of the studied nucleus. This model is able to predict the behavior of a very simple spin system. If our spin system of interest has more than one spin, and these spins are coupled via J-coupling, the behavior of the spin system cannot be correctly predicted using the Bloch equations anymore: A quantum mechanical description of NMR is necessary.

1.1.3 Quantum Mechanical Description of NMR

A quantum mechanical description of NMR provides a full description of the phenomena observed in NMR including spin relaxation and J-coupling. The evolution of a quantum mechanical system can be described using the Schrödinger equation:

$$\frac{d}{dt}\Psi(t) = -i\hbar^{-1}\hat{H}\Psi(t)$$

With $\Psi(t)$ being the wave function describing the system, $\hat{H}$ the Hamiltonian, $\hbar$ the reduced Plank constant and $i^2 = \sqrt{-1}$. For convenience, we may work with so-called natural units in which we can write $\hbar = 1$. In addition, the Dirac notation will be used. We can express the wave function as a linear combination of the elements of a conveniently chosen orthonormal basis of the given Hilbert space:

$$|\Psi\rangle = \sum_{i=1}^{N} c_i |\psi_i\rangle$$

With $|\psi_i\rangle$ the elements of the basis, c_i the associated coefficients and N the number of eigenstates in the system. For example, a wavefunction in a single spin-$\frac{1}{2}$ system can be expressed as a combination of the vectors $\begin{pmatrix}1 & 0\end{pmatrix}^T$ and $\begin{pmatrix}0 & 1\end{pmatrix}^T$ describing the two eigenstates of the spin $|\alpha\rangle$ and $|\beta\rangle$. In NMR theory, we prefer to describe the evolution of our system in the Liouville space composed by operators acting on the wavefunction as elements, since it allows us to describe the time evolution of the spin ensemble with the density operator (*vide supra*). In this framework, the hat notation is reserved for so-called super-operators, i.e. operators acting on our elements of Liouville space. For a single-spin system, a Cartesian basis of Liouville space is composed by the Pauli matrices $\{I_x, I_y, I_z, E/2\}$ with the following matrix representations:

$$I_x = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix};\ I_y = \frac{1}{2}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix};\ I_z = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix};\ E/2 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

Another similar basis composed by the following operators is often used: $\{I^+, I^-, I_z, E/2\}$. Where $I^{\pm}$ is defined as:

$$I^{\pm} = \frac{1}{2}\left(I_x \pm iI_y\right)$$

This so-called "shift basis" is used to treat relaxation since its basis elements have convenient rotation properties. In addition, such operators are convenient to describe coherence evolution. A basis for a multiple spins system can easily be derived from the tensor products of elements of the single-spin basis. For instance, a two-spin IS system will have the following basis operators in a Cartesian basis:

$$\{2I_xS_x, 2I_yS_x, 2I_zS_x, S_x, 2I_xS_y, 2I_yS_y, 2I_zS_y, S_y, 2I_xS_z, 2I_yS_z, 2I_zS_z, S_z, I_x, I_y, I_z, E/2\}$$

Where we can calculate any operator representation of this basis as a function of the Pauli matrices. For example, we have: $I_iS_j = 2I_i \otimes S_j$. These operators are thus represented with a 4x4 matrix. The factor 2 comes from normalization. Here, E is the 4x4 identity matrix. We can then describe a spin system using the density operator σ. Any density operator can be represented as a linear combination of the basis elements that we just introduced. The matrix elements of σ are defined as:

$$\sigma_{ij} = \langle\psi_i|\Psi\rangle\langle\Psi|\psi_j\rangle = c_ic_j^*$$

From this expression, we see that the diagonal elements of the density operator correspond to the populations of each energy level. The non-diagonal elements correspond to the so-called "coherences" between the different spin states. The presence of non-zero coherence terms implies that there is some overall spin polarization in the transverse plane. From the density operator we can determine the expectation value of an observable property P in a system:

$$\langle\overline{P}\rangle = Tr\left(\sigma P\right)$$

It can also be demonstrated from Schrödinger equation that the time evolution of the density matrix can be expressed as:

$$\frac{d}{dt}\sigma(t) = -i\left[H, \sigma(t)\right] = \hat{L}\sigma(t)$$

With $[H, \sigma(t)] = H\sigma(t) - \sigma(t)H$ the Liouvillian superoperator acting on the density operator. This so-called Liouville-von Neumann equation is one of the most important equations in NMR. It allows us to compute the time evolution of a spin system under a given magnetic field. If the Hamiltonian is time-independent, the solution of this equation is:

$$\sigma(t) = exp\,(-iHt)\,\sigma(0)exp\,(iHt)$$

If the Hamiltonian is time-dependent, it is often possible to perform a convenient frame transformation allowing us to use this solution. Provided that the Hamiltonian is known, it is possible to predict the evolution of a given spin system. In the general case, the total Hamiltonian has numerous contributions, from the static magnetic field, the local electronic environment, and all the other possible interactions that are taking place in the system. For example, we can have:

$$H_{tot} = H_{static} + H_{CS} + H_Q + H_{DD} + H_J$$

With H_{static} the Hamiltonian for the applied static magnetic field, H_{CS} for the chemical shift, representing the small chemical shielding performed by the electronic environment, modifying the magnetic field perceived by the nucleus. In the general case, the chemical shift is a tensor, which means that there is a possible anisotropic contribution yielding time-dependent variations of the local magnetic field upon tumbling of the molecule. H_Q is for quadrupolar interactions. This term is non zero only if we have spin quantum numbers higher than $1/2$. H_{DD} is the term reporting on the through-space interactions that can take place between different neighboring dipoles. H_J is the term reporting on the J-coupling, an electron-mediated through-bond indirect dipolar coupling between different nuclei of the same molecule.

From now on, we are going to consider that we are working with isotropic liquids, which means that there is a uniform distribution of the molecular orientations in space. In this case, the time-dependent dipolar and quadrupolar interactions are averaged out by the rapid tumbling of the molecule. Because of this rapid isotropic tumbling, The J-coupling is averaged to a constant value called scalar coupling and the chemical shift Hamiltonian experiences a similar averaging mechanism. We therefore end up with a total averaged isotropic Hamiltonian as follows:

$$H_{tot}^{iso} = H_{static} + H_{CS}^{iso} + H_J^{iso}$$

This Hamiltonian can usually be expressed as follows for a two-spin system in the Cartesian basis:

$$H_{tot}^{iso} = \omega_I I_z + \omega_S S_z + \pi J 2 I_z S_z$$

With J being the scalar coupling between the spins I and S. Both the static magnetic field and the chemical shift are encoded in the Larmor frequency ω_X for a spin X. It can be expressed as:

$$\omega_X = -\gamma_X\,(B_{static} + B_{induced}) = -\gamma_X B_{static}\left(1 + \delta^{iso}\right)$$

With B_{static} and $B_{induced}$ the static magnetic field of the spectrometer and the magnetic field induced by the chemical shielding of the electronic environment surrounding the given nucleus respectively. δ^{iso} is the isotropic chemical shift, usually expressed in ppm.

The averaged Hamiltonian can be used to describe most of the averaged observable NMR phenomena occurring up to the ms timescale in isotropic liquids including spin precession and J-coupling, as described here. Relaxation mechanisms however are described by rapid rotations from 10 ps to 10 ns timescales and cannot be described with an averaged Hamiltonian. To describe relaxation, the effects of the time-dependent Hamiltonian has to be examined in more detail.

1.2 Theory of NMR Spin Relaxation

We saw in the previous section that the return to equilibrium of a spin system was mediated by spin relaxation rates. We briefly discussed the fact that these mechanisms took place upon coupling of the spin system with its dynamic environment. It is clear from our previous description that the value of these rates are highly dependent on the dynamics of the spin system and its environment. Therefore, it might be possible to extract valuable information from an analysis of NMR spin relaxation rates. In this section, we are going to address the theory of NMR spin relaxation and discuss how we can interpret these rates in terms of protein dynamics and interactions.

1.2.1 Origins of spin relaxation

We identified two spin relaxation mechanisms so far: longitudinal relaxation, corresponding to the magnetization build-up in a magnetic field, and transversal relaxation, corresponding to the progressive dephasing of the coherences leading to a loss of overall magnetization in the transverse plane. Relaxation mechanisms appear because of rapid magnetic field fluctuations occurring at timescales from 10 ps to 10 ns. Let's express our overall Hamiltonian as a function of the time dependent and time independent contributions:

$$H_{tot} = H_{static} + H^{iso} + H^{aniso}(t)$$

The averaged isotropic contributions to the Hamiltonian are time-independent. The remaining time dependent contribution is thus an anisotropic contribution that averages to zero over the ensemble in an isotropic liquid. This term is composed with the interactions that are important for relaxation. We can split this anisotropic Hamiltonian into a longitudinal contribution $H_z^{aniso}(t)$ and a transverse contribution $H_{xy}^{aniso}(t)$:

$$H^{aniso}(t) = H_z^{aniso}(t) + H_{xy}^{aniso}(t)$$

The term $H_{xy}^{aniso}(t)$ corresponds to transversal fluctuations of the magnetic field. These local fluctuations, if containing the frequency corresponding to the energy difference between two spin states, can induce transitions from one state to another. Longitudinal relaxation is caused

by these magnetic fluctuations. In addition to this phenomenon, the spin state transitions induced by transverse magnetic field fluctuations broaden the energy levels of the different possible spin states as a result of a Heisenberg uncertainty principle: The lifetime of a spin in a given state being not infinite, the energy difference between each spin states cannot be perfectly defined, yielding an intrinsic dephasing of the overall transverse magnetization. This term is thus responsible for both longitudinal and transversal relaxation processes

The term $H_z^{aniso}(t)$ corresponds to longitudinal fluctuations of the magnetic field. These fluctuations induce local variations in the larmor frequencies of the spins. As a result, a dephasing of the spins in the transverse plane occurs leading to a loss of coherence. This is the main mechanism for transverse relaxation. It is the so-called "Adiabatic" term, since it is not responsible for any transition in energy levels. Both longitudinal and transversal magnetic field fluctuations are thus responsible for transverse relaxation.

In order to extract information on the dynamics of our molecule, we must first understand what are the interactions that lead to relaxation in the studied spin system. Then, we must know how to calculate our relaxation rates assuming a given dynamic behavior. To do this, we can derive an expression for the relaxation superoperator by looking at the behavior of the time-dependent anisotropic Hamiltonian.

1.2.2 Semiclassical Theory of NMR Spin Relaxation

The physical description of NMR spin relaxation involves non-trivial mathematical manipulations that are useful to address in order to introduce the important assumptions that are made in this theory. A simple but sophisticated enough semi-classical formulation provides the wanted results but a more complete full quantum mechanical description is possible and has been addressed [29, 33]. In this so-called Bloch-Wangsness-Redfield theory, we are going to treat the spin system quantum mechanically, and the magnetic field fluctuations classically [33, 34]. For this, we introduce in the Liouville von Neumann equation a time-dependent Hamiltonian $H_1(t)$ corresponding to a stochastic magnetic fluctuation averaging to 0 over time and with a magnitude very small with respect to the static magnetic field associated with the time-independent Hamiltonian H_0.

$$\frac{d\sigma(t)}{dt} = -i\left[H_0 + H_1(t), \sigma(t)\right]$$

We can then conveniently transform this equation into the so-called *interaction* frame to get rid of H_0 by defining $\sigma^T(t) = e^{iH_0t}\sigma(t)e^{-iH_0t}$ and $H_1^T(t) = e^{iH_0t}H_1(t)e^{-iH_0t}$, which yields after some algebra:

$$\frac{d\sigma^T(t)}{dt} = -i\left[H_1^T(t), \sigma^T(t)\right]$$

Integrating this differential equation yields:

$$\sigma^T(t) = exp\left(-i\int_0^t \left[H_1^T(t'),\right] dt'\right)\sigma^T(0)$$

To address NMR spin relaxation, we have to consider the entire system. We thus take the ensemble averaged density matrix $\overline{\sigma^T(t)}$. The random fluctuations that are responsible for relaxation occur at very fast timescale: picoseconds to nanoseconds typically. On the other hand, relaxation time constants are usually in the second timescale, therefore, we can consider the evolution of the density matrix in the interaction frame and the stochastic fluctuations to be statistically independent, which allows us to separate the averaged terms. After performing a Taylor expansion up to second order and some operator algebra, one obtains the following equation. The overbars indicate ensemble average:

$$\overline{\sigma^T(t)} = \overline{\sigma^T(0)} - i\int_0^t \overline{\left[H_1^T(t'), \sigma^T(0)\right]}dt' - \frac{1}{2}\int_0^t \int_0^{t''} \overline{\left[H_1^T(t'), \left[H_1^T(t''), \sigma^T(0)\right]\right]}dt''dt'$$

Later, we are going to ignore the overbars for the density operator. After differentiation and recalling that the ensemble average of $\hat{H}_1^T(t)$ is 0, one obtains the following equation. We also introduce $\tau = t - t'$:

$$\frac{d\sigma^T(t)}{dt} = -\int_0^t \overline{[H_1^T(t), [H_1^T(t-\tau), \sigma^T(0)]]}d\tau$$

At this point, we cannot easily go into further derivations without making more assumptions. As said earlier, the timescales for magnetic fluctuations are much faster than the timescale that is of interest for NMR spin relaxation. At times significantly greater than the typical magnetic fluctuation correlation time, the value in the integral in the previous equation is close to 0. Therefore, we allow ourselves to extend the limit of our integral to infinity. In addition, the timescales associated with relaxation-active mechanisms are much smaller than the relaxation time constant. At those considered timescales, the density matrix evolution is very small, and we can allow ourselves to replace $\sigma^T(0)$ with $\sigma^T(t)$ in the equation. It is legitimate to argue at this point that considering timescales smaller than the relaxation times is not justifiable since relaxation is what we are interested in. To correct this we can consider a time period T higher than the relaxation time constant. We can divide this time period into multiple smaller durations t much shorter than the relaxation time periods that satisfy our assumptions. If we proceed piece by piece to evaluate relaxation at time T, we obtain the expected result. Finally, a last adjustment needs to be made. The limitation of the semi-classical treatment of this problem is that it implicitly assumes an infinite temperature for the lattice, leading to no overall magnetization at equilibrium. It has been shown that introducing the final equilibrium magnetization σ_0 described in the first section overcomes this problem and leads to the same result as a fully quantum mechanical treatment. We thus replace $\sigma^T(t)$ with $\sigma^T(t) - \sigma_0$ and obtain the following equation:

$$\frac{d\sigma^T(t)}{dt} = -\int_0^\infty \overline{[H_1^T(t), [H_1^T(t-\tau), \sigma^T(t) - \sigma_0]]}d\tau$$

From here, we can allow ourselves to go back to the laboratory frame. In order to do so effectively, we must express the stochastic Hamiltonian in a convenient way. We may express the Hamiltonian as a linear combination of irreducible spherical tensors of our operator space. Such a transformation can appear as a complication, but the irreducible spherical tensor transformations under rotations are fairly simple and allow a simple transformation back to the laboratory frame. The elements of the shift basis of Liouville space defined earlier are all spherical tensors. We define the irreducible spherical spin operators of rank k, T_k^q, and we express them as a combination of elements of a spherical basis of our operator space:

$$T_k^q = \sum_p T_{kp}^q$$

From this, we can express the stochastic Hamiltonian as:

$$H_1(t) = \sum_{q=-k}^{k} \sum_p (-1)^q F_k^{-q}(t) T_{kp}^q$$

F_k^q are random functions of spatial variables that carry the time-dependent stochastic behavior of the Hamiltonian. The $(-1)^q$ terms are taken out of the random functions by convention and are sometimes inside the F_k^q terms in older textbooks [29]. It can be shown that the elements T_{kp}^q are eigenfunctions of the Liouvillian operator, with eigenvalues ω_p^q:

$$\hat{L}_0 \left(T_{kp}^q \right) = \left[H_0, T_{kp}^q \right] = \omega_p^q T_{kp}^q$$

From this property we can write:

$$exp\left(i\hat{L}_0 t \right) T_{kp}^q = exp\left(i\omega_p^q t \right) T_{kp}^q$$

Using the properties of this useful class of operators and assuming that the random functions F_k^q and $F_k^{q'}$ are statistically independent, which implies that only terms with $q' = -q$ survive, we can derive the following expression:

$$\frac{d\sigma^T(t)}{dt} = -\sum_{q=-k}^{k} \sum_{p,p'} exp\left(i(\omega_p^q - \omega_{p'}^q)t \right) \left[T_{kp}^{-q}, \left[T_{kp}^q, \sigma^T(t) - \sigma_0 \right] \right]$$
$$\times \int_0^\infty \overline{F_k^q(t) F_k^{-q}(t-\tau)} exp\left(-i\omega_p^q \tau \right) d\tau$$

Finally, a so-called *secular approximation* is made to obtain the final result. The terms $\left|\omega_p^q - \omega_{p'}^q\right|$ that are sufficiently greater than zero can be neglected since the factor $exp(i(\omega_p^q - \omega_{p'}^q)t)$ would oscillate rapidly and average to zero faster than the timescales of relaxation. If there is no degeneracy in the eigenfrequencies, only the terms $p = p'$ would actually survive. We end up with the following expression:

$$\frac{d\sigma^T(t)}{dt} = -\sum_{q=-k}^{k}\sum_{p}\left[T_{kp}^{-q},\left[T_{kp}^{q},\sigma^T(t)-\sigma_0\right]\right]\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp\left(-i\omega_p^q\tau\right)d\tau$$

The integral can be separated into its real part and its imaginary part. The term inside the integral is the correlation function of the $F_k^q(t)$ depending on spatial variables and expressed as: $\overline{F_k^q(t)F_k^{-q}(t-\tau)}$. Since a correlation function is real and even, the real part of its Fourier transform is even as well, and the imaginary part is odd. Therefore, we can rewrite the real parts $j^q(\omega_p^q)$ as the real parts of the Fourier transforms of the correlation functions of $F_k^q(t)$ the following way:

$$j^q(\omega_p^q) = 2Re\left[\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp\left(-i\omega_p^q\tau\right)d\tau\right]$$

$$j^q(\omega_p^q) = Re\left[\int_{-\infty}^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp\left(-i\omega_p^q\tau\right)d\tau\right]$$

And we can rewrite the imaginary parts as follows:

$$k^q(\omega_p^q) = 2Im\left[\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}exp\left(-i\omega_p^q\tau\right)d\tau\right]$$

The fact that the imaginary part is an odd function permits convenient simplifications that allow us to rewrite our differential equation in the laboratory frame as follows:

$$\frac{d\sigma(t)}{dt} = -i\left[H_0,\sigma(t)\right] - i\left[\Delta,\sigma(t)\right] - \hat{\Gamma}\left(\sigma(t)-\sigma_0\right)$$

with Δ being the so-called dynamic frequency shift coming from the imaginary part of our integrand. This part can be conveniently incorporated into H_0, yielding the final Liouville-von Neumann equation:

$$\frac{d\sigma(t)}{dt} = -i\left[H_0,\sigma(t)\right] - \hat{\Gamma}\left(\sigma(t)-\sigma_0\right)$$

$\hat{\Gamma}$ is the relaxation superoperator, expressed as:

$$\hat{\Gamma} = \frac{1}{2}\sum_{q=-k}^{k}\sum_{p}\left[T_{kp}^{-q},\left[T_{kp}^{q},\right]\right] Re\left[\int_{-\infty}^{\infty}\overline{F_k^q(t)F_k^{-q}(t-\tau)}exp\left(-i\omega_p^q\tau\right)d\tau\right]$$

The relaxation superoperator is thus a sum of products of two terms. One term is a double commutation operator acting on the density operator. It acts like selection rule. There is relaxation only if these terms yield non zero values. The second term is called the power spectral density function, defined as the Fourier transform of the correlation function of F_k^q. This correlation function is key to extracting valuable information from spin relaxation rates. Relaxation rates are thus expressed as a linear combination of spectral density functions.

1.2.3 Correlation Function and Spectral Density Function

In isotropic liquids and at our usual temperatures (high temperature limit), it can be shown that the spectral density functions $j^q(\omega)$ can all be expressed as a function of one single spectral density function, such that [35]:

$$j^q(\omega) = -1^q j^0(\omega)$$

Additionally, we recall that the mechanisms of interest for relaxation here are related to rotations, and rotations are associated with traceless rank 2 tensors. Therefore, only rank 2 spin operators are considered here. These mathematical considerations lead to the definition of a unique spectral density function $j(\omega)$ such that:

$$j(\omega) = Re\left[\int_{-\infty}^{\infty}\overline{F_2^0(t)F_2^0(t-\tau)}exp\left(-i\omega\tau\right)d\tau\right]$$

The spectral density function can be interpreted as the (non normalized) density of probability for a fluctuation of F_2^0 to occur at a given frequency. The autocorrelation function $\overline{F_2^0(t)F_2^0(t-\tau)}$ is associated with the ensemble and time averaged correlation of the function F_2^0 with itself over time. If we assume that this function is associated with the orientation of a rigid molecule undergoing rotational Brownian motion, a fast tumbling of the molecule will be associated with a correlation function that decays to zero rapidly, since the orientation of each molecule becomes quickly uncorrelated. On the other hand, a molecule that tumbles slowly will have a correlation function that converges slowly to zero.

The random function $F_2^0(t)$ can be factored by a function that depends on spatial variable and physical constant $d_0(t)$ with another function related to time-dependent polar angles $\Theta_t = (\theta(t), \phi(t))$ expressed with respect to the magnetic field axis and associated with a vector that points toward the principal axis of the considered interaction. We usually choose the modified second-order spherical harmonic Y_2^0 for this function. The modified second-order spherical harmonics

Y_k^q are obtained by multiplying the conventional normalized spherical harmonics $\mathcal{Y}_k^q$ by a factor of $\sqrt{4\pi/(2k+1)}$. We thus have, using the symmetry properties of the correlation function:

$$\overline{F_2^0(t)F_2^0(t-\tau)} = \overline{d_0(t)d_0(t+\tau)Y_2^0(\Theta_t)Y_2^0(\Theta_{t+\tau})} = G(\tau)$$

The function $d_0(t)$ depends on the interaction that is considered. Among the interactions responsible for relaxation, we have for example the dipolar interactions for which $d_0(t)$ would depend on the distance between the two considered spins and their gyromagnetic ratios. The principal axis for this interaction would be the vector pointing from one nucleus to the other. Another interaction, related to the possible anisotropy of the chemical shift (CSA), would imply a $d_0(t)$ term that depends on the magnetic field and the principal values of the chemical shift tensor. The principal axis would consequently point toward the principal axis of this tensor.

In some cases, the terms in d_0 are time-independent. For example, in a protein backbone amide system comprising two $\frac{1}{2}$-spins: the nitrogen ^{15}N and its proton, the two main mechanisms responsible for relaxation of the nitrogen magnetization are the dipolar interaction between these two spins and the CSA of the nitrogen nucleus. In this case, we can assume the distance between the two spins to be time-independent, as well as the other variables. If we do so, we can factorize the correlation function as follows:

$$G(\tau) = \overline{d_0(t)d_0(t+\tau)Y_2^0(\Theta_t)Y_2^0(\Theta_{t+\tau})} = d_0^2\overline{Y_2^0(\Theta_t)Y_2^0(\Theta_{t+\tau})} = d_0^2C(\tau)$$

With $C(\tau)$ the normalized time autocorrelation function of the modified second rank zero order spherical harmonic Y_2^0. We can also define the orientational spectral density function $J(\omega)$ such that:

$$j(\omega) = d_0^2 Re\left[\int_{-\infty}^{\infty} \overline{Y_2^0(\Theta_t)Y_2^0(\Theta_{t+\tau})} exp\,(-i\omega\tau)\,d\tau\right] = d_0^2 J(\omega)$$

The auto-correlation function is always a decaying function that converges to zero in a liquid. Consequently, the corresponding spectral density function will also yield a decaying function. The form of these functions depend on the global and local dynamic properties of the axis of interest in the considered molecule.

1.2.4 Analytical expression for relaxation of a spin system

To get an analytical expression for relaxation in a given spin system, one must know the mechanisms that are going to provoke relaxation. The main relaxation-active processes are through-space dipolar interactions between nuclei, chemical shift anisotropy, J couplings and quadrupolar couplings. In an isotropic liquid, for a given relaxation mechanism, the relaxation superoperator can be written as follows:

$$\hat{\Gamma} = \frac{1}{2} d_0^2 \sum_{q=-2}^{2} \sum_{p} \left[T_{2p}^{-q}, \left[T_{2p}^{q}, \right]\right] Re\left[\int_{-\infty}^{\infty} \overline{Y_2^0(\Theta_t) Y_2^0(\Theta_{t+\tau})} exp\left(-i\omega\tau\right) d\tau\right]$$

If we express the system's density operator as a vector in one basis of Liouville space, we can express the relaxation superoperator $\hat{\Gamma}$ that contains all the relaxation contributions as a matrix. Considering only relaxation, the equation for the evolution of the system can be written as:

$$\frac{d\sigma(t)}{dt} = -\hat{\Gamma}\left(\sigma(t) - \sigma_0\right)$$

This equation implies that each element of the density matrix relaxes with its own relaxation rate. The matrix representation of the relaxation operator in Liouville space is called relaxation matrix or Redfield kite in the literature[31, 36]. The diagonal elements of this matrix are the autorelaxation rate constants for each elements of the used basis. The off-diagonal elements, if non-zero, correspond to cross relaxation rate constants between one operator and another. Cross relaxation occurs in cases where the two considered operators have the same coherence order and degenerate transitions. In other terms, if the two operators oscillate at very different frequencies, the exchange in polarization between the two coherences is averaged out, leading no apparent cross-relaxation. For the protein backbone amide system, the Redfield kite is shown in Figure 1.1. To calculate a relaxation rate constant, one just takes the corresponding matrix element of the Redfield matrix. For a relaxation rate Γ_{ij} between basis operators B_i and B_j, this translates as:

$$\Gamma_{ij} = \frac{\langle B_i | \hat{\Gamma} B_j \rangle}{\langle B_i | B_i \rangle}$$

If we look at the ^{15}N-labelled protein backbone amide spin system, the two main relaxation-active mechanisms are the ^{15}N $-^{1}$ H dipolar interaction and the ^{15}N Chemical shift anisotropy. These mechanisms are now going to be introduced

The dipolar interaction

Every nucleus with a non-zero spin possesses a magnetic moment and reacts to the presence of another magnetic moment. When two nuclei with non-zero spin are close to each other, these two nuclei will interact though a dipole-dipole interaction. The strength of this interaction of course depends on the distance between the two nuclei and the size of their magnetic moment. The Hamiltonian for a dipolar interaction between two nuclei can be written as:

$$H_{dipolar} = d_{ij}\left(3\left(I_i \cdot e_{ij}\right)\left(I_j \cdot e_{ij}\right) - I_i \cdot I_j\right)$$

with

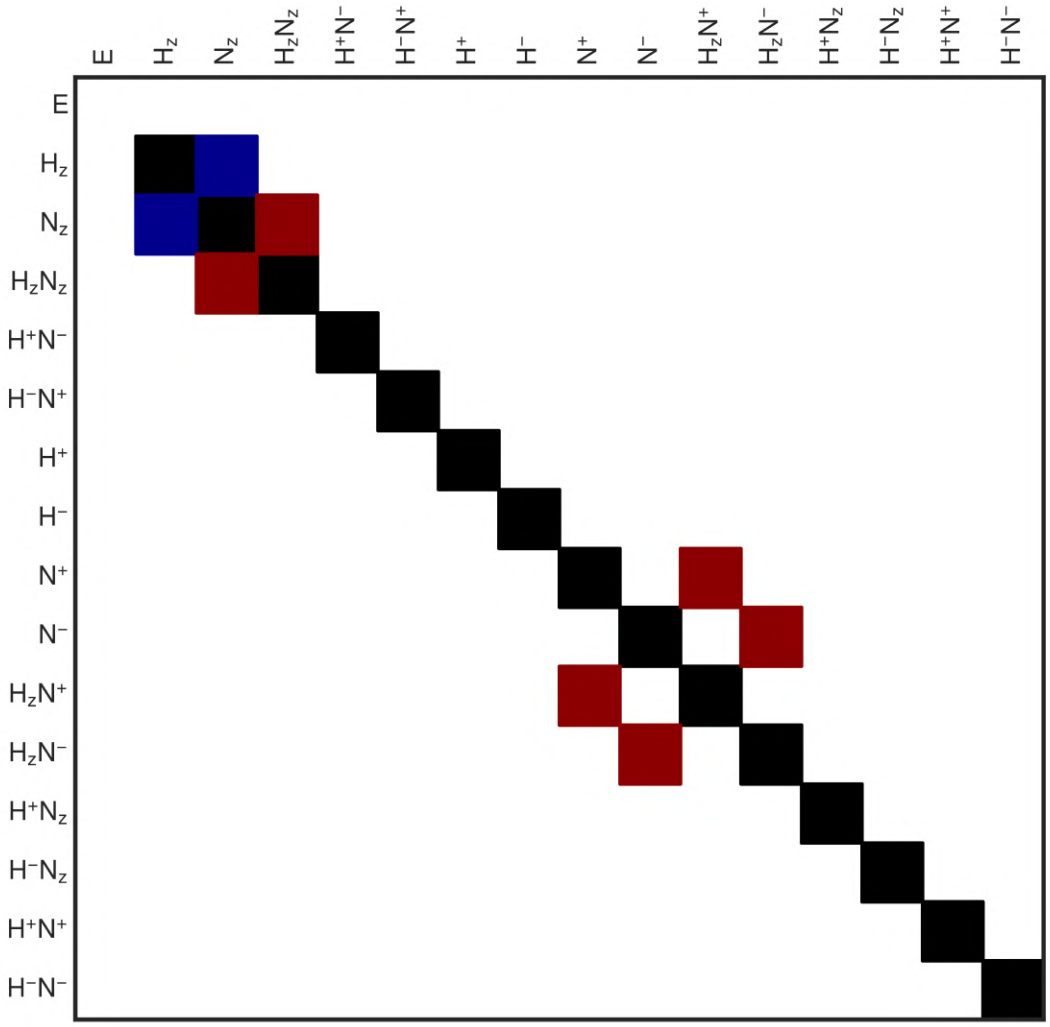


Figure 1.1: Redfield matrix for the backbone amide spin system. The black squares correspond to the autorelaxation rates. The blue squares correspond to the dipolar longitudinal cross relaxation rate and the red squares come from Dipole-CSA cross-correlated relaxation rates.

$$d_{ij} = -\frac{\mu_0 \gamma_i \gamma_j \hbar}{4\pi r_{ij}^3}$$

Where e_{ij} is a unit vector pointing in the direction of the axis between the nuclei i and j associated with their spin operator I_i and I_j and their gyromagnetic ratio γ_i and γ_j. r_{ij} is the distance between the two nuclei. The factor d_{ij} corresponds to the factor d_0 in the expression of the relaxation superoperator presented earlier. For the N-H system of the protein backbone amide, we would have:

$$d_0^{dipolar} = -\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}$$

Where r_{NH} is the N-H bond distance.

The Chemical Shift Anisotropy

The chemical shift comes from the local electronic environment. The electrons of the molecule indeed modify the local magnetic field. The static external magnetic field creates a current in the cloud of electrons. This current induces a magnetic field on its own that in turn alters the experienced magnetic field for the nuclei. We can write the experienced magnetic field for a given nucleus as follows:

$$\vec{B}_{experienced} = \vec{B}_{static} + \vec{B}_{induced} = \vec{B}_0 + \delta\vec{B}_0$$

Where δ is chemical shift rank 2 tensor. If we time-average the chemical shift Hamiltonian in an isotropic liquid, the observed effective chemical shift becomes a simple scalar. However, at the relaxation-active timescales, the chemical shift is a tensor and the effective shielding depends on the local orientation of the chemical shift tensor's principal axis. The chemical shift tensor can be written:

$$\delta = \delta^{iso} + \delta^{aniso}$$

With $\delta^{iso} = \sigma_{iso}E$ where E is the identity matrix and $\sigma_{iso} = (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})/3$ where σ_{ij} are the matrix elements of the chemical shift tensor. The traceless anisotropic component of the chemical shift tensor is the part responsible for relaxation.

For protons, the CSA is usually negligible, but it is important for some other nuclei like the amide ^{15}N. If we assume that the CSA is uniaxial, we can write the following for the prefactor associated with the relaxation superoperator associated with the protein backbone ^{15}N:

$$d_0^{CSA} \propto \Delta\sigma\gamma_N B_0$$

B_0 is the spectrometer magnetic field and $\Delta\sigma = \sigma_{\parallel} - \sigma_{\perp}$ assuming an axially symmetric CSA with principal values $\sigma_{zz} = \sigma_{\parallel}$ and $\sigma_{xx} = \sigma_{yy} = \sigma_{\perp}$. A notable property of the CSA is that its contribution to relaxation increases with the magnetic field of the spectrometer.

1.2.5 Relaxation in the rotating frame

Applying a constant rf field in an NMR experiment has interesting consequences to the system. The rf field is defined as a rotating magnetic field component on the transverse plane, associated with a Hamiltonian $H_{rf}(t)$. To remove the time dependence of the total Hamiltonian, a transformation into the rotating frame associated with the rf field is necessary before the transfer into the interaction frame. Ignoring the scalar coupling, this means that we transfer the system into a double tilted frame. In this frame, the magnitude of the magnetic field experienced by a given nucleus can be written as:

$$B_{eff} = \sqrt{b_1^2 + \Delta B_0^2} = B_1/sin(\theta)$$

Where b_1 is the magnitude of the rf field and ΔB_0 is the reduced static magnetic field experienced in the double tilted frame, which depends on the chemical shift of the considered nucleus associated with the frequency ω_0 with the relation: $\Delta B_0 = (\omega_{rf} - \omega_0)/\gamma = \Omega/\gamma$. The term θ corresponds to the angle between the effective field B_{eff} and the z-axis of the laboratory frame (Figure 1.2) defined as:

$$tan(\theta) = \frac{\omega_1}{\Omega}$$

As a result of this, the terms ω_p^q in the spectral density functions become $\omega_p^q + \omega_p^{q(rf)}$ with:

$$\sum_{i=1}^{N} \omega_{rf,i} \left[I_{zi}, A_{kp}^q \right] = \omega_q^{q(rf)} A_{kp}^q$$

Where $\omega_{rf,i}$ is the rotating frame frequency associated with the i^{th} spin and N the number of irradiated spins in the spin system. With an rf field of strength ω_1 with a correlation time τ_c in which $\omega_1\tau_c \ll 1$, we can consider that $j^q(\omega_p^q + \omega_p^{q(rf)}) \simeq j^q(\omega_p^q)$. We can then calculate the relaxation rate constants as follows:

$$\Gamma'_{ij} = \frac{\left\langle U^{-1}A'_iU|\hat{\Gamma}\left(U^{-1}A'_jU\right)\right\rangle}{\langle A'_i|A'_i\rangle}$$

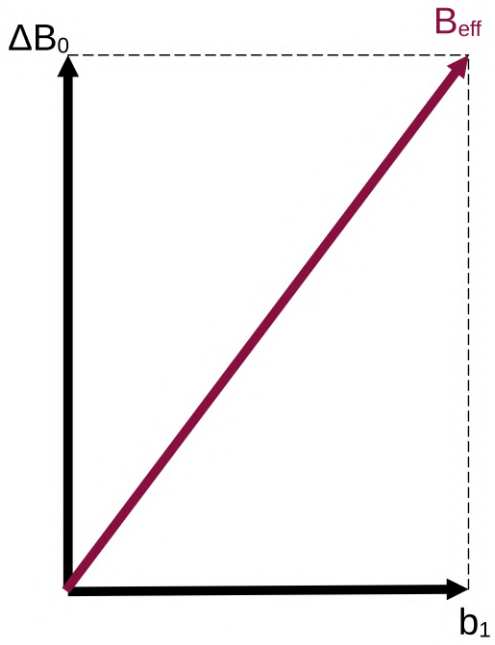


Figure 1.2: Representation of the effective magnetic field experienced by a nucleus (red) in the rotating frame with a static magnetic field B_0 aligned with the z-axis and an orthogonal b_1 field.

Where A_i' is our operator A_i expressed in the rotating frame and U is a unitary transformation operator defined as:

$$U = exp\left(i\sum_{i=1}^{N}\theta_i I_{yi}\right)$$

We assumed an rf field applied along x. θ_i corresponds to the tilt angle between the laboratory frame z-axis and the rotating frame z'-axis associated with the spin i. We can define the longitudinal relaxation rate in the tilted rotating frame $R_{1\rho}$ as the relaxation of the magnetization along the axis z' of the tilted frame. From our considerations, it follows that $R_{1\rho}$ can be expressed as a function of R_1, R_2 and the tilt angle θ for a given nucleus with the following expression assuming no chemical exchange contribution to R_2:

$$R_{1\rho} = R_1 cos^2\theta + R_2 sin^2\theta$$

$R_{1\rho}$ is often used combined with R_1 to obtain R_2. Another application of $R_{1\rho}$ is it's measurement with different rf field strengths and rf frequencies to characterize chemical exchange processes. Similarly, we could define the transverse component $R_{2\rho}$.

1.2.6 Longitudinal and Transversal auto-relaxation rates, application to protein backbone ^{15}N

R_2 is the rate for which the magnetization in the transverse plane with respect to the magnetic field decays to zero. The derivation for this rate using the formula derived with the semi-classical theory yields, for two dipolar-coupled nuclei N and H with a CSA on the nucleus N:

$$R_2 = \frac{1}{20}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2 \left(\sigma_\parallel - \sigma_\perp\right)^2 (4J(0) + 3J(\omega_N))$$

We interpret this expression for the protein backbone amide but this formula is valid for any dipolar-coupled nuclei with a CSA on the nucleus of interest. We can identify in the formula the two contributions from the dipolar interaction and the ^{15}N CSA. We can also identify the 5 angular frequencies in the spectral density function for which this relaxation rate is sensitive: 0, ω_N, $\omega_H + \omega_N$, ω_H and $\omega_H - \omega_N$ in ascending order. These are the eigenfrequencies ω_p^q derived from the semiclassical treatment. In fact, the formula for all the relaxation rates in this system contain spectral density function values evaluated exclusively for some of these eigenfrequencies. If we examine spectral density functions (figure 1.5), we see that $J(0)$ is the largest value in $J(\omega)$. Usually, in the case of biomolecules where tumbling is slow with respect to smaller molecules, R_2 is dominated by $J(0)$. In the simple case of a rigid molecule exhibiting isotropic tumbling, $J(0)$ is simply proportional to τ_c. R_2 would in this case be proportional to the correlation time of the molecule. Figure 1.3 illustrates the evolution of R_2 as a function of τ_c. Most of the time, the spectral density function cannot be modeled with a simple single Lorentzian, but it is a reasonable approximation to say that the distribution of R_2 over the sequence gives an estimate of the evolution of the tumbling of the N-H axis along the sequence of the protein.

On the other hand, R_1 is the build up of ^{15}N magnetization along the magnetic field axis. The derivation yields:

$$R_1 = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2 \left(\sigma_\parallel - \sigma_\perp\right)^2 J(\omega_N)$$

R_1 is expressed here as a linear combination of $J(\omega_N)$, $J(\omega_H)$, $J(\omega_H + \omega_N)$ and $J(\omega_H - \omega_N)$. For this rate, $J(\omega_N)$ is the largest term. If we look at the evolution of R_1 as a function of the correlation time in our simple rigid molecule example, we see that R_1 is more difficult to interpret with a simple look. R_1 exhibits a maximum value for a specific correlation time, then it decreases if τ_c is increased or decreased with respect to this value (Fig. 1.3).

Measurement

The measurement of relaxation rates is performed with the following scheme: First, the proton spin is excited. Since it has a high gyromagnetic ratio compared to the nitrogen of carbon nuclei, we obtain a significantly larger magnetization compared to a direct nitrogen or carbon excitation. The magnetization is then transferred to the nucleus of interest, typically an INEPT

scheme. Then, the magnetization is prepared for relaxation evolution. For R_1 and $R_{1\rho}$, the magnetization must be along the z-axis prior to relaxation. After the relaxation period, we let the magnetization of our spin of interest to evolve with an incremented delay to encode the indirect dimension to obtain a 2D spectrum. Finally, the magnetization is transferred back to proton for acquisition, for example with a reverse INEPT scheme. In order to obtain our relaxation rate of interest, different measurements of the NMR signal with different relaxation delays are performed. To obtain a relaxation rate expressed as an exponential decay of the magnetization for example, one performs this experiment with several relaxation delays, and then fit the intensity of the NMR signal as a function of the relaxation delay with a decaying exponential function. The rate of decay corresponds to the relaxation rate.

1.2.7 Cross-relaxation rate

The Redfield matrix in (figure 1.1) for the protein backbone amide NH two-spins system highlights the presence of a cross-relaxation term between the operators H_z and N_z. The derived expression for this rate writes as follows:

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 \left(6J\left(\omega_H + \omega_N\right) - J\left(\omega_H - \omega_N\right)\right)$$

σ_{NH} only depends on J evaluated at high frequencies and thus reports on fast timescales dynamics exclusively. This cross relaxation is possible since both operators H_z and N_z correspond to a zero-order coherence pointing along the same constant axis and thus not oscillating. This term is the manifestation of a longitudinal magnetization exchange through dipolar coupling between the proton and the nitrogen-15. This effect is called the Nuclear Overhauser Effect (NOE). It is the basis of the NOE SpectroscopY (NOESY) experiment in which two close dipolar-coupled spins exchange magnetization during a mixing delay, allowing for instance to obtain the structure of a protein through restraint-based optimization. In our case, we take advantage of this effect to obtain high frequency information on the spectral density function by measuring the rate of magnetization exchange.

Measurement

To obtain this rate on a protein, we measure the steady-state NOE enhancement that we will write nOe. The principle is to record two similar experiments where we observe a signal proportional to the ^{15}N magnetization. In one experiment, we saturate the protein's amide protons with a weak selective radiofrequency field for a time sufficiently longer than the longitudinal relaxation time to reach a steady-state. nOe is extracted by comparing the z-magnetization of the steady-state nitrogen spin with and without saturation of the proton. Taking into account auto-relaxation and cross-relaxation, the overall nitrogen z-magnetization evolves according to the following equation:

$$\frac{dN_z^{ss}}{dt} = -R_1\left(N_z^{ss} - N_z^0\right) + \sigma_{NH} H_z^0$$

Where N_z^{ss} stands for the steady-state z-magnetization of the ^{15}N nucleus and N_z^0 and H_z^0 are the equilibrium z-magnetizations of ^{15}N and ^{1}H respectively. Since we consider a steady-state, this derivative equals 0. Therefore, if we record two identical experiments with the exception

of the saturation time present in only one of them, and where the signal is proportional to the steady-state nitrogen z-magnetization, the ratio between the signal intensity of both experiments equals:

$$\frac{I_{saturation}}{I_{reference}} = \frac{N_z^{ss}}{N_z^0} = 1 + \frac{\sigma_{NH}}{R_1}\frac{H_z^0}{N_z^0} = 1 + \frac{\sigma_{NH}\gamma_H}{R_1\gamma_N}$$

Since this parameter also depends on R_1, extracting σ_{NH} also requires the measurement of the longitudinal relaxation rate.

1.2.8 Relaxation interference and cross-correlated relaxation rates

When more than one time-dependent Hamiltonian is causing relaxation of a given spin, interference between these different Hamiltonians may occur. We can rewrite the overall stochastic Hamiltonian as:

$$H_1(t) = \sum_i \sum_{q=-k}^{k} \sum_p (-1)^q F_{ik}^{-q}(t) T_{ikp}^q$$

Where i stands for the different interactions responsible for relaxation. With this expression, a given relaxation rate expressed as the matrix element of the Redfield kite is written as:

$$\Gamma_{rs} = \sum_i \Gamma_{rs}^i + \sum_{i,j}^{i\neq j} \Gamma_{rs}^{ij}$$

Where Γ_{rs}^i is the relaxation rate constant due to the ith interaction and Γ_{rs}^{ij} is the relaxation rate constant coming from the cross-correlation between the ith and jth interactions. We can define the cross-correlation spectral density function $j_{ij}^q(\omega)$ associated to the cross-correlation term as follows:

$$j_{ij}^q(\omega) = Re\left[\int_{-\infty}^{\infty} \overline{F_{ik}^q(t) F_{jk}^{-q}(t+\tau)} exp\left(-i\omega\tau\right) d\tau\right]$$

When two relaxation mechanisms are correlated, the corresponding cross-spectral density function doesn't quench: $j_{ij}^q(\omega) \neq 0$ and cross-correlated relaxation might occur. In protein backbone amides, the cross-correlation effect of interest comes from the correlation between the ^{15}N CSA and the ^1H-^{15}N dipolar interaction. This effect yields a longitudinal and a transversal component, η_z and η_{xy} respectively (Figure 1.1). The derived expressions for these rates are as follows:

$$\eta_z = \frac{1}{15} P_2\left(\cos\theta\right)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)\left(\sigma_\parallel - \sigma_\perp\right)\omega_N\left(6J\left(\omega_N\right)\right)$$

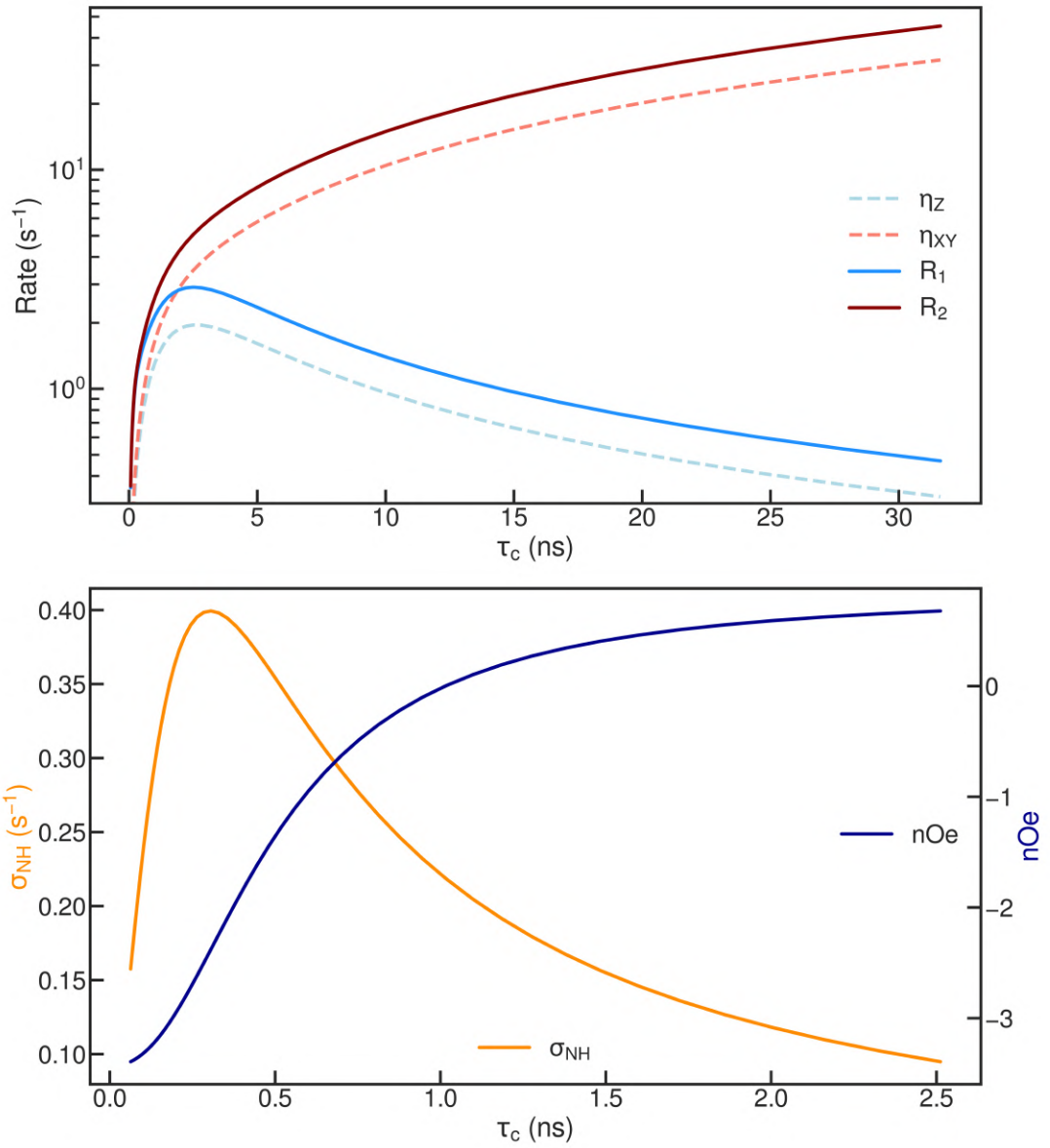


Figure 1.3: ^{15}N relaxation rates as a function of the rotational correlation time in the case of a single Lorentzian spectral density function. Left: R_1 (Blue), R_2 (Red), η_{XY} (Dashed red) and η_Z (Dashed blue) as a function of the correlation time. Right: Dipolar crossrelaxation rate (Orange) and nuclear Overhauser enhancement (Blue) as a function of the correlation time. The calculations were done with a magnetic field of 600 MHz proton frequency, a CSA tensor of -172 ppm, an NH bond distance of 1.02 Angström and an angle between the CSA and the NH bond vector of 22°.

$$\eta_{xy} = \frac{1}{15} P_2(\cos\theta) \left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3} \right) (\sigma_\parallel - \sigma_\perp)\, \omega_N\, (4J(0) + 3J(\omega_N))$$

Where θ is the angle between the CSA principal axis and the dipolar interaction principal axis, and P_2 is the second order Legendre function $P_2(x) = \frac{1}{2}(3x^2 - 1)$. η_{xy} is often dominated by the term $J(0)$ and yields a similar pattern as R_2. η_z on the other hand depends on $J(\omega_N)$ and so gives similar information to R_1 about the spectral density function (Fig. 1.3). We will see later that it can be of interest to measure these rates as a complement, especially η_{xy}. The drawback of using these relaxation rates is that it requires a good knowledge of the relative orientation of the CSA principal axis with respect to the NH bond vector, represented by the angle θ here as well as the level of anisotropy of the CSA tensor. In principle, these values are estimated around 22.5° and 172 ppm for the angle and the tensor anisotropy respectively for backbone amide nitrogens, but it can vary along the sequence and induce a bias in relaxation data analysis. The optimization of this parameter for each site can be included in a relaxation analysis provided that enough relaxation data is present.

measurement

If we examine the Redfield kite (Figure 1.1), we see that these rates translate to a magnetization exchange between the operators N_z and H_zN_z for the longitudinal contribution and between $N^\pm$ and $H_zN^\pm$ for the transverse contribution. The most accurate way of measuring cross-relaxation rates is to use symmetrical reconversion [37, 38]. The principle is to measure four similar experiments, allowing us to observe the rate of exchange between in-phase ($N^\pm$) and anti-phase ($H_zN^\pm$) terms. In each experiment, we prepare nitrogen magnetization, then we evolve it during a relaxation period where we allow cross-relaxation. In one experiment, we evolve in-phase magnetization and we observe the residual in-phase through a signal intensity proportional to it. In another experiment, we observe the produced anti-phase instead. The two other experiments are similar but starting with evolving anti-phase magnetization (Figure 1.4).

The chemical shift evolution taken apart, the components N^+ and N^+H_z of the density operator evolve as follow:

$$\frac{dN^+}{dt} = -i\pi J_{NH} 2H_zN^+ - \left(R_{2,N}^{DD} + R_2^{CSA}\right) N^+ - \eta_{xy} 2H_zN^+$$

$$\frac{d2H_zN^+}{dt} = -i\pi J_{NH} N^+ - \left(R_{2,NH}^{DD} + R_2^{CSA}\right) 2H_zN^+ - \eta_{xy} N^+$$

Where $R_{2,N}^{DD}$ and $R_{2,NH}^{DD}$ are the in-phase and antiphase transverse relaxation rate of the ^{15}N nucleus from the dipolar interaction respectively and R_2^{CSA} is the transverse relaxation rate from the ^{15}N CSA. This can be translated into a matrix equation. If we ignore the J-coupling evolution (if for example we refocus the J-coupling during the given relaxation period), we end up with the following:

$$\frac{d}{dt}\begin{pmatrix} N^+ \\ H_zN^+ \end{pmatrix} = -\begin{pmatrix} R_{2,N}^{DD} + R_2^{CSA} & \eta_{xy} \\ \eta_{xy} & R_{2,NH}^{DD} + R_2^{CSA} \end{pmatrix} \begin{pmatrix} N^+ \\ H_zN^+ \end{pmatrix}$$

If from this equation we calculate the intensity ratio associated with the signal intensity from the anti-phase magnetization I_{ia} with the signal intensity from the residual in-phase magnetization I_{ii} in an experiment starting with in-phase magnetization, we would obtain the following result:

$$\frac{I_{ia}}{I_{ii}} = tanh(\eta_{xy}T)$$

By symmetry, we obtain:

$$\frac{I_{ai}}{I_{aa}} = tanh(\eta_{xy}T)$$

With T being the duration of the relaxation delay. If we combine the four experiments, we obtain:

$$\eta_{xy} = \frac{1}{T} arctanh\left(\sqrt{\frac{I_{ia}I_{ai}}{I_{ii}I_{aa}}}\right)$$

The advantage of measuring four experiments instead of only two is that such symmetrical scheme cancels out some errors due to pulse miscalibrations and other uncontrolled sources of error [37]. An implementation of this experiment is presented in figure 1.4.

Apart from providing insight into the spectral density function, cross-correlated relaxation rates are also exploited in the widely used TROSY sequences. If we define H^{α} and H^{β}, corresponding to the proton magnetizations in its two eigenstates responsible for the doublet in an uncoupled nitrogen spectrum, we obtain the following matrix equation:

$$\frac{d}{dt}\begin{pmatrix} N^{+}H^{\alpha} \\ N^{+}H^{\beta} \end{pmatrix} = -\begin{pmatrix} i\pi J_{NH} + \overline{R_2} + \eta_{xy} & \left(R_{2,N}^{DD} - R_{2,NH}^{DD}\right)/2 \\ \left(R_{2,N}^{DD} - R_{2,NH}^{DD}\right)/2 & -i\pi J_{NH} + \overline{R_2} - \eta_{xy} \end{pmatrix}\begin{pmatrix} N^{+}H^{\alpha} \\ N^{+}H^{\beta} \end{pmatrix}$$

If we assume that $2\pi J_{NH}^2 >> (R_{2,N}^{DD} - R_{2,NH}^{DD})/2$, the matrix becomes diagonal. We see that the two peaks of the doublet don't have the same relaxation rate. In a decoupled HSQC, we in fact will have four peaks whose relaxation rates differ because of cross-correlated relaxation. One peak ends up with a narrower linewidth. It is thus possible, by looking only at this narrow peak, to improve the quality of the spectra in so-called TROSY (Transverse Relaxation Optimized SpectroscopY) sequences.

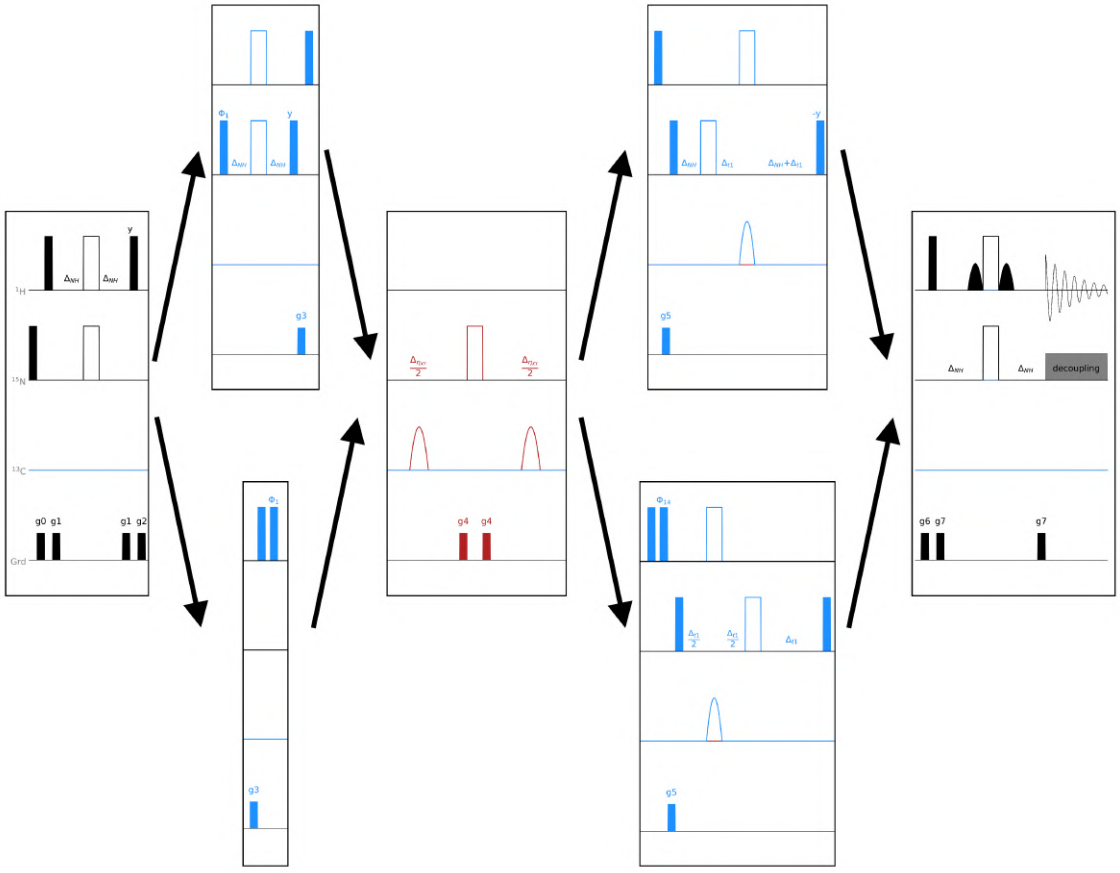


Figure 1.4: Symmetrical reconversion pulse sequence for the measurement of DD/CSA transverse cross-correlated relaxation rates in backbone amide nitrogens. The first block corresponds to the excitation of the proton and transfer to nitrogen. The second blocks (Blue) corresponds to the conversion to in-phase (top) or selection of antiphase (bottom) magnetization prior to the constant time relaxation period in the third block (Red). Then, conversion or selection of magnetization is performed along with chemical shift evolution on the nitrogen in a fourth block (Blue) before transfer back to proton and acquisition in the last block. All the gradients are purging or artifact correction gradients, filled and unfilled bars are 90 and 180 degree hard pulses respectively with a phase x unless stated otherwise. The shaped filled and unfilled domes are selective pulses on water for water suppression and adiabatic decoupling pulses on carbon respectively. Δ_{NH} is a delay typically set to $1/(4J_{NH})$, $\Delta_{\eta_{XY}}$ is the relaxation period and Δ_{t1} is the incremented indirect dimension evolution period. Φ_1 is [x, x, x, x, -x, -x, -x, -x], Φ_{14} is [x, x, -x, -x] and the receiver phase is [x, -x, -x, x, -x, x, x, -x].

1.3 Analysis and Interpretation of ^{15}N NMR spin relaxation

A set of relaxation rates such as R_1, R_2 and σ_{NH} measured at different magnetic fields provides valuable information on the spectral density function. Most of the relaxation analysis methods are based on obtaining an approximate solution for the spectral density function so that it fits the relaxation data. From high field NMR spin relaxation rates at different magnetic fields, one still has little information compared to the intrinsically complex nature of the dynamics of proteins (especially flexible parts). The information is restricted to specific locations in the spectral density function. The main challenge of High-field NMR spin relaxation analysis is to find an analysis framework that gives a comprehensive understanding of the protein's dynamics and that is simple enough to be justifiable given the limited number of parameters that are known. Many explicit models were derived in early studies and proposed a way to express the rotational correlation function in protein systems assuming a given shape for the protein [39] or a given type of motion like restricted diffusion and jump models [40–43] before a more general model-free method was proposed by Lipari and Szabo in 1982 [44]. This method has been extensively used and extended over the years until now and provided a valuable tool for understanding protein dynamics from 10 ps to 10 ns timescales [45–50]. In the meantime, numerous different techniques were introduced as an attempt to propose a different approach to NMR spin relaxation analysis. In this section, we discuss relaxation analysis techniques, by first examining the simplest case of a rigid spherical molecule tumbling in an isotropic liquid. Then, the model-free analysis and its extension to Intrinsically Disordered Proteins is introduced. The use of this method to model IDP dynamics at different temperatures and viscosity will then be presented. Finally, alternative and complementary methods are discussed.

1.3.1 Relaxation of rigid molecules undergoing rotational diffusion

Let's consider a simple spin ensemble of spherical molecules of radius R associated with an NMR signal relaxing via an interaction with a principal axis that depends only on the orientation of the spherical molecule. This spherical molecule undergoes isotropic rotational Brownian motion. We define $P(\Theta, t|\Theta_0)$ the probability for the interaction principal axis to be with an orientation defined by the polar angles Θ at time t knowing that it was with an orientation Θ_0 initially. This probability function obeys the rotational diffusion equation:

$$\frac{dP(\Theta, t|\Theta_0)}{dt} = D_r \Delta P(\Theta, t|\Theta_0)$$

Δ is the angular Laplacian operator and D_r is the rotational diffusion constant that can be expressed with the rotational Stokes-Einstein equation:

$$D_r = \frac{k_B T}{8\pi R^3 \eta}$$

Where k_B is Boltzmann's constant, T is temperature and η is the shear viscosity of the solution experienced by the molecule. The initial conditions translate to the Dirac function $P(\Theta, 0|\Theta_0) = \delta(\Theta - \Theta_0)$. We may look for solutions of the diffusion equation in the form of an expansion of spherical harmonics since any square-integrable function on a sphere can be expressed this way:

$$P(\Theta, t|\Theta_0) = \sum_k \sum_{l=-k}^{k} c_k^l(t) Y_k^l(\Theta_t)$$

Where the c_k^l are time-dependent coefficients. This form is convenient since the spherical harmonics are eigenfunctions of the angular Laplacian: $\Delta Y_k^l(\Theta) = -k(k+1)Y_k^l(\Theta)$. The diffusion equation thus simplifies to the following:

$$\frac{dc_k^l(t)}{dt} = -D_r k(k+1) c_k^l(t)$$

From this very simple first-order differential equation it follows that:

$$c_k^l(t) = c_k^l(0) exp\left(-tD_r k(k+1)\right) = c_k^l(0) exp\left(-\frac{t}{\tau_k}\right)$$

Where we defined the correlation time τ_k as:

$$\tau_k = \frac{1}{D_r k(k+1)} = \frac{8\pi R^3 \eta}{k_B T k(k+1)}$$

The final solution can then be inferred from the fact that Dirac functions can be easily expressed as a function of spherical harmonics:

$$\delta(\Theta - \Theta_0) = \sum_{k,l} Y_k^{l*}(\Theta_0) Y_k^l(\Theta)$$

From this we can identify $c_k^l(0) = Y_k^{l*}(\Theta_0)$ which yields the final solution for P:

$$P(\Theta, t|\Theta_0) = \sum_{k,l} Y_k^{l*}(\Theta_0) Y_k^l(\Theta_t) exp\left(-\frac{t}{\tau_k}\right)$$

As defined earlier, in NMR, spin relaxation is calculated from the correlation function associated with the relaxation-active interactions in the spin system ensemble. Looking at the orientational correlation function associated with the current example of a nucleus in a rigid spherical molecule undergoing Brownian motion in an isotropic liquid, the correlation function would be expressed as follows, by definition of the ensemble average:

$$C(\tau) = \overline{Y_2^0(\Theta_0)Y_2^0(\Theta_\tau)} = \int\int Y_2^0(\Theta_0)Y_2^0(\Theta_\tau)P(\Theta, t|\Theta_0)P(\Theta_0)d\Theta d\Theta_0$$

The expression for $P(\Theta, t|\Theta_0)$ comes from Brownian motion as described earlier and the *a priori* probability $P(\Theta_0)$ of finding the relaxation-active interaction's principal axis at a given orientation is constant and equal to $1/4\pi$ from the isotropic properties of the solution. The expression thus becomes:

$$C(\tau) = \frac{1}{4\pi}\int\int Y_2^0(\Theta_0)Y_2^0(\Theta_t)\sum_{k,l} Y_k^{l*}(\Theta_0)Y_k^l(\Theta_t)exp\left(-\frac{t}{\tau_k}\right)d\Theta d\Theta_0$$

Since the spherical harmonics are orthogonal to each other and that they are modified from their original normalized form by a factor of $\sqrt{4\pi/5}$, it follows that the correlation function is a mono-exponential function expressed as:

$$C(\tau) = \frac{1}{5}exp\left(-\frac{t}{\tau_2}\right)$$

The correlation time τ_2 depends here on physical parameters that depend on the size of the molecule, the nanoviscosity of the solution and the temperature as:

$$\tau_2 = \frac{4\pi R^3\eta}{3k_BT}$$

This rotational correlation time corresponds by definition to the average time required for the orientation of the axis of interest to rotate by one radian. The corresponding orientational spectral density function, Fourier transform of this correlation function, is thus a single Lorentzian expressed as:

$$J(\omega) = \frac{2}{5}\frac{\tau_2}{1+\tau_2^2\omega^2}$$

Unfortunately, the fate of the factors $1/5$ for the correlation function and $2/5$ for the spectral density function, not specific to this particular case, vary in the literature. Sometimes, it is injected in the $d_0(t)$ factor in order to have a normalized correlation function, while sometimes it is kept in the orientational correlation function and spectral density function respectively to keep track of the origin of these factors. For consistency with the literature associated with this thesis [51–53], the first mentioned convention will be used unless stated otherwise. An illustration of these functions are given in (figure 1.5) for correlation times of 0.2 and 2 ns in the case of a mono-exponential correlation function. For a shorter correlation time, the correlation function

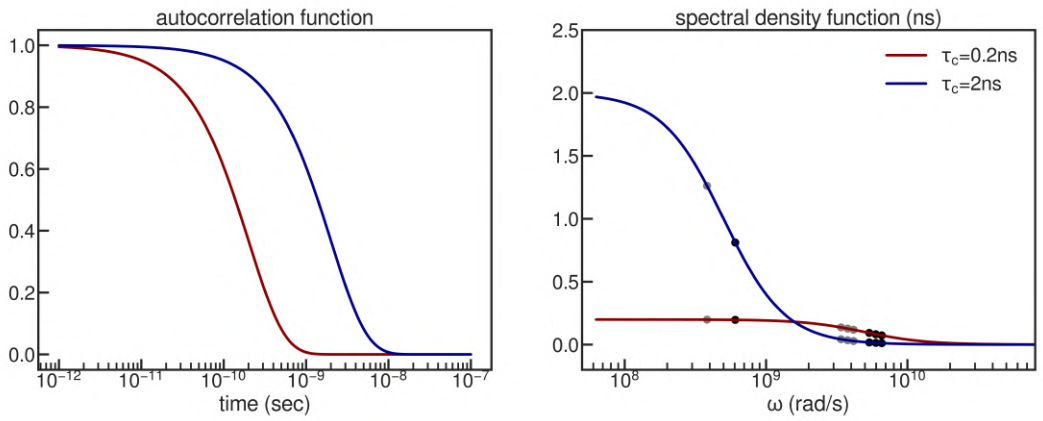


Figure 1.5: Left: Normalized time autocorrelation function for a single exponential decay with a correlation time of 0.2 (Red) and 2 ns (Blue). Right: Associated spectral density functions. The grey and black dots correspond to the spectral density function evaluated at the eigenfrequencies corresponding to a magnetic field of 600 MHz and 950 MHz proton frequency respectively.

converges to zero faster. A shorter correlation time also exhibit a spectral density function that is spread over a larger range of frequencies. Physically, for this system, it means that for a molecule that has a shorter correlation time, the probability for the motion to contain higher frequencies is higher and *vice versa* for low frequencies.

This simple example is a demonstration of the wealth of information NMR spin relaxation contains. An example of application of these expressions is the treatment of water molecules in a simple solution. Water molecules represent most of the time the majority of the molecules in a biological solution. The behavior of the water molecules must therefore translate the properties of the overall solution to a good extent. Anatole Abragam pointed out [29] that the longitudinal relaxation rates of water protons may be proportional to the nanoviscosity of the solution provided that some assumptions are verified. This proportionality relationship allows an NMR spectroscopist to obtain valuable information on the nanoviscosity in a solution. NMR spin relaxation in water protons arises from the dipolar couplings with the neighboring proton of the same molecule as well as with the protons from the surrounding water molecules. Let's first consider relaxation from intra-molecular dipolar coupling. The axis of interest is the proton-proton vector in the water molecule. Let's assume that water molecules are rigid and spherical with rotation properties expressed with a rotational diffusion equation. The equation would give:

$$\frac{dP(\Omega,t)}{dt} = D_r \Delta P(\Omega,t)$$

Where Δ is the angular Laplacian operator, $P(\Omega,t)$ is the probability of finding the proton-proton axis in the direction Ω at a time t, R is the radius of the spherical molecule and D_r is the rotational diffusion constant expressed according to the Stokes-Einstein equation as described earlier. This equation was derived earlier and we showed that it yields a Lorentzian spectral density function. From this and the BWR theory of NMR spin relaxation, the longitudinal relaxation rate associated with the intramolecular proton-proton dipolar coupling is expressed as:

$$R_1^{intra} = \frac{3}{4}\left(\frac{\mu_0 \hbar \gamma_H^2}{4\pi r_{HH}^3}\right)^2 \left(\frac{\tau_c}{1+\omega_H^2\tau_c^2} + \frac{4\tau_c}{1+4\omega_H^2\tau_c^2}\right)$$

If $\omega_H \tau_c \ll 1$, then the denominators in the Lorentzians are very close to 1 and the expression for the relaxation rate may be simplified as:

$$R_1^{intra} = \frac{15}{4}\left(\frac{\mu_0 \hbar \gamma_H^2}{4\pi r_{HH}^3}\right)^2 \tau_c$$

This means that in this case, the longitudinal relaxation rate is proportional to the rotational correlation time. In addition, we have from the diffusion and Stokes-Einstein equation:

$$\tau_c = \frac{1}{6D_r} = \frac{4\pi R^3 \eta}{3k_B T}$$

Which means that the intramolecular contribution of the longitudinal relaxation rate is proportional to the nanoviscosity provided that the correlation time is much smaller than the proton Larmor period.

The intermolecular contribution to relaxation in water protons is more complicated to describe. Abragam describes this contribution using the translational motions of water molecules and by deriving the translational diffusion equation. For the intermolecular dipolar interaction between two distant spins, since the distance between the two spins is time-dependent, this distance cannot be separated from the orientational part in the correlation function as done in a previous section. This yields the following correlation function:

$$G(t) = \left\langle \frac{Y_2^q(\Theta_0)}{r_0^3} \frac{Y_2^q(\Theta_t)}{r^3(t)} \right\rangle = \int\int \frac{Y_2^q(\Theta_0)}{r_0^3} \frac{Y_2^q(\Theta_t)}{r^3(t)} P(r,t|r_0) d^3r_0 d^3r$$

With $P(r,t|r_0)$ the probability of being at a distance r from one spin considering the initial distance r_0. It can be expressed using the translational diffusion equation where the initial probability is $P(r,0|r_0) = \delta(r-r_0)$. This equation yields the following solution as derived by Abragam:

$$P(r,t|r_0) = (8\pi D t)^{-\frac{3}{2}} exp\left(-\frac{(r-r_0)^2}{4Dt}\right)$$

With D the translational diffusion constant. By integrating the correlation function for each spin (considering a density of N spins per cm^3) to obtain the spectral density function, Abragam obtained the following expression for the intermolecular contribution of water proton longitudinal relaxation:

$$R_1^{inter} = \frac{3\gamma_H^4 \hbar^2 \pi^2 N \eta}{k_B T}$$

Once again, to obtain this expression, one must assume that $\omega_H \tau_c \ll 1$. In these conditions, the relaxation of water protons can be considered proportional to nanoviscosity. We have, for A and B two proportionality constants and η the nanoviscosity:

$$R_1^{water} = R_1^{inter} + R_1^{intra} = A\eta + B\eta = (A + B)\eta$$

We can define a parameter ρ called solvent friction, expressed as a function of a nanoviscosity value of reference η_0 and an effective nanoviscosity for a given solution with a different nanoviscosity η, such that:

$$\rho = \frac{\eta - \eta_0}{\eta_0} = \frac{R_1 - R_{1,0}}{R_{1,0}}$$

Where R_1 and $R_{1,0}$ are the longitudinal spin relaxation rates of the water protons in solutions associated with nanoviscosities of η and η_0 respectively. The required approximation allowing us to simplify the spectral density function and obtain this proportionality relationship implies that the spectral density function is constant around ω_H, which translates to a field-independent R_1 of water protons. A good practice to make sure that this approximation is valid is thus to measure the R_1 of water at several magnetic fields. At a magnetic field of 600MHz proton frequency, $\omega_H = 3.770.10^9 rad.s^{-1}$. For the relationship to be strictly valid, the water rotational correlation time must thus be much smaller than $0.265ns$. At 298K, the rotational correlation time of water molecules in a simple water solution has been measured to be around 1.7ps [54], which falls within the required condition.

Of course, not all rigid molecules are undergoing isotropic rotational diffusion. In the general case, a rigid molecule with any non spherical shape undergoes anisotropic diffusion, which means that the diffusion equation becomes:

$$\frac{dP}{dt} = div\left(D_r \overrightarrow{grad}\,(\Delta P)\right)$$

Where D_r is now a diffusion tensor with three eigenvalues corresponding to its principal components D_{XX}, D_{YY} D_{ZZ}. Woessner [39] showed that in this case, the correlation function can be expressed as a sum of exponentials as follows:

$$C(\tau) = \frac{1}{5}\sum_{i=1}^{5} A_i exp\left(-\frac{t}{\tau_i}\right)$$

Where the A_i are coefficients that depends on the relative orientation of the principal axis with respect to the molecular or diffusion tensor frame and τ_i are the corresponding correlation times. This gives the following Lorentzian:

$$J(\omega) = \frac{2}{5}\sum_{i=1}^{5} A_i \frac{\tau_i}{1+\omega^2\tau_i^2}$$

Of course, the more asymmetric the tensor is, the more different from each other the correlation times are, and in the isotropic limit everything is reduced to a single exponential. This description is useful to analyze relaxation rates of rigid folded proteins [55]. However, it is rare that relaxation can only be described by pure overall motion since the motion of proteins is complex and occurring at different scales in space and time. Numerous complicated explicit models that are difficult to distinguish from each other were introduced in the past to attempt an interpretation of NMR relaxation rates, before an analysis method finally provided a "model-free" description of protein dynamics.

1.3.2 Model Free analysis of Folded Proteins

The Model Free Analysis for NMR spin relaxation rates of macromolecules in solution was introduced in 1982 by Lipari and Szabo [44], quickly after similar work from Halle and Wennerstrom [56]. The strength of this method is that it requires no initial assumption on the type of motion occurring in the considered molecule, an explicit geometric model is unnecessary. The more general required assumptions for the theory were presented later on by Bertil Halle [57]. Let's consider a folded protein were relaxation is triggered by overall tumbling and internal motion. If we assume that the overall tumbling and the internal motion are statistically independent, the correlation function associated to spin relaxation $C(t)$ can be expressed as the product of the correlation function associated to the overall tumbling of the protein $C_O(t)$ with the correlation function for the internal motion $C_I(t)$:

$$C(t) = C_O(t)C_I(t)$$

The second required assumption is that the overall tumbling of the protein is isotropic. In this case, the overall rotational tumbling of the protein can be expressed as a single exponential correlation function $exp(-t/\tau_c)$ as demonstrated earlier, where τ_c is the correlation time associated with the overall tumbling. No assumptions have been made for the internal auto-correlation function so far. For a folded protein, the internal correlation function doesn't decay to zero, but rather towards an asymptotic value S^2 called order parameter. This order parameter represents the amount of restriction for the internal motion. An order parameter of 1 would mean a fully restricted condition where no internal motion can occur. An order parameter of 0 would mean a fully unrestricted motion where the principal axis vector can sample the entire rotational spherical space. These considerations lead to the following Model Free formula for the correlation function:

$$C(t) = exp(-t/\tau_c)\left(S^2 + (1 - S^2)C_I(t)\right)$$

If we now assume that the internal motion correlation function can be modeled with a single exponential, we obtain:

$$C(t) = exp(-t/\tau_c)\left(S^2 + (1 - S^2)exp(-t/\tau_I)\right) = S^2 exp(-t/\tau_C) + (1 - S^2)exp(-t/\tau)$$

With $\tau^{-1} = \tau_C^{-1} + \tau_I^{-1}$. The corresponding spectral density function would correspond to a sum of Lorentzians:

$$J(\omega) = S^2 \frac{\tau_c}{1 + (\omega\tau_c)^2} + (1 - S^2)\frac{\tau}{1 + (\omega\tau)^2}$$

This formula is relevant for folded proteins only. In the case of more dynamic molecules, the method must be extended to account for the increased complexity of the proteins motion. Clore and coworkers proposed an extension of this analysis where the internal motion, more complex, is represented by two components instead of one [45].

A rigorous theoretical basis of the different model-free analyses was given by Bertil Halle in 2009 [57]. He introduces different approaches with different assumptions for these approaches. An extensive mathematical description of each case was presented by decomposing the interaction axis of interest into several components by defining several reference frames. He defines the laboratory frame (L) and the global molecular frame (G) first. Then, he introduces the local frame (I) associated with the internal motion of the protein. To account for possible anisotropy of the internal motion, he introduces an alignment frame (D) fixed with respect to the molecular frame, but aligned with the potential of mean torque (POMT) of the internal motion. Finally, he defines the interaction frame (F), usually aligned (dipolar interaction for example) or slightly rotated (CSA interaction for example) with the internal diffusion frame. In his paper, the so-called MF-A refers to the model described above where there's two statistically independent motions with the slow motion undergoing isotropic tumbling. The so-called MF-B yields the same formula as MF-A but comes from a different set of assumptions called "adiabatic approximation", including a separation of timescales for motion decoupling and an uniaxial potential of mean torque (POMT) for the internal motion with respect to the alignment frame (D). He also defined a hybrid version containing three dynamic modes with a combined superposition approximation for the two slowest components and an adiabatic approximation for the two internal motions. Such extension to three dynamic modes instead of two corresponds to the model proposed by Clore and coworkers [45].

1.3.3 Model Free analysis of Intrinsically Disordered Proteins

In disordered proteins, the concept of overall rotational tumbling is expected to be less relevant. To illustrate this fact, let's consider an IDP containing a highly extended N-terminus part containing a high amount of Glycine residues (the smallest and most flexible aminoacid). On the other end, the C-terminus appears more globular-like and rich in heavy and bulky amino-acids. In such

a case, we can expect the N-terminus chain of the protein to tumble rapidly, while the more bulky part would undergo slower tumbling and thus exhibit higher correlation times. In the general case, it is thus not possible to define a single overall correlation time for the protein. Instead, we may define later a sequence specific correlation time associated with the motion that quenches the correlation function. In addition, the internal motion is more complex and thus cannot be expressed by a single exponential. For intrinsically disordered proteins, the correlation function is often modeled with three exponentials corresponding to three separated timescales. This yields a spectral density function expressed as a sum of three Lorentzians:

$$J(\omega) = \sum_{k=1}^{3} A_k \frac{\tau_k}{1 + \omega^2 \tau_k^2}$$

Where τ_k is the correlation time associated with the kth Lorentzian and $\sum A_k = 1$. The relevance of this expression can be demonstrated by examining the considerations taken by Halle [57]. All the considerations that were made in Halle's paper assumed the existence of a fixed global molecular frame (G). For Intrinsically Disordered Proteins, such consideration is irrelevant and another description of the theory is thus necessary. For IDPs, we allow ourselves to describe the N-H bond vector with three dynamic modes possessing three distinct timescales verifying the adiabatic approximation as defined by Halle as proposed by Brüschweiler and colleagues and applied to IDPs in our laboratory by Abyzov & Salvi *et al.* [51]. We are going to write the correlation function of interest as:

$$C(\tau) = \overline{Y_2^0(\Theta_0^L) Y_2^0(\Theta_\tau^L)}$$

With Θ_0^L and Θ_t^L the polar angles associated to the orientation of the interaction tensor of interest (Dipolar or CSA) in the laboratory frame at lag times 0 and τ respectively. We note that the factor $1/5$ associated with the orientational correlation function is encoded in this expression. We will now define several reference frames that will allow us to transform our random function from the laboratory frame (L) to the interaction frame. First, we define the slow motion frame (S) corresponding to the fixed frame with respect to the slowest process quenching the orientational correlation function. Then, we define the intermediary motion frame (I) corresponding to the fixed frame with respect to the intermediary process. In between, we define a director frame aligned with the POMT of the intermediary process with respect to the frame (S). We then define the fast motion frame (I') fixed with respect to the fast process. We also define an alignment frame (D') between (I) and (I'). Finally, the interaction frame (F) is pointing towards the principal axis of the relaxation-active interaction tensor. With such definitions, our random function in the interaction frame can be transformed into the laboratory frame with the following equation, using the convenient transformation properties of spherical harmonics:

$$Y_2^0(\Theta_t^L) = \sum_{k,l} \sum_{m,n} \sum_{p,q} D_{0k}^{2*}(\Omega_t^{LS}) D_{kl}^{2*}(\Omega^{SD}) D_{lm}^{2*}(\Omega_t^{DI}) D_{mn}^{2*}(\Omega^{ID'}) D_{np}^{2*}(\Omega_t^{D'I'}) D_{pq}^{2*}(\Omega^{I'F}) Y_2^q(0,0)$$

Where k, l, m, n, p and q are summed from -2 to +2 and the $D_{ij}^{2}(\Omega_t^{AB})$ are Wigner rotation matrices evaluated at time t at Euler angles Ω^{AB} between the frames A and B. We assumed an isotropic tumbling for the slowest component. Since the orientation $\Omega^{I'F}$ between the fastest motion axis and the interaction axis is assumed constant and $Y_2^q(0,0) = \delta_{q0}$, we can write the random process as follows by defining the geometrical coefficient $\sigma_p^* = D_{p0}^{2*}(\Omega^{I'F})$:

$$Y_2^0(\Theta_t^L) = \sum_{k,l}\sum_{m,n}\sum_{p} D_{0k}^{2*}(\Omega_t^{LS}) D_{kl}^{2*}(\Omega^{SD}) D_{lm}^{2*}(\Omega_t^{DI}) D_{mn}^{2*}(\Omega^{ID'}) D_{np}^{2*}(\Omega_t^{D'I'}) \sigma_p^*$$

With $\sum_p |\sigma_p|^2 = 1$. The adiabatic approximation implies that for two given degrees of freedom, a time interval exists for which the slow process can be considered constant while the orientation associated with the fast process at the end of this interval becomes independent of its initial orientation. The stochastic process expressed in the laboratory frame can be expressed as:

$$Y_2^0(\Theta_t^L) = V_s(\Theta_t^L) + V_i(\Theta_t^L) + V_f(\Theta_t^L)$$

$V_s(\Theta_t^L)$ is the slow component of our random function expressed as:

$$V_s(\Theta_t^L) = \left\langle Y_2^2(\Theta_t^L) \right\rangle_{T_i}$$

Where the brackets correspond to an average over a time T_i corresponding to a time interval where the slow process is constant and the intermediary process is uncorrelated with its initial condition as described in the adiabatic approximation. $V_i(\Theta_t^L)$ is the component associated with the intermediary process, expressed as:

$$V_i(\Theta_t^L) = \left\langle Y_2^0(\Theta_t^L) - V_s(\Theta_t^L) \right\rangle_{T_f} = \left\langle Y_2^0(\Theta_t^L) \right\rangle_{T_f} - V_s(\Theta_t^L)$$

Where the brackets correspond to an average over a time T_f corresponding to a time interval where the slow and intermediary processes are constant while the fast process is uncorrelated with its initial condition as described in the adiabatic approximation. The second equality arises from the statistical independence of the two processes which arise from the timescale separation. $V_f(\Theta_t^L)$ is the component associated with the fastest processes. It is expressed as:

$$V_f(\Theta_t^L) = Y_2^0(\Theta_t^L) - V_i(\Theta_t^L) - V_s(\Theta_t^L) = Y_2^0(\Theta_t^L) - \left\langle Y_2^0(\Theta_t^L) \right\rangle_{T_f}$$

The associated orientational correlation function can be expressed as:

$$C(\tau) = G_{ss}(\tau) + G_{si}(\tau) + G_{sf}(\tau) + G_{is}(\tau) + G_{ii}(\tau) + G_{if}(\tau) + G_{fs}(\tau) + G_{fi}(\tau) + G_{ff}(\tau)$$

With $G_{ab}(\tau) = \overline{V_a(\Theta_0^L)V_b(\Theta_\tau^L)}$ the auto and cross-correlation functions for the different processes. Due to the timescale separation, all the cross-correlations vanish and we end up with the following correlation function:

$$C(\tau) = G_{ss}(\tau) + G_{ii}(\tau) + G_{ff}(\tau)$$

The slow fluctuating component associated with the slowest correlation function can be expressed as:

$$V_s(\Theta_t^L) = \sum_{k,l}\sum_{m,n}\sum_{p} D_{0k}^{2*}(\Omega_t^{LS})D_{kl}^{2*}(\Omega^{SD})\left\langle D_{lm}^{2*}(\Omega_t^{DI})\right\rangle D_{mn}^{2*}(\Omega^{ID'})\left\langle D_{np}^{2*}(\Omega_t^{D'I'})\right\rangle \sigma_p^*$$

If we assume now as done in Halle's MF-B that the local POMT's are uniaxial in both frames *D* and *D'*, we can write:

$$\left\langle D_{lm}^{2*}(\Omega_t^{DI})\right\rangle = \delta_{l0}\left\langle D_{0m}^{2*}(\Omega_t^{DI})\right\rangle$$

and

$$\left\langle D_{np}^{2*}(\Omega_t^{D'I'})\right\rangle = \delta_{n0}\left\langle D_{0p}^{2*}(\Omega_t^{D'I'})\right\rangle$$

Which simplifies our expression for the slow process:

$$V_s(\Theta_t^L) = \sum_{k} D_{0k}^{2*}(\Omega_t^{LS})D_{k0}^{2*}(\Omega^{SD})\sum_{m}\sum_{p}\left\langle D_{0m}^{2*}(\Omega_t^{DI})\right\rangle D_{m0}^{2*}(\Omega^{ID'})\left\langle D_{0p}^{2*}(\Omega_t^{D'I'})\right\rangle \sigma_p^*$$

And $G_{ss}(\tau)$ can be expressed as:

$$G_{ss}(\tau) = S^2 G_{slow}(\tau)$$

With S a generalized order parameter define here as:

$$S^2 = \left| \sum_{m,p} \sigma_p \left\langle D^2_{0p}(\Omega^{D'I'}) \right\rangle D^2_{m0}(\Omega^{ID'}) \left\langle D^2_{0m}(\Omega^{DI}) \right\rangle \right|^2$$

This parameter can be factored as:

$$S^2 = \left| \sum_{p} D^2_{p0}(\Omega^{I'F}) \left\langle D^2_{0p}(\Omega^{D'I'}) \right\rangle \right|^2 \left| \sum_{m} D^2_{m0}(\Omega^{ID'}) \left\langle D^2_{0m}(\Omega^{DI}) \right\rangle \right|^2 = S_f^2 S_i^2$$

S_f^2 correspond to the scaling of the correlation function by the fast component, and S_i^2 correspond to the scaling due to the intermediary process. The slow component correlation function can be expressed as:

$$G_{slow}(\tau) = \sum_{k} \sum_{k'} D^2_{k0}(\Omega^{SD}) D^{2*}_{k'0}(\Omega^{SD}) \left\langle D^2_{0k}(\Omega^{LS}_0) D^{2*}_{0k'}(\Omega^{LS}_\tau) \right\rangle$$

From the isotropic solution and the orthogonality, normalization and unitary properties of the Wigner rotation matrices [58], it follows that the slow correlation function G_{slow} decays from 1/5 to 0. We can note that the uniaxiality of the local POMT in frame D' is not necessary for decoupling the slow component normalized correlation function $G_{slow}(\tau)$ with a generalized order parameter. In this case, the generalized order parameter becomes:

$$S^2 = \left| \sum_{m,n,p} \sigma_p \left\langle D^2_{np}(\Omega^{D'I'}) \right\rangle D^2_{mn}(\Omega^{ID'}) \left\langle D^2_{0m}(\Omega^{DI}) \right\rangle \right|^2$$

The intermediate component is expressed as:

$$V_i(\Theta^L_t) = \left\langle Y^0_2(\Theta^L_t) \right\rangle_{T_f} - V_s(\Theta^L_t)$$

Which translates, after factorization, to the following:

$$V_i(\Theta^L_t) = \sum_{k,l,m,n,p} D^{2*}_{0k}(\Omega^{LS}_t) D^{2*}_{kl}(\Omega^{SD}) \left(D^{2*}_{lm}(\Omega^{DI}_t) - \left\langle D^{2*}_{lm}(\Omega^{DI}) \right\rangle \right) D^{2*}_{mn}(\Omega^{ID'}) \left\langle D^{2*}_{np}(\Omega^{D'I'}) \right\rangle \sigma^*_p$$

The associated correlation function becomes:

$$G_{ii}(\tau) = \sum_{k,k'}\sum_{l,l'}\sum_{m,m'}\sum_{n,n'}\sum_{p,p'} \times$$

$$\left\langle D^{2*}_{0k}(\Omega_0^{LS}) D^{2*}_{kl}(\Omega^{SD}) \left(D^{2*}_{lm}(\Omega_0^{DI}) - \left\langle D^{2*}_{lm}(\Omega^{DI})\right\rangle\right) D^{2}_{0k'}(\Omega_t^{LS}) D^{2}_{k'l'}(\Omega^{SD}) \left(D^{2}_{l'm'}(\Omega_t^{DI}) - \left\langle D^{2}_{l'm'}(\Omega^{DI})\right\rangle\right)\right\rangle$$

$$\times D^{2*}_{mn}(\Omega^{ID'}) D^{2}_{m'n'}(\Omega^{ID'}) \left\langle D^{2}_{np}(\Omega^{D'I'})\right\rangle \sigma_p \left\langle D^{2*}_{n'p'}(\Omega^{D'I'})\right\rangle \sigma^*_{p'}$$

For brevity we are going to define X_m as:

$$X_m = D^{2*}_{mn}(\Omega^{ID'}) \left\langle D^{2*}_{np}(\Omega^{D'I'})\right\rangle \sigma_p$$

We may write:

$$V_i(\Theta_t^L)_{k,l,m} = D^{2*}_{0k}(\Omega_t^{LS}) D^{2*}_{kl}(\Omega^{SD}) D^{2*}_{lm}(\Omega_t^{DI}) X_m - D^{2*}_{0k}(\Omega_t^{LS}) D^{2*}_{kl}(\Omega^{SD}) \left\langle D^{2*}_{lm}(\Omega_t^{DI})\right\rangle X_m$$

This element is the time dependent component $D^{2*}_{lm}(\Omega_t^{DI}) X_m$ of the motion where the offset from the slow motion has been removed. The intensity of this component still depends on the orientation of the global slow motion frame at a given time, but thanks to the isotropy of the solution, every *a priori* orientation of the slow motion has the same probability, which allows us to write the correlation function associated to the intermediary motion as follows:

$$G_{ii}(\tau) = \left\langle V_i(\Theta_0^L) V_i(\Theta_\tau^L)\right\rangle = \sum_{l,l'}\sum_{m,m'}\sum_{n,n'}\sum_{p,p'} \left\langle D^{2*}_{lm}(\Omega_0^{DI}) D^{2}_{l'm'}(\Omega_\tau^{DI})\right\rangle$$

$$\times D^{2*}_{mn}(\Omega^{ID'}) D^{2}_{m'n'}(\Omega^{ID'}) \left\langle D^{2}_{np}(\Omega^{D'I'})\right\rangle \sigma_p \left\langle D^{2*}_{n'p'}(\Omega^{D'I'})\right\rangle \sigma^*_{p'}$$

Assuming the uniaxiality of the local POMT in the alignment frame D', we obtain:

$$G_{ii}(\tau) = S_f^2 \sum_{l,l'}\sum_{m,m'} \left\langle D^{2*}_{lm}(\Omega_0^{DI}) D^{2}_{l'm'}(\Omega_\tau^{DI})\right\rangle D^{2*}_{m0}(\Omega^{ID'}) D^{2}_{m'0}(\Omega^{ID'}) = S_f^2 G_{inter}(\tau)$$

With:

$$S_f^2 = \left|\sum_p D^2_{p0}(\Omega^{I'F}) \left\langle D^2_{0p}(\Omega^{D'I'})\right\rangle\right|^2$$

From the properties of the Wigner rotation matrices again, it follows that $G_{inter}(0)$ is equal to $1/5$. However, this correlation function doesn't decay to zero, but towards an asymptotic value. By noting that the Euler angles Ω_τ^{DI} become independent of their initial condition after a long time, we can write by invoking the uniaxiality of the POMT in the frame D:

$$G_{inter}(\infty) = \left| \sum_m D^2_{m0}(\Omega^{ID'}) \left\langle D^2_{0m}(\Omega^{DI}) \right\rangle \right|^2 = S_i^2$$

Therefore, we can write, by defining the normalized intermediary correlation function $C_{inter} = (G_{inter} - S_i^2)/(1 - S_i^2)$:

$$G_{ii}(\tau) = \frac{1}{5}(1 - S_i^2) S_f^2 C_{inter}(\tau)$$

Finally, using the same principle, we can derive the fast correlation function $G_{ff}(\tau)$ that can be shown to be expressed as:

$$G_{ff}(\tau) = \frac{1}{5}(1 - S_f^2) C_{fast}(\tau)$$

The total correlation function can finally be expressed as:

$$C(\tau) = \frac{1}{5}\left[S_f^2 S_i^2 C_{slow}(\tau) + S_f^2(1 - S_i^2) C_{inter}(\tau) + (1 - S_f^2) C_{fast}(\tau) \right]$$

If we make the further assumption that the three correlation functions can be expressed as single exponentials, we can write the associated spectral density function as a sum of three Lorentzians as follows:

$$J(\omega) = \frac{2}{5}\left[S_f^2 S_i^2 \frac{\tau_{slow}}{1 + \omega^2 \tau_{slow}^2} + S_f^2(1 - S_i^2) \frac{\tau_{inter}}{1 + \omega^2 \tau_{inter}^2} + (1 - S_f^2) \frac{\tau_{fast}}{1 + \omega^2 \tau_{fast}^2} \right]$$

This formula may be written equivalently in terms of amplitudes of each three dynamic modes rather than order parameters. The associated correlation function would be written as follows:

$$C(\tau) = A_1 C_1(\tau) + A_2 C_2(\tau) + A_3 C_3(\tau)$$

The validity of the assumption that the POMT in the alignment frames associated with the two

fast processes is uniaxial can be argued. However, even not perfect, the fitted amplitudes and timescales of each dynamic modes provide a relevant sketch of the dynamics of intrinsically disordered proteins. The extracted order parameters can be seen as effective asymptotic values of each term corresponding to the dynamic mode associated with fast and intermediate motion. This provides information on the level of motional restriction at the timescales associated with the fast and intermediate correlation times. The validity of this model has been confirmed several times on Intrinsically Disordered Proteins [50–52]. In this thesis, the physical meaning of the IDP Model-Free parameters are treated in several chapters. The model-free analysis provided insight into the dynamics of countless systems from folded to intrinsically disordered proteins [50–52, 59–64].

1.3.4 Temperature and Viscosity Dependence of IDP Backbone Dynamics

Intrinsically Disordered Proteins are much more in contact with the solution's solvent than folded proteins where only the surface is accessible. Their dependence on the solvent properties is thus important and we can expect the dynamics of IDP to be highly coupled with the solvent. Abyzov, Salvi and coworkers performed extensive NMR spin relaxation measurements of the disordered C-terminus of the Nucleoprotein of Sendai Virus (SeV Ntail) at different temperatures [51]. Using an approach previously applied to the study of ^{15}N and ^{13}C relaxation of proteins in the solid state [65], they performed a Model-Free analysis and modeled the associated correlation times with an Arrhenius relationship as a function of temperature such that:

$$\tau_k(T) = \tau_{k,\infty} exp\left(-\frac{E_{a,k}}{RT}\right)$$

With $\tau_{k,\infty}$ the correlation time for infinite temperature, $E_{a,k}$ the residue-specific activation energies associated to the given dynamic mode, T the temperature and R the ideal gas constant. The amplitudes of each dynamic modes were also temperature dependent. In addition to making the Model-Free fitting more robust, this analysis allowed them to propose the type of motion that dominates each dynamic mode in the Model-Free analysis. The slowest timescale exhibited a bell-shaped pattern with short-range correlations over the chain, indicating that chain-like motion might dominate the slow dynamic mode. The intermediary dynamic mode was associated with peptide-plane fluctuations since the extracted activation energies corresponded to the expected range for this given type of motion. The fastest dynamic mode highlighted almost neither sequence nor temperature dependence for the correlation time and was associated to libration and vibrations of the N-H bond. This study shows that Model-Free analysis of IDPs provides an unprecedented insight at atomic resolution on the behavior of the protein chain's backbone dynamics. A few years later, Adamski and coworkers studied the effect of viscosity on IDP backbone dynamics by adding viscogens to the solution [52]. They had noted that it was possible to obtain information on the solution nanoviscosity by looking at water relaxation, and proposed to model the correlation times τ_k as a function of the nanoviscosity of the viscous solution η and the nanoviscosity of the reference solution η_0. They defined the solvent friction $\rho = (\eta/\eta_0) - 1$ and expressed the correlation time as:

$$\tau_k(C) = \tau_{k,0}\,(\varepsilon_k \rho(C) + 1)$$

With, for the kth dynamic mode, $\tau_{k,0}$ the correlation time of reference corresponding to no addition of viscogen and ε_k a friction coefficient. C is the concentration of viscogen and $\rho(C)$ is the solvent friction associated with the concentration C of viscogen. This relationship has important implications. It implies that the correlation times associated to the protein's backbone dynamics is linearly coupled with the behavior of the water molecules, with a proportionality coefficient ε_k. They noted that the fast timescale had no viscosity dependence, while the intermediary timescale's ε_2 was around 1, indicating that it was directly coupled with the water's friction. It was also observed that the slow timescale ε_3 was always around 3 to 4 times bigger than ε_2, indicating that the size of the group of atoms associated with the slow motion was bigger than the size for the intermediary motion in agreement with the length-scale dependence of experienced viscosity in complex environments (*vide supra*). This observation supports what has been hypothesized in the previous work on the type of motion associated to each dynamic modes. They also tested two different viscogens: Dextran40 and PEG10000, two long chain polymers, and observed that the coefficient ε_k was the same for both viscogens, indicating that these coefficients might be independent on the type of viscogen used to alter the viscosity of the solution. A few measurements were performed in ovocyte cells and the predictions made with the model reproduced the range of the experimental *in cellulo* data, indicating that the model can work in physiological conditions. Assuming that the temperature and the viscosity-dependence mechanisms are orthogonal to each other, one can model the correlation times extracted from model-free analysis as follows:

$$\tau_k(T) = \tau_{k,0,\infty}\left(\varepsilon_k \rho(C) + 1\right) exp\left(-\frac{E_{a,k}}{RT}\right)$$

With $\tau_{k,0,\infty}$ the correlation time for infinite temperature and zero solvent friction.

1.3.5 Spectral Density Mapping

One useful method to analyse NMR spin relaxation data is Spectral Density Mapping. If one has measured enough relaxation rates on a system, the spectral density function evaluated at specific frequencies can be estimated [66–69]. With ^{15}N relaxation data exclusively, it is possible to have an estimate of $J(0), J(\omega_N)$ and, depending on the method, $J(\omega_H), J(0.870\omega_H)$ or $J(0.858\omega_H)$ [69]. The principle relies on assuming a certain shape of the spectral density function at the high frequency values around $J(\omega_H)$ allowing to estimate the spectral density function at a single effective high frequency value $J(\varepsilon\omega_H)$ with $\varepsilon \in [0, 1]$, instead of all the three $J(\omega_H + \omega_N), J(\omega_H), J(\omega_H - \omega_N)$. This simplification allows the derivation of expressions for the spectral density functions evaluated at the given frequencies as a function of the measured relaxation rates. For example, Farrow and coworkers proposed a method based on the assumption that the spectral density function at high frequencies could be approximated as [69]:

$$J(\omega) = \lambda_1/\omega^2 + \lambda_2$$

Where the λ_x are real constants. This approximation is shown to be most accurate for folded proteins with slow overall motion and fast internal motion with respect to the proton Larmor period. A second requirement is the following relationship:

$$6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) = AJ(\omega_q)$$

If this is valid, this system of equations gives $A = 5$ and $\omega_q = 0.870\omega_H$. With the same principle, we also obtain the following relationships, where the combination of high-frequency terms are simplified to a simpler single term:

$$J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N) = 7J(0.921\omega_H)$$
$$J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H) = 13J(0.955\omega_H)$$

With these simplifications, the three measured relaxation parameters R_1, R_2 and nOe can be expressed as:

$$R_1 = \frac{d^2}{10}\left(3J(\omega_N) + 7J(0.921\omega_H)\right) + \frac{2}{15}c^2J(\omega_N)$$

$$R_2 = \frac{d^2}{20}\left(4J(0) + 3J(\omega_N) + 13J(0.955\omega_H)\right) + \frac{c^2}{45}\left(4J(0) + 3J(\omega_N)\right)$$

$$nOe = 1 + \frac{d^2\gamma_H}{10\gamma_N R_1}5J(0.870\omega_H)$$

With $d = (\mu_0\hbar\gamma_H\gamma_N)/(4\pi r_{NH}^3)$ and $c = \Delta\sigma\omega_N$. Three simplification methods were proposed from this system of equations. The first and simplest method consists of considering the high frequency values of the spectral density function to be constant, thus yielding $J(0.870\omega_H) = J(0.921\omega_H) = J(0.955\omega_H)$. A second method consist of estimating $J(0.921\omega_H)$ and $J(0.955\omega_H)$ from $J(0.870\omega_H)$ by assuming that $J(\omega)$ is proportional to $1/\omega^2$. This corresponds to the fastest rate of decay of the spectral density function assuming it has a Lorentzian form. The required assumptions of these two methods induce a non negligible bias in the calculated $J(\omega_N)$ and $J(0)$. The third method takes advantage nOe data measured at two different magnetic fields. In this case, $J(0.921\omega_H)$ and $J(0.955\omega_H)$ can be extrapolated by doing a first order Taylor expansion around $J(0.870\omega_H)$, which translates into the following equation:

$$J(\varepsilon\omega_H) \simeq J(0.870\omega_H) + (\varepsilon - 0.870)\,\omega_H\frac{d}{d\omega}\left[J(\omega = 0.870\omega_H)\right]$$

$J(0.870\omega_H)$ is calculated as follows:

$$J(0.870\omega_H) = (nOe - 1)\,\frac{2R_1\gamma_N}{d^2\gamma_H}$$

With nOe at two magnetic fields, the derivative of $J(\omega)$ evaluated at $\omega = 0.870\omega_H$ can be estimated as:

$$\frac{d}{d\omega}\left[J(\omega = 0.870\omega_H)\right] = \frac{J(0.87\omega_H^{(1)}) - J(0.87\omega_H^{(2)})}{0.87(\omega_H^{(1)} - \omega_H^{(2)})}$$

Where $\omega_H^{(i)}$ correspond to the proton Larmor frequency at a given field strength. $J(\omega_N)$ can then be calculated as:

$$J(\omega_N) = \frac{R_1 - \frac{7d^2}{10}J(0.921\omega_H)}{\left(\frac{3d^2}{10} + \frac{2c^2}{15}\right)}$$

And finally, $J(0)$ can be calculated as:

$$J(0) = \frac{R_2 - \left(\frac{3d^2}{45} + \frac{3c^2}{45}\right)J(\omega_N) - \frac{13d^2}{20}J(0.955\omega_H)}{\left(\frac{4d^2}{20} + \frac{4c^2}{45}\right)}$$

Alternatively, one could apply the same principle with the exchange-free transverse cross-correlated relaxation rate η_{xy}, giving the following relationship for $J(0)$:

$$J(0) = \frac{\eta_{xy} - \frac{3P_2(cos(\theta))cd}{15}J(\omega_N)}{\left(\frac{4P_2(cos(\theta))cd}{15}\right)}$$

Spectral density mapping is an efficient way of obtaining relevant information on the spectral density function. Figure 1.6 gives an example of spectral density mapping performed on the residue of an intrinsically disordered protein with a spectral density function extracted from MD simulation. The results show a remarkable agreement with the real spectral density function at the evaluated frequencies, especially for $\omega = 0$ and $\omega = \omega_N$. However, sometimes, the replacement of high frequency terms with an optimized effective value induces a systematic bias that can be non negligible when the high frequency motion is significant like in Intrinsically Disordered Proteins. Kadeřávek and coworkers [70] proposed an alternative to the method of Farrow *et al.* where they approximate the spectral density function with a straight line between $J(\omega_H + \omega_N)$ and $J(\omega_H - \omega_N)$. Nothing else is assumed on the shape of the spectral density function. The slope s of this line is calculated as before with several NOE data at different magnetic fields. In this method, only the reduction associated with the NOE data is optimized with an effective frequency of $\omega_q = 0.858\omega_H$ and the bias arising from this reduction for the spectral density function estimations at lower frequencies is compensated with correction terms that are proportional to s and ω_N.

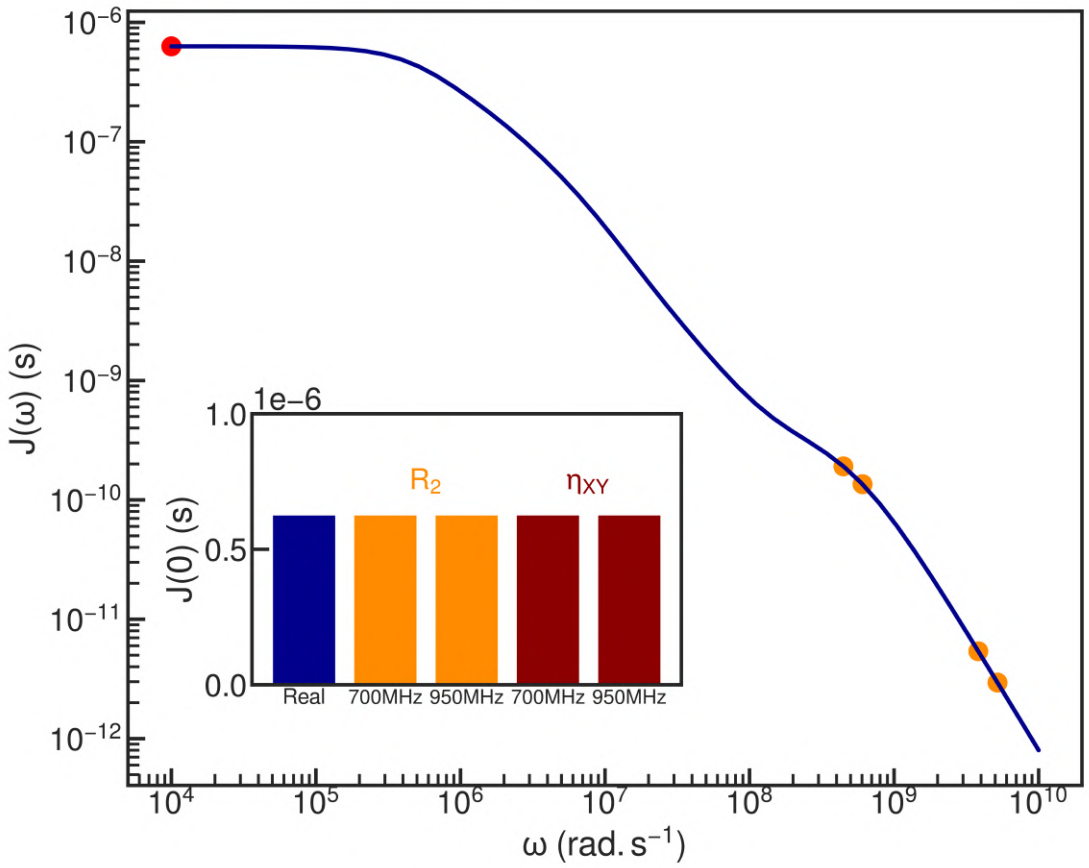


Figure 1.6: Multi-Lorentzian fit of the NH bond vector spectral density function associated with the residue 406 of an ensemble of MeV Ntail MD trajectories simulated with CHARMM36m force field. The markers correspond to the estimated values of the spectral density function from reduced spectral density mapping using method 3 of Farrow *et al.* [69] given the calculated relaxation rates at two magnetic fields (750 MHz and 950 MHz). The bars in the inserted figure are the real and estimated values of J(0) using either R_2 or η_{XY} with the 700 MHz or the 950 MHz data.

1.3.6 Other Relaxation Analysis Techniques and concluding remarks

Among the other techniques that were introduced along the years, a lot are, strictly speaking, "model free" in the sense introduced by Lipari and Szabo. Khan and coworkers proposed in 2015 to model the spectral density function as a sum of Lorentzians akin to the conventional model-free analysis but by fixing the correlation times in a logarithmic scale prior to fitting the associated amplitudes [71]. In this so-called IMPACT ("Interpretation of motions by a projection onto an array of correlation times") analysis, the distribution of correlation time for a given data set is chosen with a statistical analysis of the fitting results for different combinations of number of correlation time and covered time window. In their paper, the combination giving the best Akaike's Information Criterion (AIC) is chosen for the analysis. In their case with the data from the partially disordered chicken Engrailed 2 (146-259) measured at five different magnetic fields, they found that 6 correlation times in a logarithmic scale between 21ps and 21ns was optimal given the statistical analysis. They showed that the results of IMPACT gave similar insight with fewer measurements in a limited set of magnetic fields compared to five. These results provide information on the distribution of correlation times in the protein's motion and thus give information on the shape of the spectral density function. The goodness of fit was comparable with conventional model-free analysis with a small improvement from IMPACT. One must however be careful with the analysis of such result. As correctly stated by the authors, the correlation times in this analysis have no physical meaning , it just gives information on the correlation time distribution of the motion around these values. From this analysis, they showed that the spectral density function associated with disordered domains could exhibit a broad distribution of correlation times compared to a folded domain, yielding a smoother decay of the spectral density function.

Another, still "model-free" method, consists of modeling the correlation times with an inverse-gaussian distribution [72]. This model was introduced for low-frequency analysis of the spectral density function in the context of high-resolution relaxometry. Continuous correlation times were already widely used in other systems like polymers [73], with different distribution models exhibiting a non converging spectral density function around 0 [74]. Inverse-gaussian distribution allow a low-frequency analysis with a converging spectral density function.

A very recent approach developed for solid-states NMR mostly consists of using "dynamic detectors". It relies on calculating so-called sensitivities reporting on relaxation as a function of correlation time. The main claim of this method is to provide accurate estimation of motional timescales associated with the studied protein's residue with respect to the conventional model-free analysis. The interpretation of the results of this method remains challenging due to the severe overlapping of the sensitivities associated to the dynamic detectors that are derived during their analysis. Such method combined with MD-simulation may provide valuable insight.

So far, a significant number of analysis methods were proposed to analyze relaxation rates. From explicit models to "model-free" frameworks involving different distribution of correlation times and shapes of spectral density function. The physical limitations due to the necessary balance between spectral resolution provided by high-fields and low-frequency data provided by low-fields is progressively being overcomed by the rise of high-resolution relaxometry, a principle based on a device proposed by Redfield [75]. Such a device allows a sample to be transferred from a high field position for magnetization buildup and detection to a low field location for relaxation period. This allows the measurement of spin relaxation rates at multiple magnetic fields including low magnetic fields in the range of tens of MHz and lower, provided that accurate corrections are performed to account for the travel of the sample. The emergence of such devices already

started to give insight in more refined details of the spectral densities in numerous spin systems in biomolecules including side chains dynamics with the aid of molecular dynamics simulations for the model choice [76, 77]. They showed that the information of multifield data including low field for this system yielded deviations from a conventional model-free analysis. They proposed a multi-site jump model based on MD simulation and successfully fitted the spectral density function. The continuous rise in computer efficiency will allow the emergence over time of more molecular dynamics simulation studies. MD simulation is a powerful complement to NMR spin relaxation data and the combination of both can provide unprecedented insight. Numerous studies using MD simulation to analyze NMR spin relaxation were performed and will be discussed in the next chapter. Another important aspect concerning the current challenges related to NMR spin relaxation is the understanding of the effect of the environment on the biomolecular dynamics, namely temperature, viscosity and crowding. Protein dynamics is highly dependent on its environment and the future of NMR spin relaxation also needs to provide an understanding of the effect of these factors on protein dynamics. Some of these challenges were addressed previously [51, 52] and some aspects are treated in this thesis.

1.4 Chemical Exchange Processes in NMR

The theoretical basis of NMR for studying protein dynamics at timescales from 10 ps to 10 ns was introduced. These timescales are relatively fast with respect to timescales associated with large conformational changes, ligand binding and other biologically relevant mechanisms. From above tens of nanoseconds to higher timescales, NMR spin relaxation is blind since the relaxation-active processes already quenched the correlation function. We will see that exchange events occurring at timescales from microseconds up to seconds can be observed and characterized using chemical exchange techniques. In this section, the basis of NMR chemical exchange will be introduced, and the most used techniques taking advantage of this process to obtain information on protein dynamics from μs to ms timescales will be presented from CPMG to CEST experiments.

1.4.1 Theory of Chemical Exchange

Let's consider for simplicity a protein that possess two distinct conformational states: A and B. One visible NMR signal arising from one spin of the protein will have the angular frequency ω_A when the protein is in the state A and ω_B when the protein will be in the state B. We can define $\Delta\omega$ the difference between the two angular frequencies. Let's assume now that there is an exchange between the two states, assumed to be associated with Markovian jumps, with associated exchange rates of k_{AB} and k_{BA}. We can write the following equation:

$$A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B$$

We may define the exchange rate k_{ex} such that $k_{ex} = k_{AB} + k_{BA}$. If the exchange is very slow with respect to the angular frequency difference, we would see two peaks in the NMR spectrum corresponding to the two conformations of the protein as if they were not exchanging. On the other hand, if the exchange is extremely fast with respect to the difference in angular frequency between the two states, the exchange would be so fast that we would only observe a single peak at the weighted average angular frequency. The exchange regime is determined by the exchange rate and the difference in angular frequencies between the two states. If the two states are magnetically equivalent, we wouldn't see any chemical exchange processes by NMR. On the other

hand, if the two states exhibit a very big difference in angular frequency, it is likely that we would observe two peaks unless the rate of exchange is significant enough to be bigger than the angular frequency difference. At an intermediate regime where the angular frequency difference is close to the exchange rate k_{ex}, the Fourier-transformed signal is broadened. This results from an increase in the effective transverse relaxation rate due to the dephasing of the transverse magnetization induced by the random exchange between the two states in the sample. Let's take an ensemble of spins associated to our exchanging protein. At an initial condition, all the spins are along the x-axis. During evolution, all the coherences will precess at an angular frequency of either ω_A or ω_B. Without exchange, we will observe two relatively sharp peaks corresponding to the two distinct coherences. If there is exchange between the two states, each spin will randomly undergo exchange between the two states. This will induce a progressive dephasing of the coherence independent from the relaxation-active rotational tumbling, leading to an increased observed R_2, and thus broader peaks. The observed effective R_2^{eff} can be expressed as:

$$R_2^{eff} = R_2 + R_{ex}$$

With R_2 the transverse magnetization corresponding to the signal without exchange and R_{ex} the exchange contribution due to exchange induced coherence dephasing. This behavior can be modeled with the Bloch McConnell equation [78]. Assuming our signal without exchange can be modeled with the Bloch equation, we can extend the equation to include the exchange between the two spins using a kinetic equation as follows:

$$\frac{d}{dt}\begin{pmatrix} I_A(t) \\ I_B(t) \end{pmatrix} = \begin{pmatrix} -k_{AB} & k_{BA} \\ k_{AB} & -k_{BA} \end{pmatrix}\begin{pmatrix} I_A(t) \\ I_B(t) \end{pmatrix}$$

With I_A and I_B the signal intensities associated to the states A and B respectively. Combined with the Bloch equation, the free evolution Bloch-McConnell equation for free evolution would give:

$$\frac{d}{dt}\begin{pmatrix} I_x^A \\ I_x^B \\ I_y^A \\ I_y^B \\ I_z^A \\ I_z^B \end{pmatrix} = \begin{pmatrix} -k_{AB} & k_{BA} & -\omega_A & 0 & 0 & 0 \\ k_{AB} & -k_{BA} & 0 & \omega_B & 0 & 0 \\ \omega_A & 0 & -k_{AB} & k_{BA} & 0 & 0 \\ 0 & \omega_B & k_{AB} & -k_{BA} & 0 & 0 \\ 0 & 0 & 0 & 0 & -k_{AB} & k_{BA} \\ 0 & 0 & 0 & 0 & k_{AB} & -k_{BA} \end{pmatrix}\begin{pmatrix} I_x^A \\ I_x^B \\ I_y^A \\ I_y^B \\ I_z^A \\ I_z^B \end{pmatrix}$$

Relaxation would appear as additional diagonal elements in the matrix but was ignored in this equation for simplicity. This consideration can be easily generalized to a more complex N-states model. Figure 1.7 shows the lineshapes of Fourier-transformed NMR peaks with different exchange rates and different populations for the states A and B as calculated from the Bloch-McConnell equation. At slow exchange, we see the two sharp peaks whose intensity depend on the state populations. Then, at faster exchange rates we observe the previously described line broadening until k_{ex} is approximately equal to the angular frequency difference. In this case, we observe a coalescence effect where the two peaks are almost invisible. At faster k_{ex}, one single broad peak

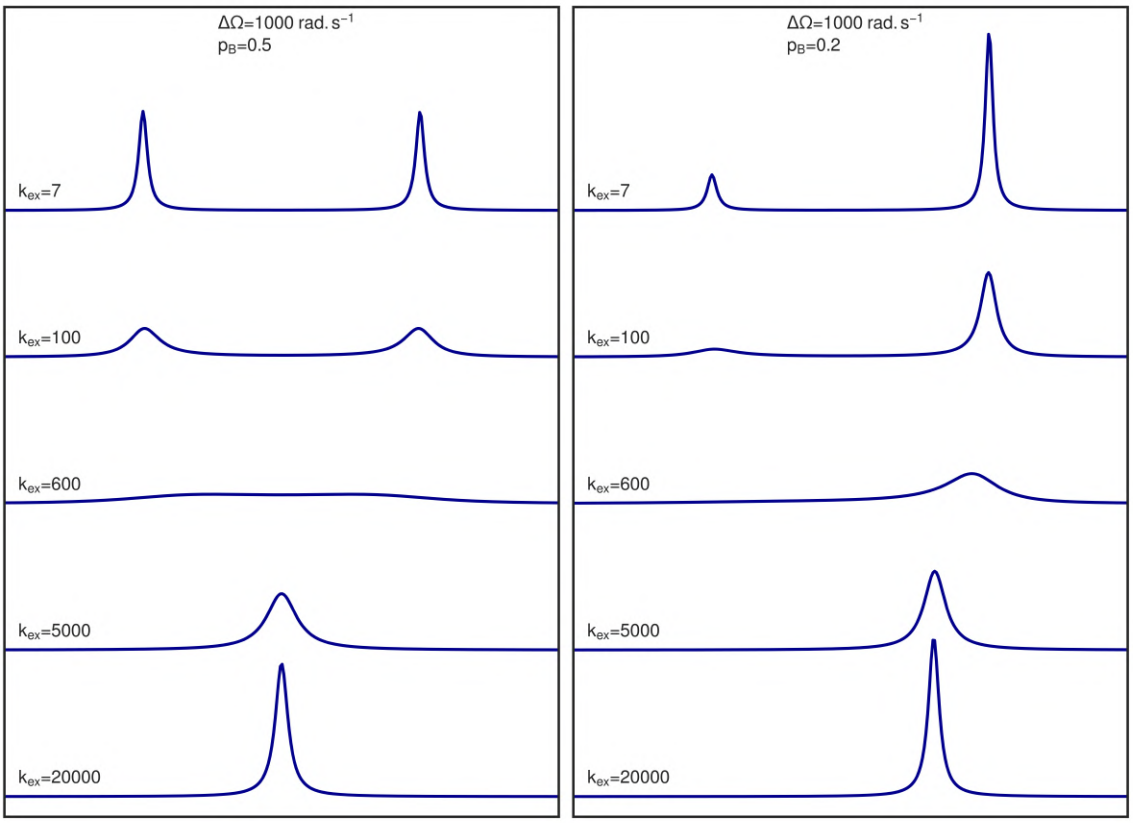


Figure 1.7: Fourier-transformed NMR signal of a single spin system exchanging between two states A and B with equal population (left) or asymmetrical populations (right), with different exchange rates from 7 to 20000 s^{-1} from top to bottom. The FIDs were obtained by numerical computation of the 2-states Bloch-McConnell equations.

is visible. The peak dispersion is the strongest at the coalescence location. Then, at increasing k_{ex}, the peak becomes sharper until the exchange contribution to R_2 becomes negligible when $k_{ex} \gg \Delta\omega$. In cases where exchange broadens our NMR signal, numerous NMR techniques allow us to obtain valuable information on our NMR signal. We will see that it is possible to obtain information on the exchange kinetics, on the chemical shift difference between the different states, on the relative populations, and on the ground R_2 that only depends on the rotational tumbling of the relaxation-active interactions.

Just like for the Bloch equation, in the case of a system involving several spins, J-coupling or any quantum mechanical effect, the Bloch McConnell equation is not sufficient anymore. Instead, the Liouville von-Neumann equation is used combined with an exchange matrix component, yielding the Stochastic Liouville equation [79, 80]. We can model a stochastic jump process between N states with the following equation:

$$\frac{dp(t)}{dt} = Kp(t)$$

Where p is the vector of size N associated with the population of all the states at a given time t, with the population for state $j \in \{1, .., N\}$ at time t being the component p_j of the vector. K is

the $N \times N$ exchange matrix composed by all the exchange rates between each states. the matrix elements are $K_{jk} = k_{jk}$ with k_{jk} the exchange rate between states j and k and $K_{jj} = -\sum_k k_{kj}$. We can write the Liouville von-Neumann equation for a given exchange state j as follows:

$$\frac{d\sigma_j(t)}{dt} = -(i\hat{L}_j + \hat{\Gamma}_j)\sigma_j(t)$$

Where σ_j is the density operator, expressed as a vector whose coordinates are expressed in a convenient basis of our Liouville space of dimension M (ex: Cartesian or Shift basis). $\hat{L}_j$ is the Liouvillian super-operator or commutation super-operator for the state j, and $\hat{\Gamma}_j$ is the relaxation super-operator, both of size $M \times M$ and expressed as a matrix. The Stochastic-Liouville equation is the combination of the markov jump equation with the Liouville equation as follows. We define an operator Λ associated with the Liouville operators for each given states with the following expression:

$$\Lambda = \sum_j^N P_j \otimes -(i\hat{L}_j + \hat{\Gamma}_j)$$

Where P_j is an $N \times N$ matrix that applies a projection into the exchange state j. The symbol $\otimes$ represents a tensor product, resulting in a final matrix of size $MN \times MN$ in a space of dimension $M \times N$. For the exchange, we define the operator Ξ defined as:

$$\Xi = K \otimes E_{spin}$$

Where K is the $N \times N$ exchange matrix and E_{spin} is the $M \times M$ identity matrix associated with the considered spin system in Liouville space. The final Stochastic-Liouville equation is thus, after defining the density operator in the Stochastic-Liouville space as $\tilde{\sigma}(t) = [\sigma_1(t), .., \sigma_N]^T(t)$:

$$\frac{d\tilde{\sigma}(t)}{dt} = (\Lambda + \Xi)\tilde{\sigma}(t)$$

We note that in a case of a single spin system or in a case with no quantum-mechanical complications, the Stochastic-Liouville equation becomes the Bloch-McConnell equation. The latter is usually enough to model most of the chemical exchange problems involved in biomolecular NMR. Although the lineshapes of the observed peaks in NMR spectra already provide helpful information [81, 82], specific techniques allow a quantitative description on the exchange occurring in our system of interest. From relaxation dispersion to saturation transfer experiments, one can characterize the exchange rates, the chemical shift differences and the state populations in a given exchanging system.

1.4.2 CPMG Relaxation Dispersion

The CPMG relaxation dispersion experiment is based on the Carr-Purcell-Meiboom-Gill spin echo pulses technique developped in the 50s [83, 84]. A CPMG sequence consists of a series of repeating spin-echo periods. Let's consider our two-state system described earlier. If we start with our overall magnetization along the x-axis, the spins will precess at a randomly exchanging angular frequency. The exchange induces a dephasing of the overall magnetization and thus line-broadening. During a CPMG sequence, each 180° pulse will invert the sense of precession of the nuclear spins. If the frequency of 180° pulses is close or higher than the rate of exchange, then the dephasing is reduced. An illustration of this mechanism is presented in figure 1.8. For ^{15}N nuclei, the frequency of 180° pulses ν_{CPMG} can usually go up to 1000Hz. ν_{CPMG} is limited by physical constraints, a too high frequency could damage the NMR probe. ν_{CPMG} also determine the maximum timescale for which CPMG is sensitive. If the rate of exchange is much faster than ν_{CPMG}, then the pulses are not applied frequently enough to reduce the exchange-induced dephasing. On the other hand, if the exchange is too slow, too few exchange events will occur during the CPMG period and nothing will be visible. The CPMG period is of course limited by the ground transverse relaxation rate due to rotational tumbling.

A CPMG experiment can be modelled with a Bloch McConnell equation, and analytical expressions for the effective transverse relaxation rate as a function of ν_{CPMG} can be derived from it. Let's consider our transverse overall magnetization $I^+(t) = I_x(t) + iI_y(t)$. The Bloch-McConnell equation considering only this transverse component can be reduced to:

$$\frac{d}{dt}\begin{bmatrix} I_A^+(t) \\ I_B^+(t) \end{bmatrix} = \begin{bmatrix} -i\omega_A - k_{AB} - R_{2A} & k_{BA} \\ k_{AB} & -i\omega_B - k_{BA} - R_{2B} \end{bmatrix}\begin{bmatrix} I_A^+(0) \\ I_B^+(0) \end{bmatrix}$$

This equation can be analytically solved, giving a solution of the form:

$$\begin{bmatrix} I_A^+(t) \\ I_B^+(t) \end{bmatrix} = \begin{bmatrix} a_{AA}(t) & a_{AB}(t) \\ a_{BA}(t) & a_{BB}(t) \end{bmatrix}\begin{bmatrix} I_A^+(0) \\ I_B^+(0) \end{bmatrix} = P(t)\begin{bmatrix} I_A^+(0) \\ I_B^+(0) \end{bmatrix}$$

With $a_{XY}(t)$ propagation functions depending on all the systems parameters. We can easily propagate our system during free evolution with the propagation matrix P. A 180° pulse would have the effect of inverting all the coherences. Propagation following the inversion pulse is then expressed with the complex conjugate of the initial propagation matrix: P^*. Propagating the system in a spin-echo of duration τ_{se} thus translates as:

$$\begin{bmatrix} I_A^+(\tau_{se}) \\ I_B^+(\tau_{se}) \end{bmatrix} = P(\tau_{se}/2)P^*(\tau_{se}/2)\begin{bmatrix} I_A^-(0) \\ I_B^-(0) \end{bmatrix}$$

Finally, the propagation of our system during a CPMG of duration T_{CPMG}, with a CPMG period $\tau_{CPMG} = 1/\nu_{CPMG}$ and an even number of pulses would be written:

$$\begin{bmatrix} I_A^+(\tau_{se}) \\ I_B^+(\tau_{se}) \end{bmatrix} = (P(\tau_{CPMG}/2)P^*(\tau_{CPMG}/2))^n \begin{bmatrix} I_A^+(0) \\ I_B^+(0) \end{bmatrix}$$

Where n is the number of applied 180° pulses. From this, the transverse relaxation rate constant can be expressed as a function of the spin-echo frequency ν_{CPMG}. Such analytical solutions involve relatively sophisticated functions. It is often easier to numerically integrate the Bloch-McConnell equation. R_2^{eff} is equal to $R_2 + R_{ex}$ when $\nu_{CPMG} = 0$, then it decreases until reaching a plateau at higher CPMG frequencies corresponding to the ground transverse relaxation rate R_2. A CPMG relaxation dispersion experiment consists of measuring the transverse relaxation rate with different CPMG frequencies to give a relaxation dispersion profile. Fitted with the appropriated model, it gives valuable information on the kinetics (k_{ex}) and thermodynamics (binding affinity, populations) of the interaction as well as the absolute value of the chemical shift difference between the different states. Usually, one must measure dispersion at several magnetic fields to have a proper fit of the different parameters. For binding studies, one can measure relaxation dispersion at different admixtures of ligand to extract the binding affinity and the populations of free and bound state. Of course, it is possible to fit highly complex exchange models, but the more complex the model, the more data is required [85]. Since the invention of the CPMG sequence, the method has been widely applied and optimized in different biological systems [86–89]. Physical limitations however makes some timescales inaccessible to CPMG dispersion. Other methods exist for chemical exchange and were shown to be either a good alternative, of a powerful complement to CPMG relaxation dispersion.

1.4.3 CEST Experiments

The Chemical Exchange Saturation Transfer experiment has been introduced in the 60s for studying small exchanging systems [90]. It has then been used in biomolecules later on from the 70s [91–93]. CEST has also been extensively applied in Magnetic Resonance Imaging (MRI) experiments as a very large signal enhancement could be achieved from the water signal by a saturation of small metabolites exchanging with water [94–96]. In 2011, Fawzi and coworkers developed a method similar to CEST called Dark-state Exchange Saturation Transfer (DEST) that takes advantage of the strong difference in transverse relaxation between two states of a molecule [97]. They used it to probe exchange between Amyloid-β monomers and protofibrils and characterized for the first time an exchange between a visible monomer and an NMR-invisible fibril state. Similarly, Vallurupalli and coworkers showed in 2012 that a CEST technique included in a multidimensional biomolecular NMR experiment could allow the extraction of exchange parameters as well as chemical shift differences with a great precision even with an invisible low populated protein's excited state. The principle of a CEST experiment is to apply a weak selective radiofrequency (rf) field (b1 field) during a given period of time to a spin system with magnetization aligned with the z-axis. The rf field is selective enough to saturate only spins in a specific narrow chemical shift range. This frequency range depends on the strength of the applied rf field. A stronger b1 field will affect a broader frequency range. Assuming that a spin at a state *A* is undergoing exchange with a state *B*, if exchange occurs during the saturation, the intensity of the signal associated with the *B* state will be altered as well, and *vice versa*. In a CEST sequence, a whole chemical shift range is scanned with such experiment and the intensity of the given signal is plotted against the chemical shift range. A dip at the original signal's chemical shift should be observed, and if exchange occurs at a proper timescale and if the chemical shift difference

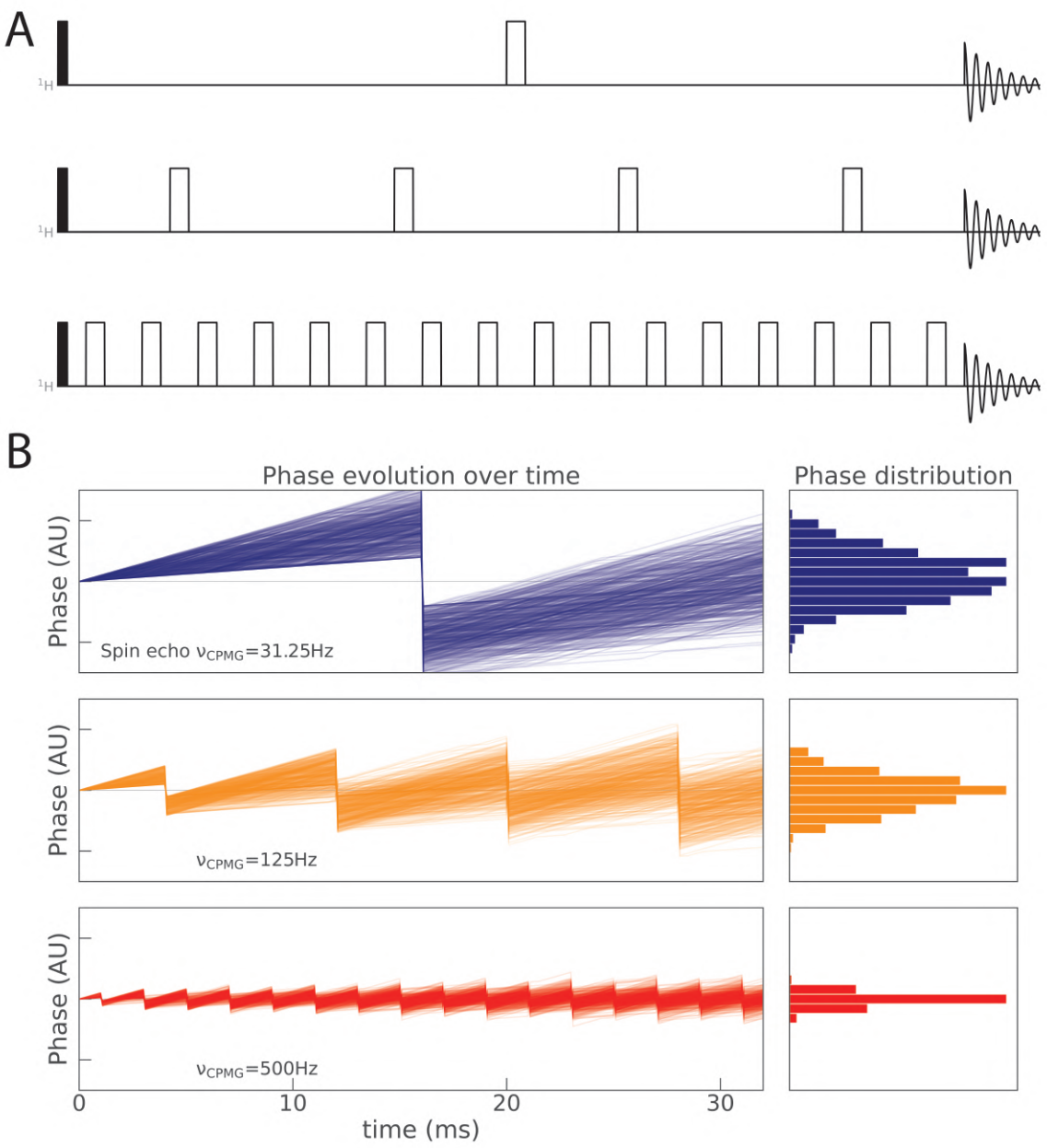


Figure 1.8: A: Basic pulse sequences of a CPMG experiment with one (top), four (middle) and sixteen (bottom) inversions during the CPMG period respectively. B: Illustration of the effect of the corresponding sequences on transverse magnetization. Simulated evolution of the phase of an ensemble of 500 oscillating spins in the transverse plane exchanging between two states in a 32 ms long CPMG experiment containing 1, 4 and 16 inversion pulses respectively from top to bottom. The two exchanging states corresponded to arbitrary angular frequencies of 0.5 and 2 rad/s respectively. One spin had a probability of 2% to switch from one state to the other each 0.1 ms. A higher frequency CPMG train (bottom) will be more efficient at preventing coherence dephasing than a simple spin-echo (top) as illustrated in this panel.

is significant enough, a second dip should be observed at the chemical shift associated with the second state. A basic CEST sequence is illustrated in Figure 1.9. Our spins of interest are along the z-axis. A weak rf field is applied at a given frequency, then the magnetization is transferred to the transverse plane for detection. The intensity associated with a given signal as a function of the frequency of the weak rf-field can be modeled with a Bloch-McConnell equation taking into account the weak b1 field:

$$\frac{d}{dt}\begin{pmatrix} E \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^A - k_{AB} & -\omega_A & \omega_1 & k_{BA} & 0 & 0 \\ 0 & \omega_A & -R_2^A - k_{AB} & 0 & 0 & k_{BA} & 0 \\ R_1^A I_{eq}^A & -\omega_1 & 0 & -R_1^A - k_{AB} & 0 & 0 & k_{BA} \\ 0 & k_{AB} & 0 & 0 & -R_2^B - k_{BA} & -\omega_B & \omega_1 \\ 0 & 0 & k_{AB} & 0 & \omega_B & -R_2^B - k_{BA} & 0 \\ R_1^B I_{eq}^B & 0 & 0 & k_{AB} & -\omega_1 & 0 & -R_1^B - k_{BA} \end{pmatrix} \begin{pmatrix} E \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \end{pmatrix}$$

Where I_i^N $(i \in x, y, z)$ is the i component of the angular momentum for state $N \in A, B$. R_1^N and R_2^N are the associated longitudinal and transverse relaxation rates respectively, k_{AB} and k_{BA} are the rate of exchange between states A and B and B and A respectively and ω_N is the offset corresponding to the signal associated with the protein state N. I_{eq}^N is the equilibrium z-magnetization for state N and E is the identity operator. Finally, ω_1 is the strength of the low-power radiofrequency field.

An advantage of the CEST experiment over the CPMG dispersion is that it provides a precise measure of the chemical shift difference including the sign of this difference. In addition, since the magnetization of interest is along the z-axis during the relaxation period, this period is limited by the longitudinal relaxation rates, which are smaller or equal to the transverse relaxation rates. Therefore, slower exchange processes can be probed with CEST.

Scanning a whole frequency range with a CEST sequence requires a significant amount of points and can be highly time consuming when implemented in a multidimensional experiment for biomolecules. Recently, Yuwen and coworkers proposed a variant of the CEST experiment in which the weak rf-field is replaced by a DANTE selective excitation scheme [98]. This scheme provokes a periodic selective multi-site excitation over the frequency space instead of a selective single-site excitation [99], allowing to sample the whole frequency space with a significantly shorter experiment time. In such experiment, the actual chemical shift associated with a dip can however not be determined with only one frequency window. Usually, at least two DANTE-CEST experiments are performed with two different frequency windows that are not a multiple of each other in order to be able to extrapolate the real chemical shifts associated with the measured dips. This multi-site excitation scheme allows a significant reduction of the spectrometer time used to probe a given exchange process at a wide range of frequencies and is thus a significant time-saving improvement over the conventional CEST experiment.

1.4.4 Other Chemical Exchange Techniques

Other NMR methods are able to probe NMR chemical exchange processes. Among them, on and off-resonance $R_{1\rho}$ dispersion allows to probe exchange processes in a significant range of timescales, from what CPMG can probe to even faster processes. The application of a rotating b_1 field for a system of exchanging states changes the effective exchange regime between the different states, and can quench the dispersion effect of exchange if the b_1 field is strong enough.

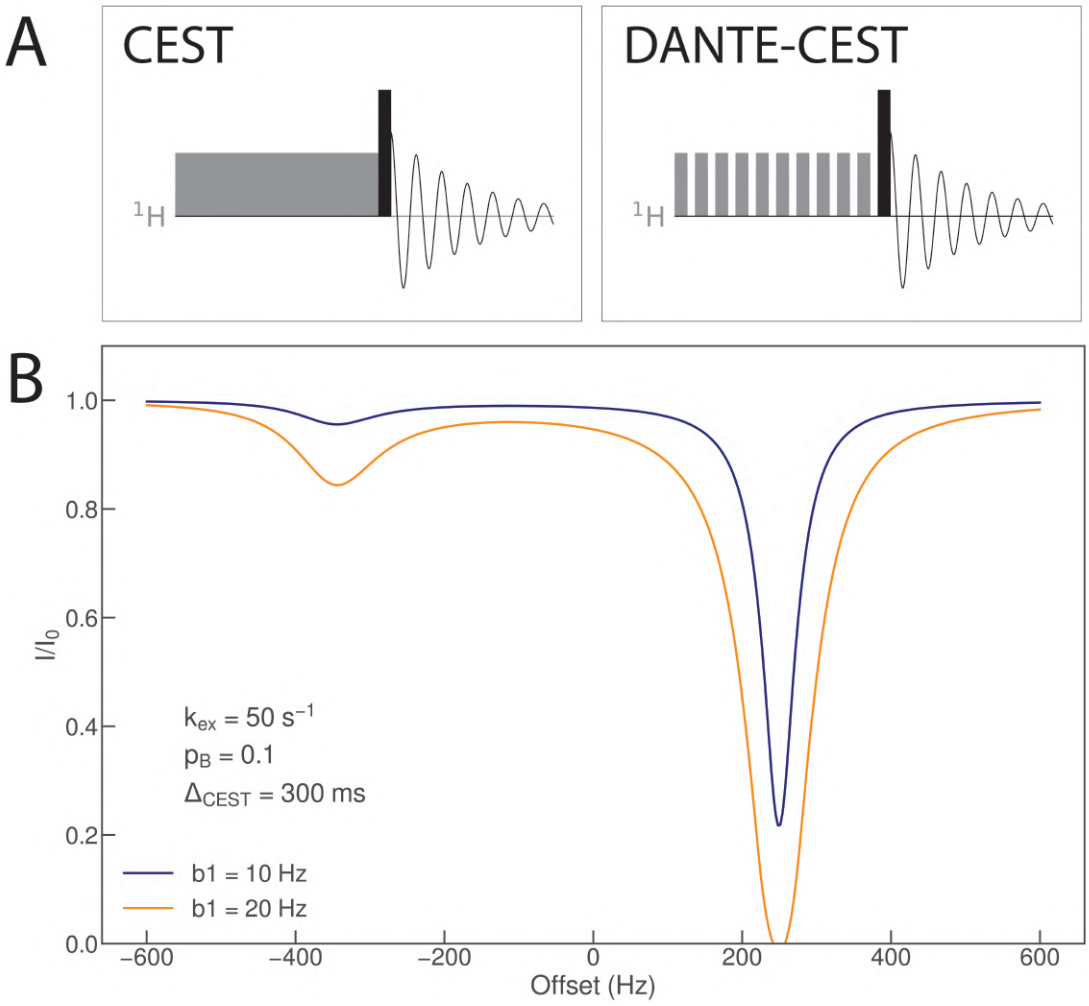


Figure 1.9: A: Basic pulse scheme for CEST (left) and DANTE-CEST (right) experiments. The grey rectangle corresponds to a low-power rf field applied for a duration Δ_{CEST} folowed by a 90° pulse and acquisition. The CPMG and CEST experiments in this thesis were performed with an HSQC encoding similar to the relaxation experiments. B: Simulated CEST profile at two different low-power radiofrequency (rf) fields. The profiles were calculated with a Bloch-McConnell equation.

Several reviews provide an extensive explanation of how $R_{1\rho}$ dispersion work [100, 101]. Additionally, ZZ-exchange [102, 103] and EXSY [104, 105] can be used to probe slower exchange processes. The choice of which technique should be used depends on the dynamic range of the given exchanging system and the exchange regime defined by the ratio between the exchange rate k_{ex} and the chemical shift difference between the different states. All these techniques provide atomic-resolution information on relevant biological mechanisms and offer a direct window towards the function of proteins in biology.

1.5 Conclusion

In this chapter, the NMR principles that are covered in this thesis for the study of protein dynamics were presented. I have shown that 10 picosecond to 10 nanoseconds dynamics could be probed by NMR spin relaxation, providing insight into the rapid dynamics of proteins from fast internal motion up to the timescales of rotational tumbling. I also presented the effect of chemical exchange processes, observable at timescales from microseconds to seconds with different NMR techniques, providing insight into the slow dynamics of large domains of proteins as well as molecular interactions. This chapter shows the unique power of NMR in providing atomic-resolution information on IDP dynamics. Several questions remain unanswered: Is it possible to derive conformational ensembles of IDPs? How can we study the behavior of IDPs in different conditions, from the test tube to more physiological environments? Can computational tools help us understanding protein dynamics? These questions will be addressed in the next chapter.

Chapter 2

Characterization of Intrinsically Disordered Proteins from the test tube to physiological environments

"Deux dangers ne cessent de menacer le monde: l'ordre et le désordre."
- Paul Valéry

Since Intrinsically Disordered Proteins don't have a fixed structure, ensemble methods are necessary to describe their conformational sampling. So far, high-resolution structural techniques such as cryo-EM and crystallography fail to obtain conformations of IDPs, despite recent progress. On the other hand, NMR provides highly valuable information on the ensemble-averaged conformational sampling of IDPs at atomic resolution, and other biophysical methods such as SAXS and fluorescence methods provide complementary low-resolution information. Computational tools like molecular dynamics simulations are powerful complements to experimental methods since they allow the sampling of the conformational space of IDPs at atomic resolution and can be combined with experimental data to derive accurate ensembles of IDPs. In this chapter, we will first describe the useful experimental observables that provide valuable information on Intrinsically Disordered Proteins. Then, we will introduce the computational methods that are used to describe the conformational sampling and the dynamic properties of IDPs. Finally, we will explore recent progress in investigating the properties of IDPs in crowded environments and in liquid-liquid phase separated biomolecular condensates.

2.1 Experimental observables

NMR is a unique tool for characterizing Intrinsically Disordered Proteins at atomic resolution. In this section, the main NMR techniques for conformational and dynamical characterization of IDPs at atomic resolution are presented. In addition, the complementary lower resolution techniques are also introduced to provide a broad view on the field.

2.1.1 NMR Observables

Chemical Shifts

Experimental values of NMR Chemical Shifts are easy to measure and provide insightful information on the averaged local conformational sampling of proteins. They report on the ensemble averaged electronic environment surrounding the considered nuclei, which is directly related to the local conformational space explored by the atom groups in the vicinity of each nucleus. IDPs undergo fast conformational exchange with respect to the chemical shift timescales, with a wide conformational space compared to folded proteins. As a result, the ensemble averaged chemical shift experienced by several different amide protons for example is going to be similar along the sequence of an IDP, unlike for folded proteins. Therefore, it is possible to identify if our protein is an IDP or a folded protein with a single 1D proton NMR spectrum. The amide region of the 1D spectrum of an IDP can exhibit a very narrow chemical shift window between 8 and 8.5 ppm while a folded protein can show amide proton peaks between 6 and 10 ppm (Fig. 2.1). Carbon and Nitrogen chemical shifts exhibit a more disperse chemical shift window due to their stronger amino-acid dependence.

While amide protons provide insight into the disordered or folded nature of our protein, backbone carbon chemical shifts are extremely valuable since they report on the secondary structure propensities of proteins [106, 107]. In practice for IDPs, experimental backbone carbon chemical shifts δ_{exp} are compared to a set of random-coil chemical shifts δ_{rc} determined for each type of amino-acid. in order to obtain the so-called *Chemical Shift Index* (CSI) or *Secondary Chemical Shift*: $\delta_{index} = \delta_{exp} - \delta_{rc}$ [108, 109]. The reference random coil values can be pH, temperature and buffer-dependent and must thus be carefully choosen [110]. CSI values close to zero will indicate a conformational sampling close to random coil, while non-zero values will indicate a propensity for helical or beta sheet structures. Positive and negative C_α CSI values will indicate propensities for alpha helices and beta sheets respectively.

Using chemical shift data, methods were developed to estimate secondary structure propensities. A secondary structure algorithm (SSP) were developed for this purpose by Forman-Kay and coworkers [111]. A set of chemical shift data for random coil, fully formed alpha helices and fully formed beta sheets is used in combination with the measured chemical shifts of several nuclei in the molecule of interest to estimate the chemical shift propensity along the sequence, giving negative values for beta sheets and positive values for alpha helices. Similarly, a "neighbor-corrected structural propensity calculator" (ncSPC) was developed to estimate such propensities, using random-coil values [112, 113].

Residual Dipolar Couplings

RDCs are extremely powerful to provide insight into the structural ensemble of folded proteins and IDPs. The dipolar coupling between two nuclei depends on the orientation of the internuclear vector with respect to the magnetic field and the internuclear distance. Therefore, ob-

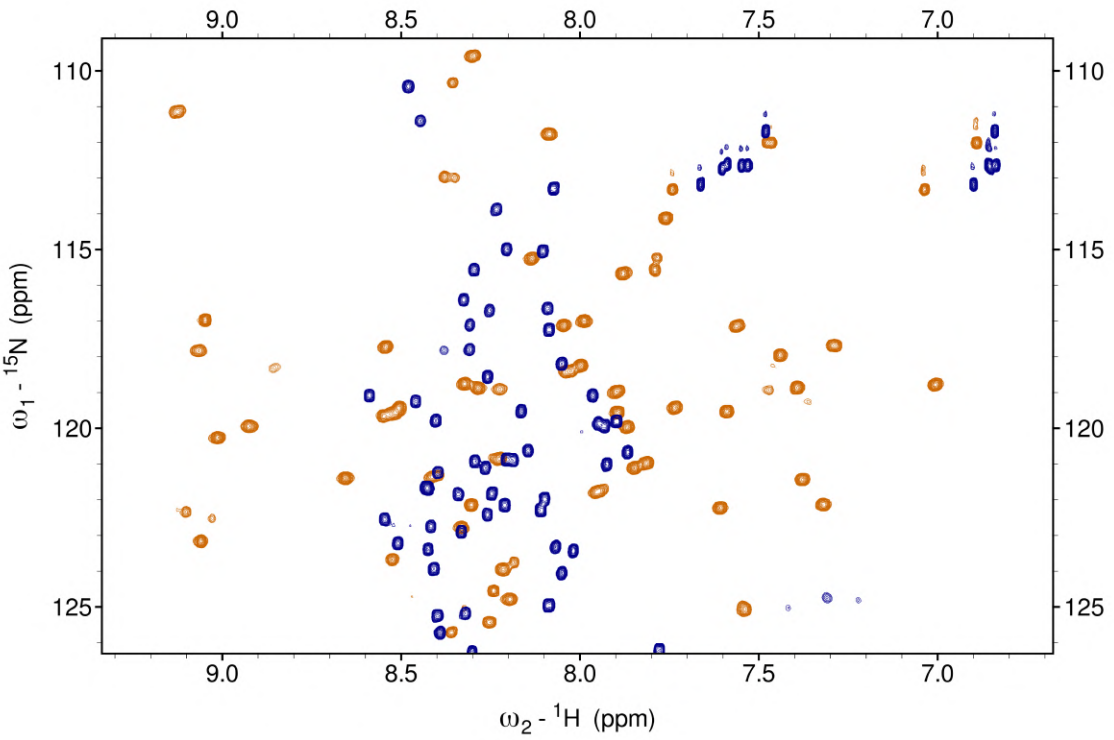


Figure 2.1: HSQC spectra of the Intrinsically Disordered MeV Ntail 465-525 (blue) and the folded XD domain of the Measles Virus (MeV) Phosphoprotein (Orange). IDPs are characterized by a narrow proton chemical shift range compared with folded proteins.

taining several RDC values for all residues along the sequence provide valuable restraints that can be used to derive structural propensities, cooperativity and long-range interactions for both folded proteins and IDPs. If we induce in a sample a weak potential of mean force that induces a slight preference for our protein to align in a particular direction, the ensemble averaged dipolar coupling component of the NMR Hamiltonian doesn't vanish anymore and is expressed as $H_{RDC} = 2\pi D I_z S_z$ in natural units with D the RDC [114]. The RDC Hamiltonian adds up with the scalar coupling Hamiltonian. Measuring RDCs is done by measuring scalar couplings with and without alignment media. The RDC is obtained by subtracting the measured scalar coupling in aligned and isotropic condition. For folded proteins, the analysis approach is straightforward: Defining a common molecular frame and assuming the considered inter-nuclear vectors related to RDC values to be rigid within this frame while defining an alignment tensor to describe the alignment of the molecule along the static magnetic field allows to easily obtain structural restraints [115, 116]. For IDPs, obtaining information is more difficult since each conformation of the IDP possesses its own alignment tensor depending on its shape. In this case, an overall RDC for a given internuclear vector in an IDP is described as an ensemble averaged RDC over all conformations, taking into account the alignment properties of each conformation in an explicit ensemble description [117–120].

RDCs are sensitive to timescales up to tens of milliseconds, which is usually the order of magnitude for chemical shifts. Therefore, RDCs report on the whole sampling of the conformational energy landscape at faster timescales. The ensemble averages obtained from RDCs can provide information on the local dynamics of the protein in terms of so-called order parameters in a similar way to NMR spin relaxation. Comparison with NMR spin relaxation order parameter proved to be useful to study the amount of dynamics occurring between tens of ns (limits of the sensi-

tivity of NMR spin relaxation) to tens of milliseconds, bridging the gap between spin relaxation and chemical shift timescales [121–124].

Paramagnetic Relaxation Enhancement (PRE)

The PRE technique allows the characterization of long-range interactions in proteins. PRE involves the incorporation of a paramagnetic label in the protein. The unpaired electrons in this paramagnetic probe will undergo relaxation-active dipolar interactions with the surrounding nuclei, affecting the measured NMR spin relaxation rate of these nuclei. Since the magnetic moment associated with unpaired electrons is relatively large, it can affect nuclei at distances up to 3.5 nm in space, allowing the detection of long-range contacts between different parts of the protein [125, 126].

PREs are measured by subtracting the measured transverse relaxation rates between the paramagnetically labelled sample and a control sample with no unpaired electrons. Another practice is to look at the intensity ratio between the paramagnetically labelled sample and the control spectra, which is sufficient for qualitative analysis.

Scalar couplings

Scalar couplings, or indirect spin-spin through-bond dipolar interactions between nuclear spins, can provide insightful information on the secondary structure propensities since their value depend on the local electronic structure and molecular conformation. Scalar couplings of backbone nuclei are related to the backbone dihedral angles. Parametrized Karplus relationships can be used to express an empirical expression between dihedral angles and the value of scalar coupling constants [127]. The ^{3}J scalar coupling between amide protons and H_α in a protein backbone are for example related to the dihedral angle Φ, which allows an identification between alpha helical and beta sheet propensities [128, 129]. Backbone scalar couplings provided insightful information on the conformational sampling in IDPs [130–137]. Side-chain scalar coupling can also provide information on side-chain conformations, although obtaining such information is more challenging in IDPs due to important motional averaging [138–143].

NMR spin relaxation and chemical exchange

As presented in the previous chapter, NMR spin relaxation provides unique information on the dynamics of a protein at atomic resolution from 10 ps to 10 ns timescales. In addition to providing information on IDP dynamics, it can be argued that NMR spin relaxation is also giving structural information through relaxation data analysis, for example with the Lipari-Szabo Model-Free order parameter, which provides residue-specific insight into the amount of restriction of internal motion [44]. In addition, NMR spin relaxation allows the detection of protein interactions. Indeed, the presence of an interaction often affects the local dynamics of the IDP and is sometimes responsible for chemical exchange contributions to R_2, which is visible by NMR peak intensity comparison, spin relaxation measurements and chemical exchange techniques. As described earlier, chemical exchange NMR can give insight into conformational changes and interactions occurring from microseconds to seconds timescales, as well as structural information on a bound or excited state, through extraction of chemical shift differences.

2.1.2 Complementary approaches

It is well accepted that the more different kind of experiments are integrated in an ensemble description of IDPs, the more reliable the resulting conformational ensemble ensemble is [144,

145]. Numerous experimental methods provide complementary information to NMR.

Similar to NMR's PRE, **Electron Paramagnetic Resonance** (EPR) spectroscopy also allows the study of long-range distances between two paramagnetic spin labels incorporated in the protein thanks to the double electron-electron resonance (DEER) experiment [146–149]. In such experiment, the sample is flash frozen prior to the measurement and a distance distribution between the two labels is obtained, giving information on long-range contacts.

Small Angle X-ray Scattering (SAXS) gives information on the average size and overall shape of the conformational ensemble of Proteins and IDPs ([150–153]). SAXS can thus for example help deciphering the presence of long-range structures within an IDP ensemble, study complex formations [154] and structural ensembles [155].

Another low-resolution observable that showed to be useful to probe conformational features and changes such as induced folding is **Infrared Spectroscopy**, and more specifically Fourier Transform Infrared Spectroscopy (FTIR). Although limited to folded proteins for a long time, progress in combining it with molecular dynamics simulation allowed the extraction of valuable information in IDP's conformational features [156, 157].

Single molecule fluorescence spectroscopy techniques also proved to be useful for studying IDPs [158]. Single molecule approaches have the advantage of detecting individual molecules, which is a powerful complement to all the previous ensemble average-based approaches. Among them, Single molecule Förster Resonance Energy Transfer (FRET) experiments provides valuable information on IDPs long-range conformation and even their dynamics and interactions [144, 145, 159–162]. In such experiment, two fluorophores, a donor and an acceptor, are attached to two different regions of the protein. The donor is excited and the acceptor's signal resulting from a so-called "Forster Resonance Energy Transfer" between the fluorophores. The overall efficiency of this transfer of course depends on the distance between these two labels. We will see later that single molecule fluorescence methods are also powerful to obtain information on protein dynamics and interactions in complex environments such as in cells or in crowded environments.

Combining all these experimental observable that report on different structural features allows an accurate description of IDPs at different levels. The question now is how to use all these data to derive an ensemble of conformation that is in agreement with experimental data.

2.2 *In silico* exploration of the energy landscape of IDPs

The main difference between structure refinement of folded proteins and ensemble determination of IDPs is that a folded protein is, in the ideal scholar case, unique, while an IDP ensemble is a set of conformations that are representative of the IDP conformational space. It is, by nature, not unique. The key challenges to derive IDP ensembles are first the effective sampling of the IDPs energy landscape, using either MD simulation techniques or other smart tools. Then, one needs to obtain experimental data that provide structural information. Then, one needs to be able to predict these experimental observables based on a given set of structures. Finally, an optimization procedure is needed to find an appropriate combination of IDP conformations that is representative of the conformational sampling of the IDP. First, the basic principles of MD simulation are going to be presented, then, we will briefly discuss the basic principles for deriving conformational ensembles from conformational sampling to refining with experimental data. Fi-

nally, we will see how MD simulation can be used along with NMR spin relaxation data to obtain insight into the dynamic properties of folded and disordered proteins.

2.2.1 Introduction to MD simulation

Computational approaches such as Molecular Dynamics simulations have become a standard tool for investigating molecules and biomolecules in a very broad range of timescales, from picoseconds with quantum mechanical calculations to microseconds and beyond for coarse-grained models. In principle, the amount of detail and the timescale of the mechanisms we are studying determine the type of simulation we are going to carry. In 2023, the paradigm is as follows: For the study of chemical reactions and mechanisms that involve only a few atoms, a quantum mechanical simulation is perfectly suited. For the study of single biomolecules at timescales reaching the microsecond, it becomes preferable to use classical atomistic simulations to improve the computational efficiency at the cost of quantum mechanical details. For studying larger systems like protein assemblies and longer timescales, coarse grained models offer a suitable compromise between accuracy and computational cost. Finally, larger systems describing mesoscopic phenomena can be studied with analytical models, at the cost of resolution and numerous assumptions.

Classical MD simulations rely on solving the classical equations of motion for a model that mimics a real system. Such "*in silico* laboratory" approach allows to study complex physical problems that are difficult to tackle analytically or experimentally, such as the motion of macromolecules in solution or on membranes for example. An advantage of *in silico* approaches is also that the control parameters can easily be monitored, and extreme conditions are easy to tackle and don't require other precaution than the cooling of the processors that run the simulations. In addition, many physical and thermodynamical observables are easy to measure and control, and the system can be calibrated to control a specific set of observables. Among such systems, the microcanonical or "NVE" ensemble is an isolated system in which we fix the number of particles, the volume of the system and the energy that this system contains. In reality, a system is never isolated. Many simulations are carried out in a system in which temperature is controlled with a thermostat, which in practice is a procedure that regularly updates the particles kinetic energies during the simulation to correct and adjust temperature towards the fixed value. The canonical ensemble or "NVT" ensemble applies this principle. Similar other ensembles exist. For example the isothermal-isobaric or "NPT" ensemble controls the pressure of the system instead of the volume. Such control can be useful to stabilize the simulation in complex systems, but in practice, the results from an NPT ensemble will not deviate from the results obtained in an NVT ensemble in similar conditions [163].

Proteins, and especially Intrinsically Disordered Proteins, are a typical example of molecules that are well suited for MD simulation due to the high amount of interactions that need to be taken into account and, for IDPs, the challenges of obtaining a relevant conformational ensemble. In a classical MD simulations, the forces that rule the system's dynamics are expressed in a "force field" that takes into account all the interactions of the system. A typical potential energy function in an atomistic protein force field is given as follows:

$$V(r_1, ..., r_N; s) = \sum_{i}^{bonds} k_i^b (r_i - r_{i,0})^2 + \sum_{i}^{angles} k_i^a (\theta - \theta_0)^2 + \sum_{i}^{dihedrals} k_i^d (1 + cos(n_i \phi_i + \delta_i))$$

$$+ \sum_{i}^{improper} \frac{1}{2} K_\xi (\xi - \xi_0)^2 + \sum_{i,j} \left[\frac{\sigma_{ij}}{r_{ij}}^{12} - \frac{\sigma_{ij}}{r_{ij}}^{6} \right] + \sum_{i,j} \left(\frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \right)$$

In this function, we can identify two types of interactions: The protein bonded interactions and the non-bonded interactions. The first type represents the interaction between two bonded atoms. They are maintained together like two balls linked by a string of coefficient k_i^b and equilibrium distance $r_{i,0}$. The angles and peptide plane improper angles are modeled with string-like forces as well with coefficients k_i^a and K_ξ and equilibrium angles θ_0 and ξ_0 respectively. Then, the dihedral angles are modelled with an empirical relationship that depends on the vicinal atoms that constitute the neighboring groups of the given dihedral angle. This relationship depends mostly on the steric exclusion and electrostatic effects of these vicinal groups. Among the non-bonded interactions, the Van der Waals interactions are often modelled with a Lennard-Jones potential while the Coulomb interactions are modelled with Coulomb's law. All these empirical force-field coefficients are usually calibrated using extensive quantum mechanical calculations [164–166]. From this, it is very easy to see what are the downsides of such approach. The classical MD force fields are a severe approximation of the complex quantum mechanical properties of the interactions that rule molecular dynamics, and since the results can only be as good as the numerical model, the resulting simulations are as accurate as the force field is to the underlying quantum mechanical interactions. In addition to this intrinsic inaccuracy, MD simulation remains nowadays highly computationally expensive, and for IDPs, sampling the whole conformational space requires a large amount of computational power. In addition, the molecular mechanisms that occur at large timescales such as domain motion and protein interactions are challenging to study in terms of reaction kinetics since the timescales that can be simulated are limited by the computational cost.

An additional challenge that arises with IDPs is that many of the modern force fields were primarily designed for folded proteins. Their accuracy with unfolded proteins is often limited. A typical problem of such force fields is the tendency of the IDP to collapse on itself during the simulation, possibly because of the fact that force fields were initially designed to stabilize compact tertiary structures. Fortunately, the development of force fields for IDPs is an active field of research and many force fields were already proposed [166–169]. An ideal force field would however be able to simulate both IDPs and folded proteins with good accuracy in order to be able to simulate multidomain proteins with IDRs and more complex conditions. We will see that despite these drawbacks and challenges, MD simulation remains extensively used in order to study IDPs, and its use can help to sample their massive conformational space.

While atomistic MD simulations provide valuable high-resolution information on proteins including IDPs, it remains nowadays too computationally expensive to use it to model large systems at timescales larger than the microsecond. To tackle this problem, numerous coarse grained models have been developed through the years with a simplified representation of the molecules of interest [170–173]. Different types of coarse grained models exist and have been reviewed [174, 175]. Typically, a coarse grained model for proteins will describe one protein residue with one or several beads that regroup several atoms. In addition to this simplification, the solvent, classically represented with rigid water models [176] can be treated implicitly. To do so, the system can for example be simulated with additional terms to represent the implicit solvent with for example the so-called Langevin equation, the Newton's second law with a time-dependent random force $\vec{\eta}$ and a friction term λ:

$$m\vec{a}(t) = -\lambda\vec{v}(t) + \vec{\eta}(t) + \vec{F}_{ff}(t)$$

Where $\vec{a}$, $\vec{v}$, m are the acceleration, the velocity and the mass of the given particle respectively and $\vec{F}_{ff}$ is the sum of the forces coming from the force field. Among the most used protein coarse grained models, we can cite the MARTINI force field [177–179], used in a wide variety of applications that are not restricted to proteins, including lipids, sugars, membranes and of course proteins. Nowadays, coarse grained models are being developed for more specific tasks, like the study of liquid-liquid phase separation, protein assemblies and many other applications (*vide supra*).

Computational approaches are powerful at describing complex systems. Can this help us to derive accurate ensembles of IDPs? Can this help us understanding complex mechanisms including IDP dynamics? Can we combine these approaches with experimental data to obtain even more accurate pictures of our systems? We will see in the next section how computation is used to describe ensembles of IDPs, before introducing the current state of the art regarding IDPs in more physiological environments.

2.2.2 Sampling the conformational space

Sampling the entire conformational space of an IDP is rather challenging because of the number of possible conformations. The structural properties of Intrinsically Disordered Proteins must be described in terms of a representative set, a so-called "ensemble" that best represent the conformational space explored by the IDP. One straightforward approach to sample IDPs energy landscape is the use of MD simulations, or Monte-Carlo simulations [180]. However, the force field inaccuracies induces biases in the sampling, and an effective sampling of a sufficiently broad area of the conformational space still remains challenging [181].

In 2005, Bernado and coworkers developed an alternative method consisting of generating random-coil conformations of Intrinsically Disordered Proteins [182] with a statistical coil model. This method, called Flexible-Meccano, is an efficient algorithm that samples the amino acid-specific Ramachandran potentials of what would theoretically describe a random coil IDP. Each amino-acid is generated iteratively with random dihedral angles taken from the given Ramachandran distributions. The dihedral angle potentials were obtained from regions of protein X-ray structures exhibiting no secondary structure. In order to avoid steric clashes during the generation, hard spheres were used for each amino-acid. This tool can be used to generate hundreds of thousands of random-coil conformations in a reasonable amount of time with a standard computer. In addition to random-coil IDPs, this tool was also designed to allow the addition of secondary structure propensities, provided that it only affects the Ramachandran space sampled in these regions. Finally, long-range contacts can be modelled by adding some long-range distance constraints to a certain percentage of generated structures. From such an ensemble, it is possible then to predict resulting RDCs, chemical shifts, PREs, scalar couplings and even SAXS data. This makes this tool highly useful for comparing experimental data to a random-coil behavior, analyzing random-coil deviations in an IDP, and more importantly sample the conformational space of a given protein to obtain an ensemble that can be used for molecular modeling including molecular dynamics simulations. We will see that Flexible-Meccano is also the starting point for more advanced "sample and select" tools that allows the generation of any ensemble matching experimental data.

2.2.3 Deriving conformational ensembles

One should note that deriving a finite ensemble to describe a continuous conformational sampling does not yield a unique solution, it is in this sense "ill-defined" in many cases unless there is a finite ensemble of structures such as a folded protein exchanging between several well defined conformations. The derivation of conformational ensembles of Intrinsically Disordered Proteins is still an active field of study [183, 184]. Nowadays, ensemble determination methods are sometimes divided into two categories: maximum entropy methods and maximum parsimony methods [183–186]. Maximum entropy methods rely on using the highest number of possible conformations that together match the experimental data and such that it minimally perturbs the initial ensemble. Maximum parsimony methods on the other hands are defined as methods based on determining the minimum number of conformations that match the experimental data. Since the conformational space of an IDP is broad and continuous, it is a high entropy system compared with folded proteins with single of few possible structures. It is therefore argued that maximum entropy methods are more suited for IDP's ensembles to sufficiently sample the whole conformational space. Numerous ensemble determination methods exist and are presented elsewhere [183, 187].

Using MD simulation to derive an experimental data based ensemble of conformations is possible by refining the conformational sampling to be in agreement with the experimental data. Methods that bias the conformational sampling includes so-called restrained molecular dynamics simulations or restrained replica-exchange MD (REMD). The force field is modified to bias the conformational sampling towards an experimentally compatible sampling. These methods remain limited by the fact that a full sampling of the energy landscape is challenging for IDPs and that force fields are inaccurate by nature, although relatively efficient sampling of the conformational space can in principle be done with enhanced sampling methods such as metadynamics [188].

Instead of biasing the sampling, a reweighting procedure can be performed. An efficient way of deriving an ensemble is to select a set of IDP conformations from a pool of generated structures such that it matches experimental data. Numerous ensemble algorithms were developed with such principle like ENSEMBLE [189, 190], ASTEROIDS [120, 126, 191] and others [192–194]. ASTEROIDS (Selection Tool for Ensemble Representations Of Intrinsically Disordered States) for example relies on a genetic algorithm that selects several ensembles from a large set of conformations generated with *Flexible-Meccano*. The genetic algorithm finds an optimal combination of conformations such that the derived ensembles match experimental data. This of course requires an extensive sampling of the conformational space. The experimental datasets that are used by ASTEROIDS to optimize the ensembles typically are diverse and complementary, with local and long range structure from RDCs, local comformational sampling from chemical shifts, long range contacts from PREs, global shape distribution with SAXS and more recently local and long-range contacts complementary to PREs with single-molecule FRET data [145]. ASTEROIDS is designed to generate a fixed number, usually very high, of conformations in the wanted ensemble. This possibility to generate a significant number of conformations allows a significant sampling of the conformational space of the IDP, therefore mechanically yielding a high entropy solution and fewer biases.

2.2.4 Combining NMR spin relaxation with MD simulation to understand IDP dynamics

Since NMR spin relaxation data gives only information on the spectral density function at specific frequencies, analyzing the data to extrapolate information on protein's dynamics requires important simplifications to overcome the intrinsic complexity of a protein's motion, as explained in the previous chapter. With MD simulation, this issue is partially cancelled since MD simulation provides a direct picture of the proteins dynamics, provided that the simulations are accurate and that the calculated relaxation is in agreement with experiments. From an MD simulation trajectory or an ensemble of trajectories, one can calculate the isotropic rotational time autocorrelation function $C(\tau)$ associated with the relaxation-active interaction. For amide nitrogen in protein backbones, this vector is simply the N-H bond vector. One just has to calculate the following expression along the trajectory:

$$C(\tau) = \left\langle Y_2^0(\Theta_0) Y_2^0(\Theta_t) \right\rangle$$

Where Y_2^0 is the modified second order spherical harmonics presented in chapter 1 evaluated at the angle Θ_t at time t representing the angle between the NH bond vector and the magnetic field.

In 1981, Levy and coworkers performed MD simulations and computed ^{13}C relaxation on amino-acid side-chains [195]. They used these simulations to check the validity of numerous models for the proteins internal motion. Following this, numerous studies used MD simulation with experimental NMR spin relaxation rates to obtain insight into the internal motions of proteins [196–198]. For instance, Bremi and coworkers proposed an MD-based protocol where they analyze relaxation to successfully describe protein's internal dynamics with a motional model. The protocol is based on analyzing MD trajectories and establishing the model before calculating the resulting relaxation from the autocorrelation function and compare with experimental data. They used it to describe a polypeptide's side chains internal motion in terms of dihedral angle motion in a harmonic potential combined with infrequent jumps between different rotamers [199]. They described this as a Gaussian Axis Fluctuation (GAF) and jump model, based on a previously derived GAF model [200]. This protocol was then successfully applied to the study of internal backbone dynamics of a cyclic decapeptide [201] as well as anisotropic peptide plane motion in ubiquitin [202].

Later on, Prompers and Brüschweiler proposed a new analysis method called reorientational eigenmode dynamics (iRED) where a principal component analysis of the covariance matrix associated with the relaxation-active spin interactions is performed from MD trajectories [203–205]. In practice, this isotropically averaged covariance matrix is diagonalized, resulting in several reorientational eigenmodes with associated amplitudes. Each eigenmode describes a correlated modulation of different relaxation-active spin interactions, with an eigenvalue that represents the variance associated with the fluctuation amplitude in the trajectory associated with this eigenmode. Each eigenmode is also associated with a variable correlation time that is simultaneously optimized to reproduce experimental relaxation data. They applied this method to the study of Ubiquitin's native state and partially folded A-state. In the native state, they found that internal and overall motion are separable given the obtained eigenmode distribution, whereas the A-state exhibits poor separability with no evident gaps between internal and overall eigenmodes.

In 2011, Xue and Skrynnikov performed high-temperature rescaled MD simulation of denatured Ubiquitin in implicit solvents and studied the resulting NMR paramagnetic relaxation enhancements and ^{15}N spin relaxation [206]. Examining the time autocorrelation function of the vector associated with the relaxation-active interactions, they identify three exponential components with correlation times of 44 ps, 1.4 ns and 9.4 ns respectively. They attributed the first correlation time to fast internal motion involving out of plane deviations of the N-H bond vector from the peptide plane and rapid vibration, librations and fluctuations involving the dihedral angles. The second was attributed to local rearrangements while the slowest correlation time was attributed to larger scale conformational changes and tumbling. They observed that their results were consistent with the Lipari-Szabo model-free analysis. It was seen in a later article that a three mode model-free analysis in Intrinsically Disordered Proteins exhibit correlation times ranging around these 3 values [51].

In addition to understanding protein dynamics, MD simulation combined with NMR spin relaxation is useful for MD force field development. Showalter and Brüschweiler proposed to use NMR Spin relaxation to validate MD simulation trajectories [207], since NMR spin relaxation provides important information on the protein's motion at fast timescale. This approach allows the correction of backbone dynamics in flexible loops regions for instance. The use of NMR spin relaxation is particularly useful for IDPs since a lot of their behavior from their dynamics to their long range contacts can be reflected in NMR spin relaxation. Still nowadays NMR spin relaxation parameters are used to validate MD force fields. For instance, Yu and coworkers used relaxation order parameters extracted from MD using the iRED method among other NMR parameters to develop the ff99SBnmr2 force field for accurate simulation of both folded and disordered proteins [168].

To tackle the inaccuracy of protein force fields, one method is to re-weight an ensemble of trajectories so that it matches the experimental data. Salvi and coworkers developed ABSURD (Average Block Selection Using Relaxation Data), a genetic algorithm that, from a pool of MD trajectory segments of a given IDP, selects the best combination of trajectories so that the relaxation data from the chosen ensemble matches the reference experimental relaxation data [208]. Remarkably, matching a given relaxation rate using this method also improves the agreement of the other rates with respect to experimental data. In practice, the ABSURD procedure works as follows: First, several MD simulations of our IDP is performed, with a length typically of several hundreds of nanoseconds. Then, each trajectory is segmented into small trajectories of 100 ns. The time auto-correlation function associated with the NH bond vector and then the theoretical field-dependent NMR spin relaxation rates are then calculated for each segment. Finally, the ABSURD algorithm selects the best combination of rates to match experiment, using one of the measured relaxation rate as a reference. For each step, ABSURD calculates the average relaxation rates for a given combination and compares it with the experimental rates with the objective of minimizing the overall sum of squares expressed as follow:

$$RSS(R_k, f) = \sum_n \left(R_{k,n}^{exp}(f) - \sum_i w_i R_{k,n}^{calc,i}(f) \right)^2$$

where the experimental rates are summed over the n residues and the i trajectories included in the combination. The calculation is done for the reference rate R_k evaluated at the magnetic field f. Typically, the optimal reference to use depends on the force field and the protein, but

one usually takes the less well reproduced rate. In the original paper introducing ABSURD, using the force field AMBER ff99SB-ILDN [167] on SeV Ntail, the best optimization was done using R_2 evaluated at 850 MHz. Using this method, one can obtain exceptional reproduction of spin relaxation dates with respect to experimental data, allowing a more detailed analysis of the trajectories. Although this method provides well-behaved models of the autocorrelation functions of NH bond vectors, it does not inherently provide more insight into the component dynamics. In order to extract this information, and inspired by earlier work from Bremi *et al.* [199], Salvi and coworkers then used this method to propose an explicit model to describe relaxation-active motions in IDPs [209]. They proposed to separate the NH rotational correlation function into three components assuming that they are statistically independent. They first identify the correlation function associated with fast bond and angles fluctuations of the NH bond vector with respect to the peptide plane $C_f(t)$. Then, they express the peptide plane fluctuations in terms of dihedral angles fluctuations using a vesion of the GAF model [201], represented by the correlation function $C_{\phi\psi}(t)$. Finally, they identify the correlation function of the C_α-C_α bond vector $C_{tumb}(t)$. This gives the following correlation function:

$$C(t) = C_f(t)C_{\phi\psi}(t)C_{tumb}(t)$$

The advantage of this model is that it allows a separate analysis of the different types of motions in IDPs. Segmental motion for example is directly encoded in $C_{tumb}(t)$ while the effect of dihedral angles fluctuations is directly in $C_{\phi\psi}(t)$. Remarkably, the model correlation function reproduces the original ABSURD-optimized ensemble of trajectories simulated with CHARMM36m [166] correlation function relatively well. They found that $C_f(t)$ converges rapidly after a few ps and can be replaced by the corresponding order parameter S_f^2 while the $C_{\phi\psi}(t)$ and $C_{tumb}(t)$ terms, analyzed with a Tikhonov regularization, exhibited multi-exponential behavior for both correlation functions with high order terms occurring at similar timescales around 10s of nanoseconds. This approach allow them to identify correlated segmental motions and the different contributions to NMR spin relaxation. With this method, they studied the importance of the solvent on the description of IDP dynamics and found, from multi-temperature MD simulations of SeV Ntail with different water models, that the water model is key for an accurate description of IDPs dynamics by MD simulation [210]. They found that TIP4P/2005 [211] is so far (in 2019) the best water model along with the force field CHARMM36m for a relatively accurate description of IDP's dynamics, and more specifically the intrasegmental motion and pisosecond to nanosecond dynamics of IDPs in MD simulations.

Besides backbone dynamics in IDPs, MD simulation was very recently used in combination with NMR high resolution relaxometry and high field spin relaxation measurement experiments to describe methyl-bearing side chain motion. The method lies on the use of MD simulation to derive a relevant explicit motional model with unknown parameters, that can then be fitted to experimental relaxation data [212]. Cousin and coworkers used such method with the ff99SBnmr1-ILDN force field and TIP3P water model [213] to describe the dynamics of isoleucine side chains with a multi-site rotameric jump model [76], but some residues exhibited inaccurate model fitting due to force field inaccuracies. Later on, Bolik-Coulon and coworkers, with the same system, performed MD simulation with a modified Amber ff99SB*-ILDN force field [214] and TIP4P-2005 [211] along with DFT calculations (Density Functional Theory) to obtain precise isoleucine side-chain rotamer conformations in water [77]. This improvement allowed them to obtain an accurate model that describes the isoleucine side-chain's motion and determine the population of each rotamers and their associated exchange rate. In addition, this description gave clue to the

link between order parameter and conformational entropy, here linked to the rotamer states distribution.

Since the rise of MD simulation, a lot of progress has been made, from describing internal motion in polypeptide and folded proteins. Recent advances provided insight into more challenging systems such as the backbone motion of IDPs and methyl side chains. These examples show how useful MD simulation can be to describe complicated mechanisms in IDPs and proteins in general. The latter examples also show how important it is to have accurate MD simulation force fields to obtain an accurate model of protein's motion. Despite a lot of effort in force field and water model developments for Disordered Proteins, optimizing MD simulation for IDPs remains an active field of study, and it can be expected that further improvements in the future in combination with NMR spin relaxation and relaxometry measurements will improve our understanding of IDP dynamics. Besides improving force field for conventional simulations, a growing interest for physiological environments in addition with a constantly increasing computational power invites the force-field developers and computational biologists to look into larger systems, since biologically relevant mechanisms occur in crowded and complex environments. Numerous challenges remain in this perspective since the computational power available remains relatively limited, which forces scientists to look into Coarse-Grained solutions where extracting extremely sensitive atomic-resolution information such as NMR spin relaxation is almost impossible.

2.3 IDPs with macromolecular crowding: Towards cellular environments

Cellular media are extremely complex for, their properties can also vary depending on the location in eukaryotic cells (cytosol, nucleus, mitochondria, etc.). The cytoplasm for example is composed of hundreds of millimolars of inorganic ions including K^+, Mg^{2+}, Ca^{2+} and Cl^-, but also hundreds of millimolars of metabolites including peptides, ATP and sugars. Besides these small species, cytoplasm has a significant volumic percentage of large macromolecules ranging from 10 to 40% of the total volume. The volume of *E.coli* cytoplasms for instance contains 25% of proteins in which 10% of it are cytoskeletal filaments and 90% are soluble proteins [215, 216]. In mammalian cells, protein concentration can go up to 250 g/L. Along with proteins, the cytoplasm also contains other macromolecules including RNA, DNA, lipids and glycans that can also reach high concentrations [217]. It is clear that understanding the modulation of each parameter on IDP structural ensembles, dynamics and interactions is a massive task.

Macromolecular crowding can be described as the presence of soluble macromolecules in a solution such that they represent a significant proportion of the total weight of the solution. From a pure crowding point of view, e.g. assuming that the surrounding molecules can be modelled by inert hard ellipsoids and/or chains exerting steric repulsion forces on other molecules, one can predict that the change in free energy in the system will favor conformations that minimize the occupied space of a given specie. For IDPs, this translates as a preference towards more compact conformations [218, 219]. However, the presence of soft inter-macromolecular interactions can as a consequence favor other conformations, depending on the environment and the type of interaction [220–222]. The change in the conformational sampling of an IDP in crowded environments with respect to a dilute buffer solution will therefore depend on the protein itself and the nature of the crowder [223]. Balu and coworkers recently studied by small-angle scattering methods the conformational sampling of Rec1-resilin, an IDP, as a function of several crowding agents [224]. Their results exhibit crowder-specific conformational changes involving either extension or compaction of the protein upon crowding, confirming the previous theory. We will see

here that besides possible modulation of the global conformational sampling, macromolecular crowding and cellular environments in general can affect viscosity and translational diffusion, rotational diffusion and internal protein dynamics as well as protein-protein interactions.

2.3.1 Complex macromolecular crowding slows down translational diffusion in a length-scale dependent manner

First of all, the presence of a great number of large macromolecules in a solution is obviously increasing the viscosity and the translational diffusion properties of the molecules in the solution. While attempting to demonstrate the existence of atoms, Einstein derived in his doctoral thesis a first order approximation of the effective viscosity η of a solution comprising inert spherical particles in suspension as a function of the volume Φ occupied by these particles, assumed very small:

$$\eta = \eta_0(1 + \frac{5}{2}\Phi)$$

Where η_0 is the viscosity of the solution without any spherical particle in suspension. In a case of high concentration of these molecules (e.g. crowding condition), a non-linear model can be derived (Mendoza2009). Applied to inert sphere of hydrodynamic radius we may express the Stoke's relation expressed as:

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

Where k_B is Boltzmann's constant, T is temperature and D is the isotropic diffusion constant. Since these expressions assumed a perfect spherical particle evolving in an homogeneous fluid, it is clear that deviations are to be expected in more complex environments involving multiple species of different size and shape such as polymer solutions and cellular environments. The understanding of the viscosity and diffusion properties of species in such complex environments is in fact an active field of study [225–227].

One important aspect of the diffusion property of macromolecules in a complex environment, or more specifically their effective viscosity as known as the viscosity experienced by these molecules, is that it is dependent on the size of these macromolecules with respect to the size of the crowding molecules and/or polymers that constitute the environment [228–233]. Typically, molecules with a smaller radius of gyration r_g than monomers of radius of gyration R_g constituting a polymer solution, to take a relatively simple example, will experience deviations from the Stokes-Einstein equation and a length-scale dependent viscosity. In fact, it was shown that when $r_g \ll R_g$, the experienced viscosity approaches the viscosity of the pure solvent (no polymer), while when $r_g \gg R_g$ the experienced viscosity approaches the measured macroscopic viscosity. Holyst and coworkers for instance derived a model to describe the effective viscosity as a function of the size of the studied macromolecule and the hydrodynamic radius of the polymer [230, 231, 234]. From this, a distinction must be made between nanoviscosity experienced by small molecules at the nanoscale and macroviscosity, as known as the macroscopic friction properties of the solution experienced by relatively large probes ($r_g \gg R_g$). In an NMR study on viscosity-dependent

protein conformational exchange kinetics, Sekhar and coworkers pointed out this fact and reminded the importance of using small probes for obtaining reliable solvent-friction values [235]. Later on, Adamski and coworkers successfully used water molecules as a probe for solvent friction in polymer solutions by using the rotational correlation time extracted from longitudinal relaxation rates of water protons [52].

The diffusion properties of macromolecules in complex environments, especially in cellular environments, are very difficult to predict. So far, the consensus in the literature is that a significant decrease in the diffusivity of macromolecules in physiological environment is observed [236–238]. Interestingly, it was recently found by König and coworkers that intrinsically disordered proteins have a higher diffusivity than folded proteins of similar hydrodynamic radius in crowded cellular environments [238]. The same study found that the use of high molecular weight polymer crowders better reproduced the protein's diffusivity in cells than low molecular weight polymers. The development and improvement of experimental methods for probing molecular diffusion *in cellulo* will *in fine* lead to a better understanding of the diffusion properties of macromolecules in cells in and between the different cellular compartments [237–240].

2.3.2 Macromolecular crowding and physiological environments slow down internal protein dynamics in a length-scale dependent manner

Macromolecular crowding also affects rotational diffusion and internal protein dynamics. The same study of König *et al* also probed rotational diffusion in crowded cells using nsFCS and extracted relative chain reconfiguration times in crowded cells [238]. It showed a two-fold slow down in crowded cells. In comparison, translational diffusion time was seven times smaller in crowded cells with respect to *in vitro*. This result was interpreted to be a manifestation of the length-scale dependence of diffusivity. Again, larger objects between the nanoscale and the scale of the radius of gyration of the crowder will be subject to a higher effective viscosity, therefore exhibiting a higher translational and rotational correlation time than smaller objects. Here, the length-scale of the chain dynamics probed by nsFCS is smaller than the length-scale of the full protein, and therefore it experiences a smaller viscosity dependence compared with the full protein. This manifestation was also probed by NMR spin relaxation measurements in crowded environments. Adamski and coworkers studied the fast backbone dynamics of SeV Ntail in dextran solutions and found that the three extracted rotational correlation times exhibited different behaviors with respect to solvent friction [52]. The slowest rotational correlation time, associated with chain-like segmental motion, exhibited a higher dependence in solvent friction than the intermediary dynamic mode, attributed to be associated with peptide plane fluctuations. From the perspective of the scale-dependence of viscosity, this study provided an additional clue on the origins of the extracted dynamic modes fom NMR spin relaxation and confirmed the previous theory on the attribution of these dynamic modes [51].

The effects of macromolecular crowding on the relative translational and rotational correlation times of macromolecules are in fact highly related to viscosity. As long as solvent friction is known, these viscosity effects can be well understood and characterized [52, 235]. However, little else is known on the effect of the cellular environment on the relative dynamic properties of proteins, especially regarding exclusion effects and soft interactions, since they are expected to depend on the environment.

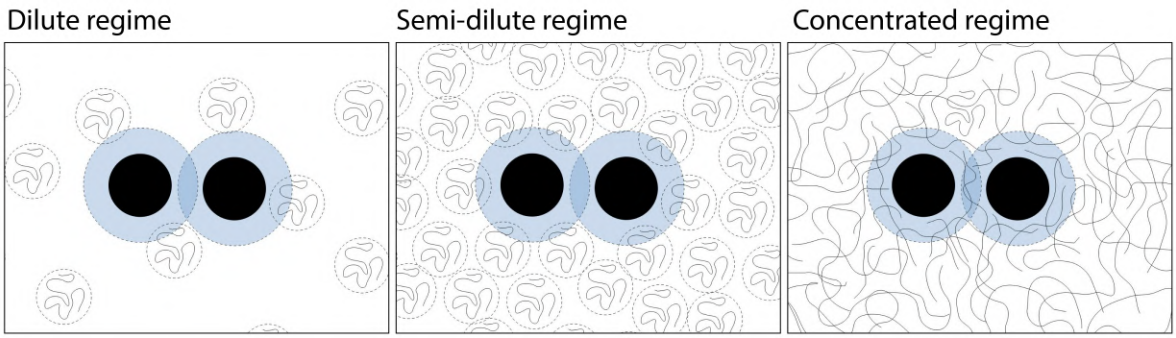


Figure 2.2: Schematic representation of depletion interactions between two spherical macromolecules in a polymer crowder solution in the dilute regime (left), the semidilute regime (center) and the concentrated regime (right).

2.3.3 Macromolecular crowding and physiological environments affect protein interactions

Because macromolecular crowding implies, in addition to an increase in viscosity and exclusion volume, the presence of soft interactions that depend on the properties of the complex milieu and the proteins of interest, the affinity and kinetic properties of protein interactions are extremely difficult to predict. Numerous binding studies have been performed in folded proteins and yielded qualitative descriptions of the effect of macromolecular crowding [241–244]. Kozer and coworkers studied the kinetic and thermodynamic properties of the interaction of TEM1-β-lactamase with the β-lactamase inhibitor protein. They observe that the evolution of the association rate k_{on} is not monotonic with respect to the crowder concentration, with an increase first followed by a decrease in k_{on} with concentration. They found that the effect of crowding on the interaction of these two species could be explained as a function of the crowding regime. Indeed, a polymer crowder can be described as being in a dilute, semi-dilute or concentrated regime depending on its concentration [228, 245]. In the dilute regime, the polymers are described as coils delimited by solvated spheres and a size-dependent radius of gyration R_g. Upon increasing the concentration of crowder, the spheres of each polymer coil cannot be considered separated and they start to overlap each-other in the so-called semi-dilute regime. In a system comprising particles in suspension with polymer crowding in the dilute and semi-dilute regime, a mechanism called depletion interaction occurs and exert an attractive force to two particles close to each other [246–248]. To describe this mechanism, let's consider a spherical particle in such crowded environment. Each of these particles have a so called depletion layer in which the center of mass of the crowder cannot enter. When two particles are close enough to each other, the depletion layer of the two particles overlap. In this overlap, no crowder can enter, and collisions with the crowder surrounding these two particles result in an attractive force. When these two particles get closer to each other, the available volume for the polymers increases, increasing entropy. The attractive interaction between the particles can be described by an osmotic pressure $\Pi = nk_BT$ where n is the molar concentration of crowder [246–248]. This entropic effect increases the association rate between two interactants, and is also responsible for the expected collapse in the global conformational sampling of IDPs upon crowding [218, 219]. In the concentrated regime however, there is much less solvent, the solution is very dense in polymer coils yielding a high level of entanglement where the polymers can penetrate the depletion layer overlap between two particles, yielding a decrease in the association rate. The depletion interaction phenomenon is illustrated in figure 2.2.

As a mean to provide a theoretical explanation of crowding effects on protein interactions, Berezhkovskii and Szabo proposed a quantitative theory of crowded-modulated biomolecular reaction rates in the case of low reactant concentration and arbitrary crowder concentration [249]. In their theory, the reactant of size R is assumed to be spherical with a diffusion constant D_0 and a reaction rate k_0 with its partner in pure solvent solution. In this framework, they attribute the decrease in the diffusivity upon crowding to two mechanisms:

1. Collisions with the crowder. To describe this, they define theoritical average "cavities" of radius R_c in which there is no crowder.

2. Non-specific reversible binding to the crowder. To describe this, they define the free diffusion constant upon crowding D_1 and the bound-to-crowder diffusion constant D_2 such that $D_0 > D_1 > D_2$.

Finally, they also describe the phenomenon of depletion interactions between two particles close to each other upon crowding, assuming that they are close enough to each other such that there is no crowder in between them. To describe this, they define a distance dependent potential of mean force $U(r)$ expressed as a square-well potential with $U(r) = 0$ if $r > R_c$ and $U(r) = -\Delta U$ otherwise. In absence of interactions between the reactants and the crowder, the final derived formula for k_{on} is as follows:

$$\frac{1}{k_{on}} = \left(\frac{1}{k_0} + \frac{1}{4\pi D_0}\left(\frac{1}{R} - \frac{1}{R_c}\right)\right)e^{-\beta\Delta U} + \frac{1}{4\pi D_1 R_c}$$

This theoretical framework describes the two important observables that can determine protein interactions in crowded environments: The reduced translational mobility due to the crowder, and the so-called "depletion interactions", described as this effective attractive effect between two particles that are close to each other in a crowded milieu [246–248].

With Berezhkovskii and Szabo's theory in mind, Zosel and coworkers studied the effect of crowding on the interaction between two IDPs that fold upon binding using single-molecule spectroscopy [250]. They found a crowder size dependent increase in the interaction affinity where higher molecular-weight PEG molecules appeared to induce a significantly stronger stabilisation of the interaction. They rationalize the increase in affinity using the theory of depletion interactions. The effect of the size of the crowder on the affinity is explained by the fact that a smaller crowder will induce smaller depletion layers than a larger crowder, which increase the affinity to a lesser extend compared to large crowders. Additionally, they notice that theoretical calculations assuming spherical particles overestimate the volume of overlap $V_{overlap}$ between the two depletion layers. This is explained by the fact that IDPs and the crowder are polymers, and they can enter each other's hydrodynamic spheres. Interestingly, they observe that the k_{on} of the interaction yield non-bijective functions with respect to the concentration of high molecular-weight PEG similarly to what was found with folded proteins [243]. Indeed, from 0 to 100 g/L of PEG with a molecular weight higher than 1000, a non negligible increase in k_{on} is observed, followed by a decrease at higher PEG concentrations. They explained this observation using Berezhkovskii and Szabo's theory in terms of a competition between a slow down due to the slow down of diffusion and an acceleration effect due to the depletion interactions. Fitting the experimental data with the equation above resulted in reasonable agreement, but technical limitations prevented good experimental observations above 100 g/L of PEG.

In this section, we observe that the effect of macromolecular crowding on the dynamic properties of proteins has different sources. First, crowding induces length-scale dependent increases in the translational and rotational properties of proteins. Second, an entropic source often named "depletion interaction" is responsible for increases in the binding affinities in protein interactions. Then, at high concentrations, the density and high concentration in a crowded polymer solution were found to cancel the depletion effects in both folded [243] and disordered proteins [250]. Finally, because crowding agents are never inert, crowder-dependent soft interactions between the crowding agents are known to occur, and besides changing the conformational sampling in IDPs, there is a possibility that they affect the protein's dynamic and interaction properties. However, these soft interactions are not always detected. PEG for example is well known for exhibiting soft interactions with proteins, but PEG remains widely used by the scientific community, and these soft interactions are often not taken into account or not detected, suggesting that they might have a negligible effect on the behavior of the studied proteins. It remains nevertheless important to use several different crowding agents in a study if possible, to account for these soft interactions.

In most of the studied examples, the main technique was based on single molecule fluorescence. This set of techniques is extremely powerful for studying protein interactions in crowded environments and in cells, since it allows the study of multiple different conditions in a reasonable amount of time. In addition, it can provide low-resolution information on the dynamics of internal rearrangements in the protein as well as information on the translational and rotational diffusion in crowded media. This method remains however limited by the use of fluorophore, which can be complicated in very highly crowded solutions [250]. NMR spectroscopy on the other hand is powerful at providing atomic-resolution information on protein dynamics at multiple timescales, as well as protein interactions. However, NMR is limited by time since it requires the equivalent of days of measurement time to extract valuable information on a protein interaction if we are to perform chemical exchange experiments or relaxation experiments, which can compromise an extensive study if measurement time is limited. The strength of NMR over fluorescence methods to study interactions in crowded environments remains nevertheless the possibility to study conditions with higher crowding concentration provided that our protein of interest tumbles sufficiently rapidly, since it does not requires the detection of fluorescence. NMR is therefore an interesting tool and while the viscosity effects on IDP dynamics were well characterized with NMR, the atomic-resolution effects of depletion interactions and concentrated polymer regimes on IDP's dynamics and interactions are not well understood yet.

2.4 Liquid-Liquid phase separation of IDPs

While the existence of membrane organelles in eukaryotic cells is widely established, the importance and prominence of membrane-less organelles were discovered only recently [251, 252]. These structures are composed of proteins, nucleic acids and other small molecules that are separated from the rest of the cell via liquid-liquid phase separation, without the need of any membrane to stabilize it. It is widely suggested that these highly concentrated intracellular compartments are essential for numerous biological functions in addition to organizing the cell [251, 253, 254]. Intrinsically Disordered Proteins are notably present inside these very important dynamic assemblies, and are believed to play a role in their stabilization [255]. We will describe in this section the features of biomolecular condensates by starting with a brief thermodynamical description to explain how these membraneless compartments can form. Then, we will discuss the intermolecular interactions that can be involved in their stabilization. Finally, we will discuss the different techniques to characterize the condensate's dynamic properties, with an emphasis

on NMR spectroscopy.

2.4.1 Thermodynamic description

Liquid-liquid phase separation is a well established phenomenon. It is a spontaneous reversible energy-driven process involving the demixing of several species resulting in a separation of these species into distinct liquid phases. For example, we can consider the phase separation of an IDP with a polymer. Thermodynamically speaking, the demixing process occurs when it is energetically favorable with respect to a mixed phase. This translates as a decrease of the system's free energy upon phase separation, either by an enthalpic contribution or an entropic contribution. Of course in a system with no interactions, the demixing is entropically favorable.

A thermodynamical description of such phase separation was described in 1942 by Flory [256] and Huggins [257]. It is a very simple lattice model that can describe the energy of mixing per unit volume in a diluted polymer solution as a function of the volume fraction of monomers and the polymer-solvent interaction parameter χ. The free energy change upon mixing ΔG_{mix} can be expressed as a function of the enthalpic ΔH_{mix} and the entropic ΔS_{mix} contributions:

$$\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix}$$

Where T is the temperature. Let's consider a very simple case where our system is described by a Cartesian lattice of volume equal to the number of sites N. In this framework, we are going to assume a quasi-solid lattice with interchangeable sites over time for each elements occupying the lattice. This lattice is filled with N_s solvent molecules occupying one site each and N_m monomers of uniform size occupying x sites each so that we have $N = N_s + xN_m$. Our lattice is big enough so that we can consider that the concentration of polymer segments in the lattice is uniform and independent from any local prior knowledge such as the presence of a segment at a given coordinate. The entropic contribution can be determined as:

$$\Delta S_{mix} = -k_B \left[N_s ln(\Phi_s) + N_m ln(\Phi_m)\right]$$

Where k_B is Boltzmann's constant, and $\Phi_i = N_i/N$ is a volume fraction. The enthalpic fraction on the other hand depends on the relative interaction free energies between solvent-solvent, solvent-monomer and monomer-monomer contacts. A favorable interaction between solvent and monomer would favor a mixed solution while favorable interactions between solvent and solvent and/or monomer and monomer would favor phase separation such that we can define a mixing energy per contact as:

$$\Delta u_{mix} = u_{sm} - \frac{1}{2}(u_{mm} + u_{ss})$$

Where u_{mm}, u_{ss} and u_{sm} are the energy terms for monomer-monomer, solvent-solvent and solvent-monomer interactions respectively. From this, the enthalpic contribution to the mixing free

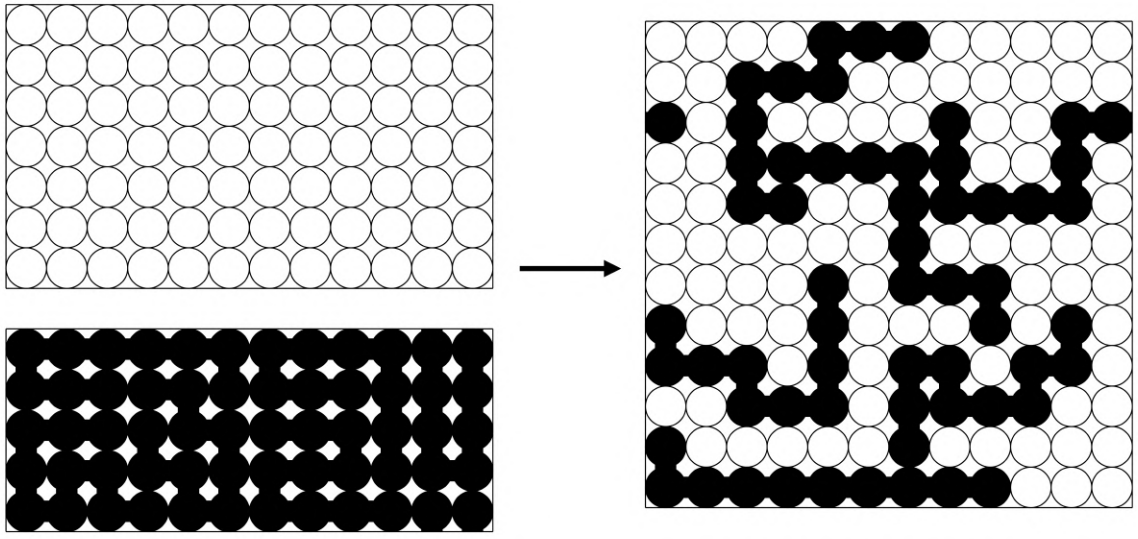


Figure 2.3: Schematic representation of the Flory-Huggins theory. The free energy of mixing is calculated between phase-separated polymer and solvent solutions (left) and the mixed phase (right).

energy is simply the demixing energy per contact multiplied by the total number of monomer-solvent contacts in the lattice, that we can express as the number of monomer sites xN_m times a coordination number z representing the number of nearest neighbors on a lattice site times the proportion of solvent sites in the lattice Φ_s. We thus have:

$$\Delta H_{mix} = \Delta u x N_m z \Phi_s = \Delta u N_s \Phi_m z = k_B T N_s \Phi_m \chi$$

Where we defined an interaction energies-dependent parameter $\chi = z\Delta u / k_B T$. The final result of Flory and Huggins for the mixing free energy as a function of the polymer volume fraction and the interaction parameter χ is thus given as follows:

$$\Delta G_{mix} = k_B T \left[N_s ln(\Phi_s) + N_m ln(\Phi_m) + N_s \Phi_m \chi \right]$$

This very simplistic model provides a remarkably good description of liquid-liquid phase separation of IDPs (Fig. 2.3). Of course, more sophisticated descriptions were developed later on such as the Overbeek-Voorn theory taking into account electrostatic interactions [258] or the random phase approximation [228, 245]. The key concept to keep in mind from these theories is that the favorability of liquid-liquid phase separation of intrinsically disordered proteins is determined by the relative intermolecular interaction energies of proteins with the solvent with respect to proteins with themselves and solvent with itself, as manifested by the parameter χ in the Flory-Huggins theory.

2.4.2 Inter-molecular interactions in biomolecular condensates

The few examples of studied biomolecular condensates of intrinsically disordered proteins were shown to exhibit residue-specific intermolecular contacts which are believed to stabilize phase separation. These interactions can be categorized in four groups [259]:

1. Coulomb interactions between residues with opposite charge
2. Electrostatic dipolar interactions between two dipoles
3. Cation-π interactions between a pi orbital bearing residue and a positively charged residue
4. π-π interactions between two pi orbital bearing residues

Of course, each possible combination of residue pairs exhibits a different interaction free energy. A common framework to describe liquid-liquid phase separation of IDPs is the "sticker-spacer" model [260], taken from the theory of associative polymers [261, 262] in which an IDP in a condensed phase will have intermolecular interactions with sticker residues while the non-interacting spacer regions will remain dynamic and flexible in the system. This framework allows a good understanding of the protein-protein interactions that may stabilize a biomolecular condensate and helps to understand the sequence dependence of biomolecular condensates' stability.

Experimental insight into these intermolecular interactions were provided by Wang and coworkers [260] where they performed an extensive study of liquid-liquid phase separation in the FUS family of proteins composed by intrinsically disordered scaffold proteins. They found that FUS phase separation is mostly governed by intermolecular interactions between tyrosine and arginine residues complemented by some regulating electrostatic interactions. They also found that glycine, glutamine and serine residues play an important role on the properties of the condensate. Gycine residues contributed to accelerating the rearrangements in the condensates while glutamine residues slowed down these rearrangements.

More direct evidence of intermolecular contacts were found by Brady and coworkers who characterized the phase separation of Ddx4 and showed that the phase diagram was highly salt dependent, suggesting that electrostatic and pi interactions [263] are important for phase separation in this system [264]. In addition, they performed ^{13}C-filtered ^{13}C-edited NOESY experiments and found a wide network of inter-molecular contacts. After counting the number of observed contacts for each amino-acids, they found that the highest number of contacts are found in the aromatic and charged residues, namely phenylalanines, lysines and arginines in descending order. Similarly, Murthy and coworkers reported intermolecular NOESY cross peaks using $^{13}C - HSQC - NOESY - ^{15}N - HSQC$ experiments on a mixture of $^{13}C^{14}N$ and $^{12}C^{15}N$ labelled FUS LC condensate [265]. In addition, they performed PRE experiments to probe long-range intermlecular contacts in the condensate. The combined results suggest that inter-molecular interactions in this condensed phase are not localized in particular areas. Complementary two-chains MD simulations and salt-dependent turbidity assays suggested that the phase separation of FUS LC is determined by a combination of nonspecific hydrogen bonds, hydrophobic effects and pi-pi interactions involving all types of pi-bearing orbitals.

Pi contacts are often associated with aromatic residues. Vernon and coworkers from the group of Forman-Kay pointed out the prevalence of pi contacts in proteins and demonstrated that a massive proportion of pi-pi stacking interactions involved non-aromatic atoms, with a proportion of 13 to 1 with respect to aromatic-aromatic stacking [266]. They used their quantitative

data to create a predictor based on the proportion of these pi-pi interactions, and demonstrated that liquid-liquid phase separation in some proteins could be predicted based on the presence of these contacts, suggesting that pi-pi interactions involving non aromatic residues are more important than previously thought.

LLPS-driving intermolecular contacts can also be regulated by post translational modifications, which can be important for biological function [267, 268]. Qamar and coworkers for instance studied cation-pi interactions in the phase separation of FUS and showed that methylation of arginine side chains affects the dynamical properties of the FUS droplets [269]. Methylation inhibits intermolecular contacts, and over-methylation resulted in a dispersed phase of a highly dynamic condensate. On the other hand, hypomethylation allows extensive intermolecular contacts leading to a hydrogels or a fibrillary gel.

In this section, the diverse intermolecular interactions that can stabilize a biomolecular condensate have been discussed. Numerous examples in the literature suggest that these contacts can be different from system to system. Besides electrostatic coulomb and dipolar interactions and pi-pi and cation-pi stacking, the driving forces of LLPS can be more difficult to detect, including hydrogen bonding and hydrophobic effects [265, 270]. Because of this high level of complexity, the prediction of the driving forces for LLPS is still a very active field of study.

2.4.3 Experimental characterization of the dynamic properties of biomolecular condensates

We saw that biomolecular condensates are membraneless compartments stabilized by energetically favorable intermolecular interactions. Experimental characterization of protein dynamics in a condensate is challenging but possible. Translational and diffusive properties of IDPs in the condensate can easily be characterized with ensemble methods like confocal fluorescence imaging, NMR diffusion experiments or single particle methods. Protein structural and conformational dynamics studies on the other hand remain possible by NMR spectroscopy and other techniques like FRET. Although invasive techniques such as residual dipolar couplings are extremely challenging in a biomolecular condensate, conventional NMR observables such as chemical shifts, NMR spin relaxation and any kind of non invasive experiments are still accessible provided that the NMR signal is still observable, which requires that the protein tumbles sufficiently rapidly such that the transverse relaxation rates yield reasonable values

One of the most widely used microscale condensate characterization method is the Fluorescence Recovery After Photobleaching (FRAP) experiment [271]. Briefly, a fluorescent dye is covalently attached to the protein of interest, then while the condensate droplet is observed by confocal fluorescence microscopy, a predefined region is bleached. The recovery of the signal is observed and, using an appropriate model, information on the condensate's fluidity and the protein's diffusion coefficients can be estimated. FRAP studies found diffusion coefficients for IDPs in condensates ranging from 0.001 to 1 $\mu m^2 s^{-1}$, which is at least 2 orders of magnitudes slower than proteins in solution [272]. Among other used methods giving similar information, we can also cite NMR diffusion experiments [265], ultrafast-scanning fluorescence correlation spectroscopy (usFCS) [273] and single particle tracking methods [274, 275].

The microscale and nanoscale diffusion properties of biomolecular condensate is very variable from system to system. Additionally, it is highly sensitive to the composition of the droplet. The presence of RNA and other compounds indeed greatly alter the diffusion properties in such systems [276] and can greatly complicate their dynamic features [277]. It was for instance found

with the disordered P granule protein LAF-1 that the presence of RNA molecules could decrease the viscosity of biomolecular condensates [278]. In addition to its composition dependence, it has been observed that some biomolecular condensates like FUS [274, 279] can evolve over time [280, 281]. Reichheld and coworkers for example studied the dynamic properties of an elastomeric protein in a condensed phase and observed with NMR spin relaxation and PFG diffusion measurements a decrease in the protein's rotational and translational motion over time, manifesting a "maturation" of the condensate in the days timescales.

While the methods for characterizing translational motions of proteins and other molecules in biomolecular condensates are well established, characterizing internal protein dynamics at atomic resolution requires non invasive methods that probe the behavior of highly disordered biomacromolecules in highly viscous and crowded environments. In principle, every IDP characterization methods presented in the first section of this chapter can be leveraged to study IDPs in condensates. NMR remains the unique atomic resolution experimental method for IDPs in the condensates, while complementary techniques like FRET and other fluorescence experiments provide additional information on protein long-range dynamics.

Burke and coworkers studied the FUS LC condensed phase. Carbon chemical shifts indicated that the local backbone sampling is identical to a good extent compared to the dilute phase [282]. Measured ^{15}N R_1, R_2 and heteronuclear nOe at one single magnetic field provided insight into the fast dynamics of this protein in the condensate. Transverse relaxation in the condensate ranged around 20 s^{-1} compared to 3 s^{-1} in the dilute condition, suggesting a significant increase in the chain segment tumbling correlation times along the chain, while nOes slightly increased as well towards a value of 0.5 s^{-1} indicating a slight decrease in the representation of high frequency ps-100 ps motion in the rotational correlation function of NH bond vector, which can be translated as a slight increase in the rigidity of the protein [282]. Their findings suggest a still relatively dynamic condensate modulated by weak inter-molecular interactions that stabilize the condensate and at the same time maintain a certain level of chain mobility. The study of elastin-like proteins [283] as well as Ddx4 [264] liquid-liquid phase separation gives similar conclusions with increased diffusion coefficient and rotational correlation times but a maintained relatively high amount of mobility, allowing the obtention of reasonably good NMR spectra in the condensate. NMR spin relaxation experiments were also measured in the condensed phase of hnRNPA2 Low Complexity (LC) domain at 65°C, yielding R_2 values comparable to the values in the dilute phase at the same temperature, but much higher nOes indicating an increased rigidity of the IDP. These interesting results suggest that conditions are possible where the tumbling correlation time of IDPs in the condensate are similar to dilute state values.

Besides analyzing fast motion of IDPs in the condensed phase, NMR spin relaxation measurements proved to be useful to probe intermolecular interactions. Kim and coworkers studied a CAPRIN1 condensed phase and observed a correlation between the increased transverse relaxation rates and so called hot spots in the protein's sequence where the residues are involved in intermolecular contacts as shown by filtered-edited NOESY experiments [284]. These results suggest that NMR spin relaxation experiments might be useful to find the regions involved in the stabilization of biomolecular condensates.

In addition to fast ps-ns dynamics, NMR can probe chemical exchange events in the condensate, as demonstrated by Yuwen and coworkers on a condensate of Ddx4 [285]. They performed off-resonance ^{15}N $R_{1\rho}$ dispersion in the condensed phase and observed a peculiar relaxation rate behavior as a function of the nitrogen carrier offset characteristic of a relaxation exchange mechanism. They characterized this exchange process with a rate ranging around 20 s^{-1} (op-

timal value 17.7 s^{-1}) and a population around 30% (optimal value 26.7%). The model fitted a relatively high relaxation difference, with a 4.5 times higher R_2 for the excited state, while no chemical shift difference was present. This type of exchange would typically not be observed in conventional CPMG or CEST experiments. The same group also successfully characterized in a condensate a solvent exchange mechanism using a ^{15}N D-CEST experiment designed to measure solvent exchange in proteins using ^{15}N deuterium isotope shift [286] coming from a change in the electronic environment at the vicinity of the amide nitrogen upon exchange from bound proton to bound deuterium [287]. Murthy and coworkers also looked for exchange processes using CPMG which probe chemical exchange and DEST which probe relaxation exchange, yielding no evidence for any exchange process in their FUS LC condensate [265].

In complement to NMR experiments, EPR, fluorescence anisotropy and FRET experiments can provide valuable information on the dynamic properties of macromolecules in a biomolecular condensate. Briefly, starting with an example from EPR, Zhang and coworkers demonstrated using EPR spectroscopy that Δtau187 remains highly dynamic in biomolecular condensate while being involved in long-range interactions with RNA [288]. Another extensively studied IDP is α-synuclein. Ray and coworkers studied α-synuclein in a biomolecular condensate using multiple techniques [281]. They used fluorescence anisotropy to characterize the rotational correlation time of the IDP in different conditions. Focusing only on the difference between the dilute state and the condensate state, the extracted rotational correlation times yielded 1 and 1.6 ns respectively. They also observed a slight increase of the rotational correlation time after 8 days which is a manifestation of an evolution of the condensate over time. They also used FRET experiments among other techniques to probe the interactions that drive LLPS. These experiments allowed them to propose a model to describe the LLPS-mediated aggregation of α-synuclein [281].

Through numerous examples, we saw that NMR is a uniquely powerful technique to probe IDP dynamics and interactions in biomolecular condensates. Recent studies used NMR and other complementary techniques to obtain valuable insight on the dynamic properties of macromolecules and their role in LLPS. There is a clear consensus on the property of these condensates that describe a non-negligible decrease of the protein's diffusion constant and and increase in the rotational correlation times, both due to the increase of viscosity and crowding as well as the presence of droplet-stabilizing inter-molecular contacts of multiple origins. However, IDPs remain sufficiently dynamic in many of these condensates where they yield reasonable good NMR signals, suggesting that these intermolecular contacts are weak enough to allow a good rotational and translational diffusion in the condensates. All these examples also show the enormous diversity of these dynamic assemblies. Indeed, each system is different and yield different features and properties depending on the composition of the droplet and the sequence of the proteins in it. In addition, biomolecular condensates can be time-dependent, as shown by many studies. As NMR spectroscopists interested by the dynamics of IDPs in these condensates, it is therefore of high importance to carefully characterize the studied biomolecular condensate and understand the composition and time-dependent properties of it before studying the important stabilizing mechanisms and the other dynamic features in the condensate.

2.4.4 Computational approaches for describing IDPs in biomolecular condensates

In the recent years, a broad range of computational approaches were developed to describe the properties of IDPs in biomolecular condensates. A significant amount of effort have especially been done in the quest of determining the sequence determinants in the stabilization of IDP's phase separation. In line with this effort, Dignon and coworkers designed a one bead per

residue coarse grained model to study the sequence modulation of the phase behavior of IDPs [289]. Using a hydrophobicity scale (HPS) to describe electrostatic and short-range interactions [290], they were able to obtain IDP's phase diagrams using slab simulations. Demonstrated with FUS LC and LAF-1 to obtain chain length and sequence mutation modulations on the phase diagram as well as intermolecular interactions, the model has been widely used and extended to take temperature effects into account [291–296]. Their model became increasingly popular to study IDP's phase separation. Several other improvements of this model were proposed later on. Latham and Zhang proposed to implement experimental data to such models with an algorithm that performs maximum entropy optimization and least-square minimization to adjust the model parameters to fit the experiments with the simulation [297]. Dannenhoffer-Lafage and Best proposed a new data-driven hydrophobicity scale model optimized from diverse data taken from different IDPs that improves the description of pi-pi interactions [298]. Tesei and coworkers also proposed a data-driven model that optimized the non-bonded interactions with a Bayesian parameter-learning procedure that optimizes radius of gyration and PRE data [299]. The group of Mittal also proposed a new hydrophobicity scale to improve their model which they validated against several IDPs [300]. Of course, besides HPS models, other coarse grained approaches were proposed to obtain phase diagrams such as the lattice-based approach proposed by Choi *et al.* based on the sticker-spacer paradigm called LASSI (LAttice simulation engine for Sticker and Spacer Interactions) [301]. Alternatively, one can also combine molecular dynamics simulations with theoretical models [302, 303].

Despite these tremendous recent efforts, little has been done concerning the dynamic properties of IDPs in biomolecular condensates by simulation. One of the reasons is probably that experimental observables such as NMR spin relaxation or chemical shifts require high-resolution simulations. Coarse Grained models often don't grasp the full detail of the IDPs conformational sampling. In this sense, it is of interest to develop higher resolution coarse grained model that are applicable to large systems and allow the extraction of numerous atomic-resolution experimental observables. Wu and coworkers introduced such coarse-grained model, called AWSEM-IDP based on a similar framework for folded proteins structure prediction [304, 305]. It is a three bead per residue model where each residue is defined by the position of the C_α, C_β and O atoms. The force field contains usual bonded and non-bonded terms as well as a "fragment memory" term that is based on a distance-dependent potential taken from an ensemble derived from either MD simulation or based on experimental data. In addition, another artificial term corrects the tendency of the IDP to collapse. This model allows the extraction of secondary structure informations that are not available with single-bead per residue models. It is likely that such models will rise as a good alternative to conventional coarse-grained model to study IDPs in complex environments.

2.5 Conclusion

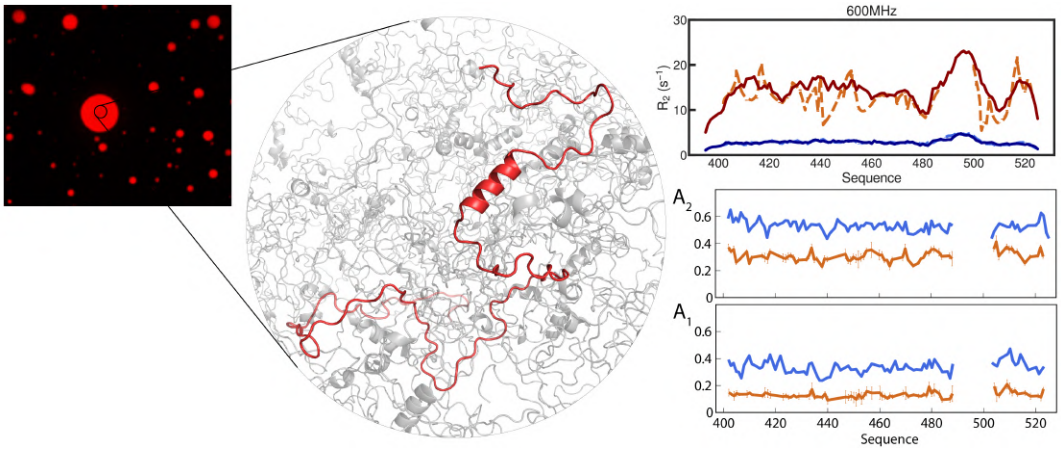
The highly dynamic class of intrinsically disordered proteins has been extensively studied both experimentally and computationally. NMR techniques combined with fluorescence and SAXS are well suited for experimental characterization in different environments, and computational tools to derive ensembles and understand the experimental data are crucial to decipher the complexity of their behavior in solution and in complex environments. While numerous ensemble determination and characterization methods exist, the understanding of how IDPs work *in vivo* is both crucial and under-explored so far, mainly because of experimental limitations. Because of their simplicity, fluorescence methods seem to dominate the field of *in vivo* nanobiology of IDPs, while *in cell* NMR techniques are more difficult and time consuming. Significant progress

has been made in the understanding of the properties of IDPs in crowded environments using fluorescence methods. However, fluorescence remains limited by the experimental constraints in terms of crowding concentrations for example. High resolution information on the dynamical properties of IDPs upon crowding is also still limited to a few studies, and understanding all the parameters requires a significant effort from the whole scientific community. Besides conventional cellular environments, a growing interest towards liquid-liquid phase separating systems can be observed from the current state of the art. This highly important mechanism is best studied with NMR at high resolution. NMR along with computational tools and complementary techniques provided extremely valuable insight into the intermolecular contacts that stabilize the condensed phases. In addition, simulation methods are constantly being developed to predict phase diagrams of IDPs. Nevertheless, high resolution information on the dynamics of IDPs in biomolecular condensate is currently lacking despite its importance to understand the rate of interconversion between conformations, since it is related to the protein's function in many cases.

While a lot of progress has been made both experimentally and computationally on the interactions and low-resolution dynamics of IDPs in crowded environments including phase separated biomolecular condensates, a significant amount of work needs to be done on understanding the effect of crowding and liquid-liquid phase separation on the dynamic properties of IDPs, from their rate of interconversion to their interactions at slower timescales. A better knowledge in this area will allow a better understanding of biological processes, improved parametrization of simulation force fields and a step towards the comprehension of how *in vitro* data are transferable in the context of the living cells.

Chapter 3

Liquid-Liquid phase separation modifies the dynamic properties of Intrinsically Disordered Proteins



Dynamics of Intrinsically Disordered Proteins is key to their numerous functions in biology. Although characterizing their dynamics in solution has been widely studied and well established, a clear picture of their functional motions and mechanisms in highly condensed phases is yet lacking despite its importance for understanding the physics and biology of these widespread phase-separated crowded systems. Here, using the C-terminal domain of the MeV Nucleoprotein, we investigated the dynamics of IDPs in a biomolecular condensate using NMR spin relaxation and molecular dynamics simulations. We observed a significant slow down in the rotational dynamics of the backbone chain in the condensate with respect to the dilute condition, and important modifications of the relative amplitudes of the dynamic modes. This amplitude modulation could be correlated with the increase of unspecific intermolecular contacts upon crowding, showing the implication of entanglement and inter-chain contacts in the dynamics of Intrinsically Disordered Proteins in condensed phase.

This collaborative work has been done along with Serafima Guseva and other coworkers and was published recently ([53] see appendix. Most of the content of this chapter is also in the article). S.G prepared the protein, performed NMR and fluorescence experiments and analyzed the experimental data. V.S analyzed experimental data, prepared, performed and analyzed MD simulations. Performed the 1D NMR experiment.

3.1 Introduction

Membrane-less organelles, intracellular compartments formed by Liquid-Liquid Phase Separation (LLPS) were recently found to be essential to cellular function in biology [306–308]. These recent findings encouraged extensive studies on the molecular mechanisms behind their formation and stability, with the aim of understanding their function in biology [252, 260, 309, 310].

LLPS is a spontaneous and therefore thermodynamically favorable demixing process of one or several species with respect to the rest of the solution, often yielding a highly viscous and concentrated liquid phase. Such phase separation was described by Flory and Huggins in 1942 with a very simple model that still allows a good thermodynamical description of phase separation [256, 257]. The demixing process of a polymer with the solvent for instance depends on the relative affinity of this polymer with itself as well as the solvent with itself compared to the affinity of the polymer with the solvent. Multivalent intermolecular interactions between several components, often including IDPs and RNAs thanks to their relative flexibility, are thought to determine the formation of membrane-less organelles.

In the current state of the art concerning biomolecular condensates, protein concentration in condensates were found to be up to 3 orders of magnitude higher than in a mixed phase. The recent studies also found that the conformational sampling of IDPs in the condensate were not dramatically affected, meaning that they stay disordered in both phases. Residue-specific intermolecular interactions and long-range contacts were also extensively studied, giving diverse results. A common finding is that these intermolecular interactions are weak, and maintain a significantly high level of dynamics, allowing rapid rearrangements which manifest as liquid-like properties on the macroscale.

The remaining high-level of dynamics in biomolecular condensate was also extensively studied since it should play an important role in the stability of the condensate and the kinetic properties of molecular reactions and interactions. The translational diffusion was probed with FRAP and NMR diffusion experiments and showed a significant slow down in the diffusion properties of the molecules in the condensate. On the other hand, reorientational and internal dynamics of IDPs was probed with PREs, EPR, FCS and FRET. The consensus being that the rotational correlation time of proteins were slowed down in agreement with translational diffusion data.

Besides experimental studies, computational approaches have been successfully used to study the behavior of biomolecular condensates. For instance, coarse-grained methods can be used to predict the phase behavior of IDPs and some condensate properties. All-atom simulations also can provide insight into the atomic-resolution behavior of proteins in a biomolecular condensate, with the high cost of computational power.

Among all the techniques, NMR is a unique tool to study the dynamic properties of IDPs in a biomolecular condensate. While it successfully probed conformational sampling, long-range contacts and intermolecular contacts at atomic resolution, it can also probe rapid backbone dynamics from tens of picoseconds to tens of nanoseconds with the help of NMR spin relaxation. The few studies that report NMR relaxation rates in condensates show a significant slow down in the rotational correlation times, in agreement with fluorescence studies. NMR spin relaxation is uniquely suited to provide detailed information on the dynamic modes of IDPs at several timescales. A high level of valuable information can be obtained with extensive NMR spin relaxation measurement at different magnetic fields.

In this study, we performed multifield NMR spin relaxation measurements of the intrinsically disordered MeV Ntail to characterize its dynamic properties in a biomolecular condensate. Additional extensive NMR spin relaxation measurements in dilute and crowded conditions allowed a relevant comparison and characterization of the effect of phase separation on the dynamic properties of IDPs. Our findings show that MeV Ntail's conformational sampling in the condensate is relatively conserved. On the other hand, we find a significant slow down in the rotational correlation times of the different dynamic modes of the IDP. In addition, the contribution of these dynamic modes to the observed dynamics is significantly modified. Using atomistic MD simulations of MeV Ntail in dilute and concentrated conditions, we could reproduce the effect of phase separation on the dynamics of MeV Ntail, and we found a correlation between the amount of non-specific inter-molecular contacts and the amount of redistribution of the IDP's dynamic modes. This combined experimental and computational study provides important insight into the dynamic properties of IDPs in biomolecular condensates.

3.2 Materials and methods

The work presented here has been performed as described in Guseva & Schnapka *et al.* [53]. The content of the materials and methods section can be found in the article in the appendix section.

3.3 Results and discussion

3.3.1 MeV Ntail liquid-liquid phase separation

The C-terminus construct composed of residues 401 to 525 of the Measles Nucleoprotein was found to phase separate upon addition of PEG10000. This construct is rich in arginines (10% of the residues) and charged residues, suggesting a condensate stabilized by electrostatic interactions. The mixing of two solutions of these polymers leads to a separation into two phases: One rich in PEG and poor in protein content, and the other one rich in MeV Ntail, yellow in apparence, and poor in PEG. Fluorescence microscopy of mixed PEG and MeV Ntail solution showed the presence of MeV Ntail in the formed droplets of condensed phase and FRAP experiments showed the liquid nature of this condensed phase (Figure 3.1A,C). PEG acts here as a trigger for phase separation. Its role is probably to modify the solvent properties, which translates into modifying the solvent-solvent and solvent-protein terms in the interaction parameter of the free energy of mixing if we interpret it in the light of Flory-Huggins theory presented in chapter 2.

The concentration-temperature phase diagram of MeV Ntail was established at different salt concentrations and exhibited an upper critical solution temperature [53]. Concentration in the condensed phase varied from 39 mM at 288 K and 118 mM NaCl to approximatively 15 mM at 308 K and 377 mM. Both temperature and salt concentration affected MeV Ntail's concentration in the condensed phase, with higher concentrations at low salt concentration and low temperature. Most interestingly, phase separation appeared to be more prominent at low salt concentration, which again suggests that electrostatic interactions are important for demixing. The results were fitted to the Flory-Huggins theory and yielded a positive interaction parameter χ suggesting that attractive interactions promote phase separation [53].

Overall, from these observations, MeV Ntail seem like an ideal model system for investigating the properties of IDPs in a biomolecular condensate.

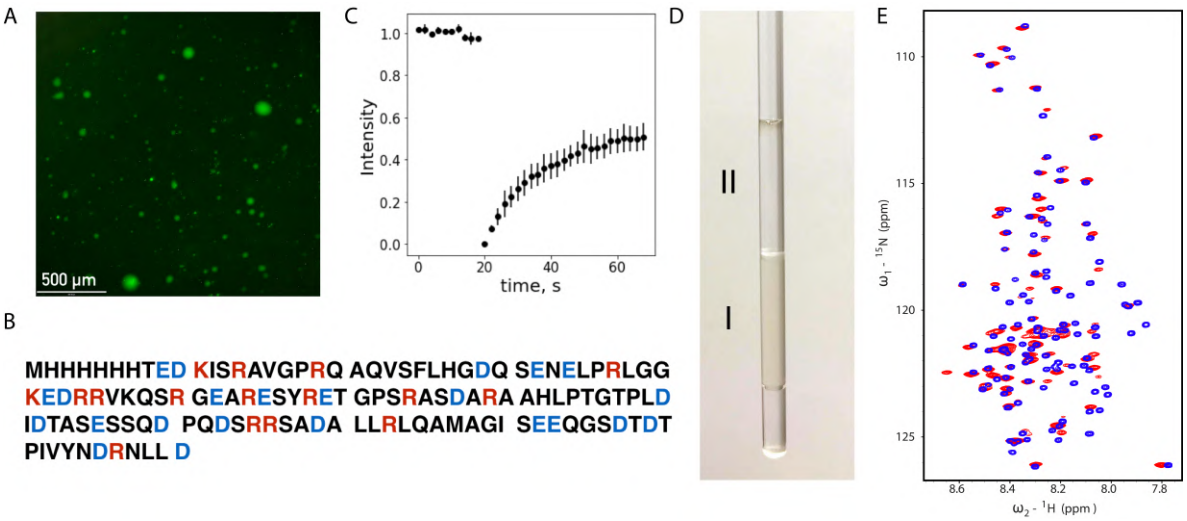


Figure 3.1: MeV Ntail phase separates upon mixing with PEG10000. A: Fluorescence microscopy image of a mixture of PEG and MeV Ntail with fluorescent-labeled MeV Ntail with FAM. B: Primary sequence of MeV Ntail. Positively charged and negatively charged residues are colored in red and blue respectively. C: FRAP data of MeV Ntail droplet. The intensity of a region of interest (Part of a dense phase droplet) is monitored during a fluorescence bleaching experiment. D: NMR sample of Ntail exhibiting the dilute phase (II) and the dense phase (I) in a 3mm shigemi tube. E: ^{1}H-^{15}N HSQC spectra of MeV Ntail in the dense phase (red, 850 MHz) and in the dilute phase (blue, 950 MHz) at 298 K. Taken with permission from [53]

3.3.2 NMR chemical shifts in the dense phase

NMR spectroscopy was performed in the condensed phase. A sample composed of a dense macrodroplet was prepared with 10% carbon 13 and nitrogen 15 isotope labeling. The comparison of the ^{1}H-^{15}N HSQC of MeV Ntail in buffer solution (dilute state) and in the condensate show no significant change in $^{1}H_N$ and ^{15}N chemical shifts in the protein (Figure 3.1E). Almost all peaks remain visible in the condensed phase, suggesting a sufficiently extensive rotational tumbling of NH-bonds throughout the protein to allow reasonable values of transverse relaxation rates. The helical propensity domain is however invisible, either due to exchange mechanisms or slowed down tumbling in this region. Carbon chemical shifts were also observed and show no significant change in the conformational sampling of MeV Ntail in the condensed phase (Figure 3.2). These results combined show that MeV Ntail remains disordered and maintains a reasonably similar local sampling of Ramachandran space in the condensed phase with respect to the dilute state.

3.3.3 NMR spin relaxation of MeV Ntail upon phase separation

To study the backbone dynamic properties of MeV Ntail, NMR spin relaxation rates including R_1, R_2, nOe and η_{xy} were measured at different magnetic fields. The dynamic properties of MeV Ntail were first characterized in the dilute state and in crowded conditions corresponding to 0, 37 and 75 g/L of PEG10000 where MeV Ntail doesn't phase separate. The results highlight the helical propensity domain with elevated R_2 and nOe values with respect to the disordered regions, and all the rates show typical values expected in IDPs (Figure 3.3).

NMR spin relaxation in the dense phase shows significantly different values compared to the dilute conditions at the same temperature, with a general sequence-dependent trend that follows what is observed in the dilute conditions as we can observe for example with elevated nOe values around residues 439 and 451 (Figure 3.3). Most notably, transverse relaxation rates are

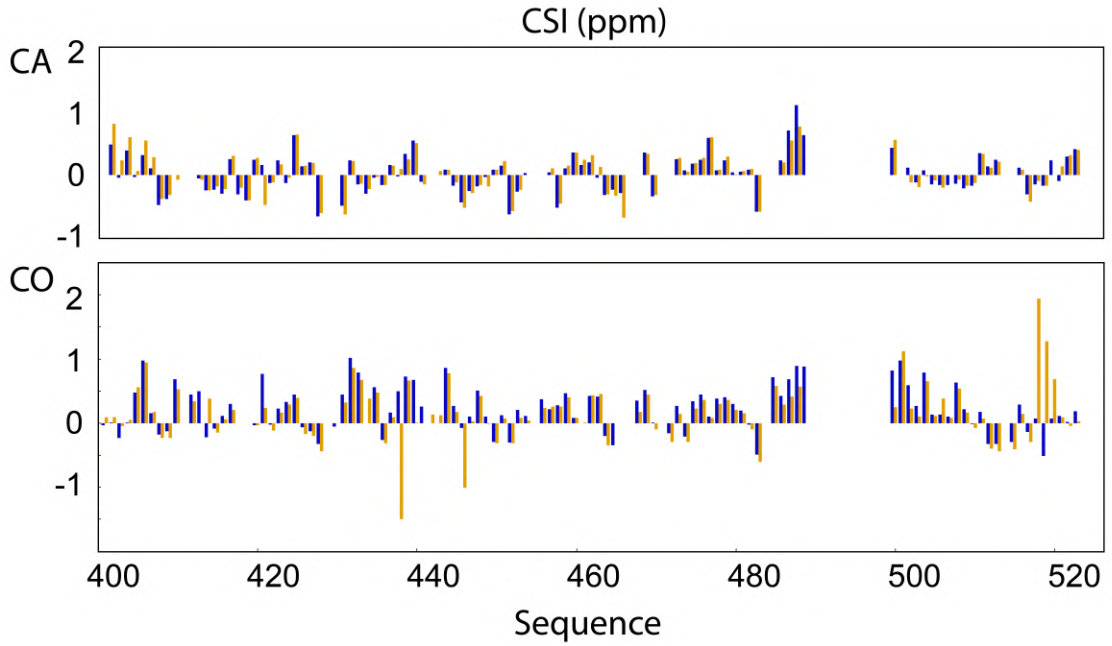


Figure 3.2: Carbon alpha (CA, top) and carbonyl (CO, bottom) secondary chemical shifts of MeV Ntail in the dilute state (Blue) and in the dense phase (Orange).

particularly higher in the dense phase, with rates ranging around 15 s^{-1}. A significant sequence dependence is also observed, including regions with more elevated R_2 values such as $^{406}KI^{407}$ or $^{516}IVY^{518}$ in the terminal regions for instance.

3.3.4 Model-Free analysis reveals the modified dynamic properties of MeV Ntail upon phase separation

Model-Free analysis was performed in both the dilute states and in the condensed phase. The spectral density function was modeled as a sum of three Lorentzians as described in previous studies [51, 52].

For the dilute state, the relaxation rates in the different PEG concentrations were fitted together with a model based on the nanoviscosity of the solution [52]. 20 experimental relaxation rates were fitted this way for each residue in the protein. The fitting results show a good agreement with experiment (Figure A.2). Experimental measurements performed at 700 MHz were used for cross validation (figure A.3). The resulting correlation times and amplitudes associated with the three dynamic modes are similar to recent analyses performed on other IDPs [51, 52], with correlation times fitted around 50 ps for fast motions, 1 ns for intermediate motions and 5 to 10 ns for motions associated with the slowest dynamic mode as we can see in figure 3.4. The viscosity coefficient obtained thanks to the different viscogen concentrations also exhibit similar results compared to other IDPs [52]. The slow motions are important to relaxation in the helical propensity domain and a region between residues 435 and 441, where faster motion is more restricted. Finally, the angle between the NH dipolar vector and CSA principal axes, also fitted, yielded similar results to other IDPs, with an average of 26.5°.

For the dense phase, Model-Free analysis performed with η_{xy}, R_1 and nOe (without R_2) at the two fields exhibits a non negligible gap between correlation times back calculated from R_2 and

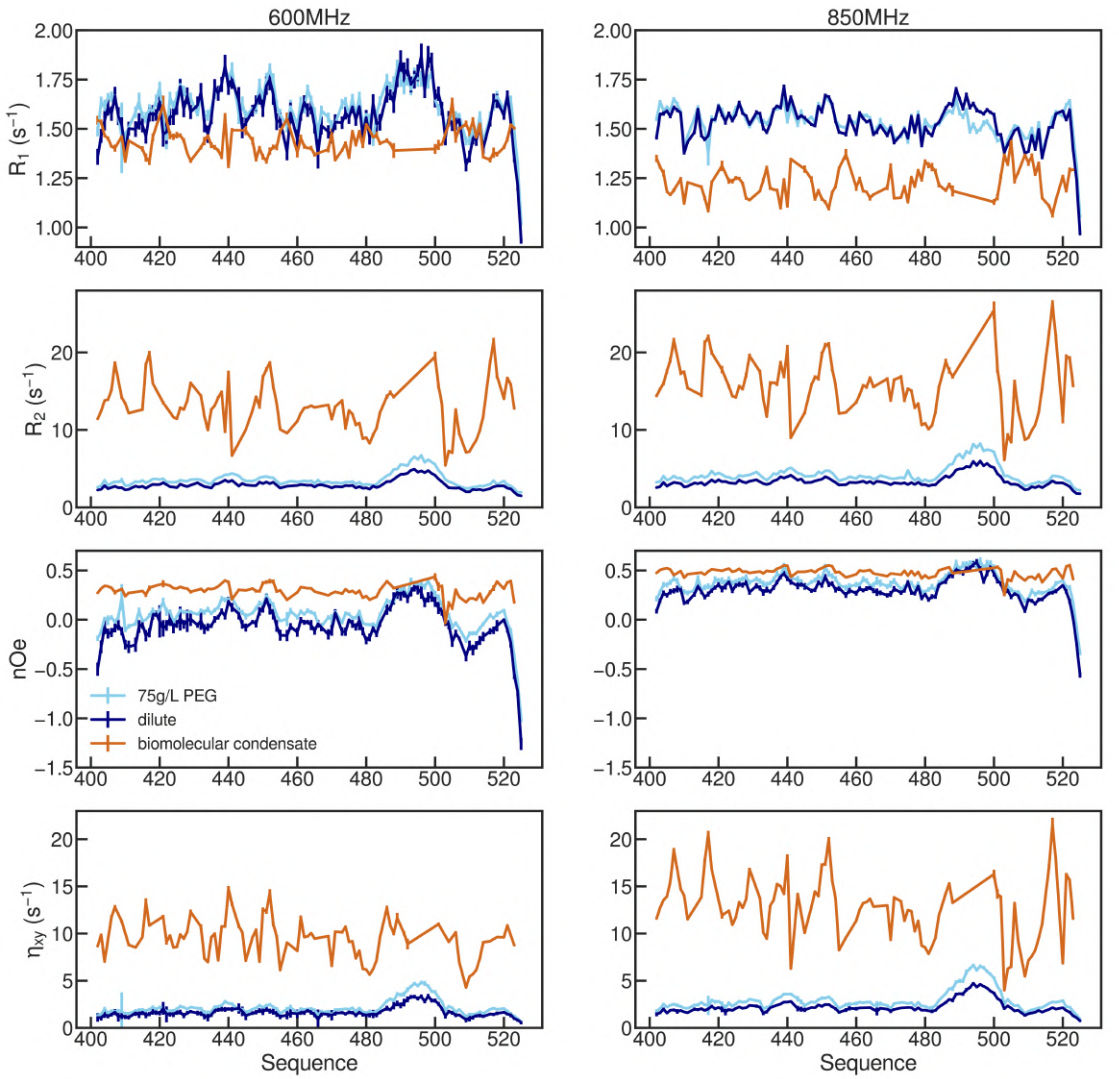


Figure 3.3: Relaxation profiles of MeV Ntail in the dilute state (dark blue), with 75 g/L of PEG (light blue) and in the dense phase (dark gold) at 600 (left) and 850 (right) MHz proton frequency.

from η_{xy}, suggesting a possible exchange contribution to R_2 (Figure A.4). The small exchange contribution to R_2 extrapolated from this analysis appeared to show weak field dependence, suggesting a slow/slow-intermediate exchange regime or a relaxation exchange. Such exchange was characterized before on another protein involved in phase separation [285]. CPMG relaxation dispersion experiments performed in our dense phase did not show any exchange contribution to R_2 and therefore no trace of chemical shift exchange (Figure A.5). An attempt to exploit nanoviscosity as a means to predict rotational correlation times in the dense phase was unsuccessful, possibly because of additional contributions to the measured R_1 of water protons, invalidating the assumption of linearity between this rate and nanoviscosity [311]. Removing one cross-correlated relaxation rates from the analysis yields accurate data compared with the excluded rate, showing an overall consistency of the data sets at the two magnetic fields (Figure A.6) despite the limited amount of experimental data.

Comparing Model-Free analysis in both the dense and dilute states reveals strong differences in the fitted correlation times. First, the fast correlation time is fitted around 100 ps in the dense phase against 50 ps in the dilute state. Then, the intermediate timescale, fitted around 1 ns in

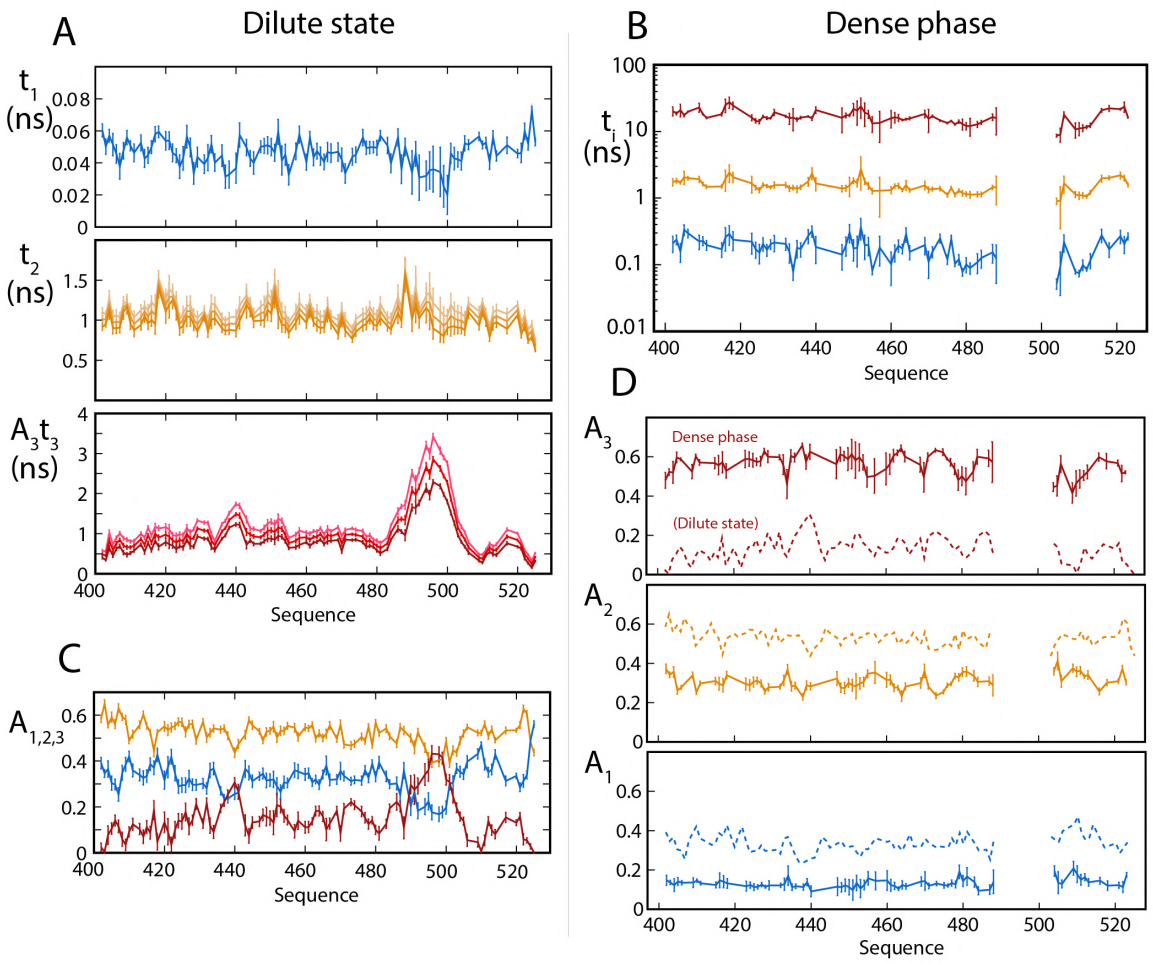


Figure 3.4: A: correlation times associated with the dilute state fitted from Model-Free analysis. B: Correlation times associated with the dense phase fitted from Model-Free analysis. C: Amplitude of the dynamic modes over the sequence in the dilute state from Model-Free analysis. D: Amplitudes of the dynamic modes in the dense phase (continuous lines) compared with the dilute state (dashed lines). Data associated with the fast, intermediary and slow dynamic modes are in blue, orange and red respectively.

the dilute state is fitted around 1.5 ns in the dense phase. Finally, the slow component, fitted between 5 and 10 ns in the dilute state samples timescales over 20 ns in the dense phase (Figure 3.4 and A.7). Remarkably, a significant variation of the amplitude associated with each dynamic mode is observed from the dilute state to the dense state. Associated with the parameters S_f^2 and S_i^2 presented in chapter 1, or more generally with the asymptotic values of the two fastest correlation functions associated with the NH bond vector, this can be interpreted as more restricted fast motions in the dense phase compared with a dilute state. As a result, the contribution to relaxation of the slowest dynamic mode is significantly more important in the dense phase.

3.3.5 MD simulations of MeV Ntail in the dilute state

In order to better understand the effect of LLPS on MeV Ntail's dynamic properties, MD simulations of MeV Ntail were performed. As we saw in chapter 2, MD simulation is a powerful tool for helping the interpretation of NMR relaxation data. Here, MD simulations of MeV Ntail were first performed in the dilute state to obtain a trajectory ensemble that reproduces the dilute state experimental data. 30 trajectories starting from a conformation taken from an ASTEROIDS en-

semble were simulated for 200 ns with CHARMM36m force field and TIP4P/2005 water model, the combination that showed to provide the best description of IDP dynamics so far [210]. Extracting the relaxation data on the resulting ensemble of trajectories yields an overestimated transverse relaxation rate, as expected from recent studies of other IDPs [208, 210]. To obtain an accurate ensemble, an ABSURD procedure was performed [208]. First, the 30 trajectories of 200 ns were divided into 90 smaller trajectories of 100 ns. The reweighting procedure was performed on this ensemble using R_2 measured at 850 MHz proton frequency as a reference, yielding a more accurate representation of MeV Ntail dynamic properties as shown in figure 3.5. The most remarkable feature of ABSURD optimization is that all the rates that were not included in the optimization procedure also show a significant improvement in the agreement with experimental data, as observed previously [208, 210]. Additionally, the local backbone sampling is in agreement with experimental data (Figure 3.7). We therefore successfully obtained an accurate representation of MeV Ntail dynamics in the dilute state with MD simulations.

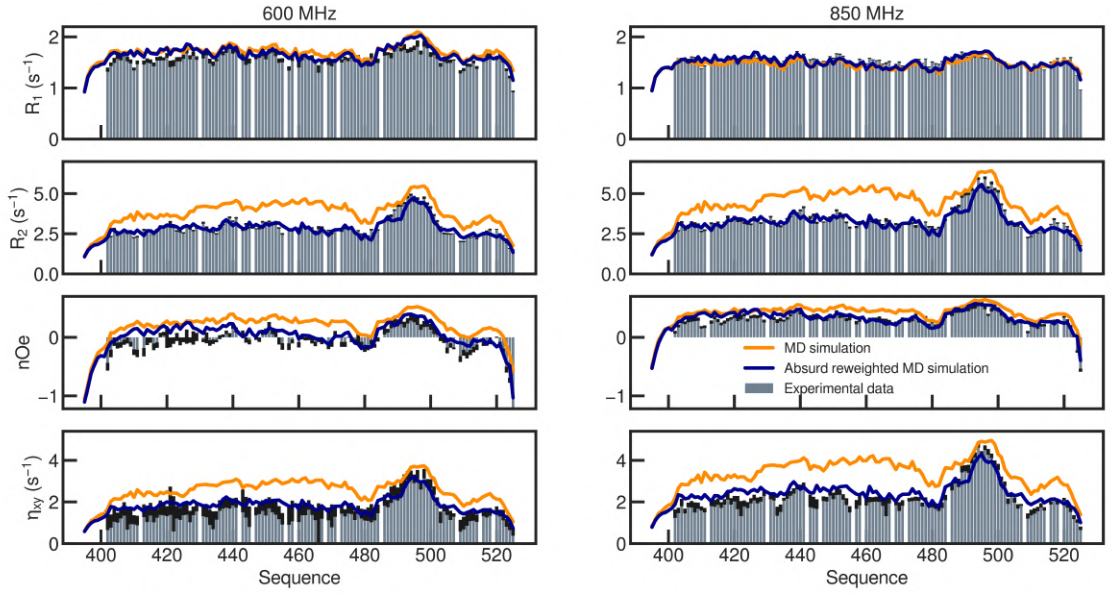


Figure 3.5: Experimental (gray bars), simulates trajectories (orange) and ABSURD reweighted trajectories (blue) of MeV Ntail relaxation rates at 600 and 850 MHz proton frequency.

3.3.6 MD simulations of self crowded MeV Ntail

Simulating large systems at atomistic resolution is a challenging task. The accuracy of the force fields being not guaranteed in the dilute conditions, obtaining good results with highly concentrated boxes of MeV Ntail was not expected. However, it is so far one of the only simulation method available that allows us to extract relaxation parameters despite the rise of numerous coarse grained models [289, 312]. To create self crowding conditions, several boxes made of hundreds copies of MeV Ntail with concentrations of 5, 14 and 20 mM were prepared (Figure 3.6). The boxes were prepared by taking random conformations of MeV Ntail taken from the ASTEROIDS ensemble and incorporating them in the box while avoiding steric clashes. Water and ions are then added such that it neutralizes the charges in the system. Above 20 mM, steric clashes are too important, preventing a proper box preparation. However, 20 mM, which corresponds to approximately 260 g/L of proteins, remains a sufficiently high concentration in the range that can be expected in biomolecular condensates. MD simulations were performed for at least 200 ns using the same force field and water model combination that was used in dilute

simulations.

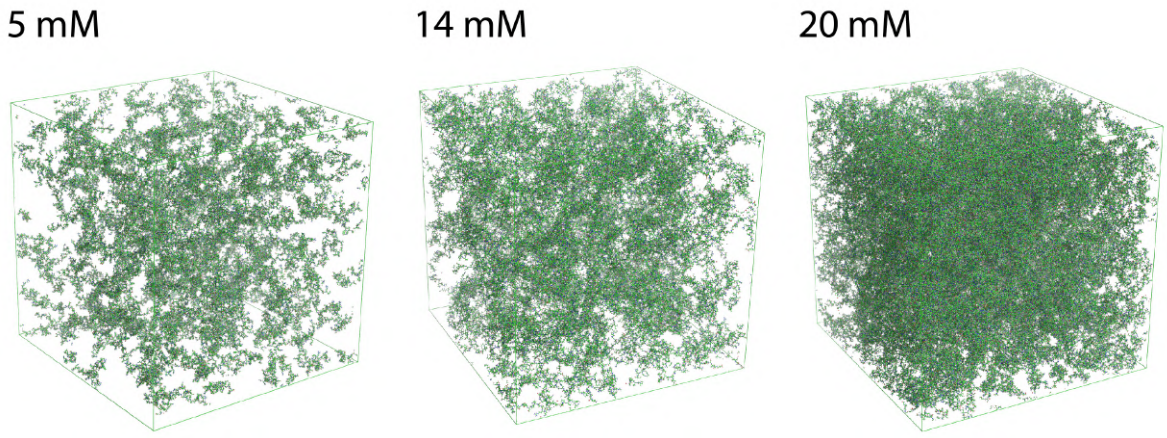


Figure 3.6: Boxes of concentrated solutions of MeV Ntail at 5 mM (left), 14 mM (center) and 20 mM (right).

First, the structural properties of MeV Ntail in each box were analyzed. Chemical shifts were predicted for each conditions and exhibited no significant change with respect to the dilute and experimental condition (Figure 3.7), indicating an accurate sampling of the local conformational space with respect to the experimental observations. Then, the radius of gyration was examined and remained stable in each simulation (Figure A.8). The radius of gyration distribution remained close to the ASTEROIDS ensemble (Figure A.9).

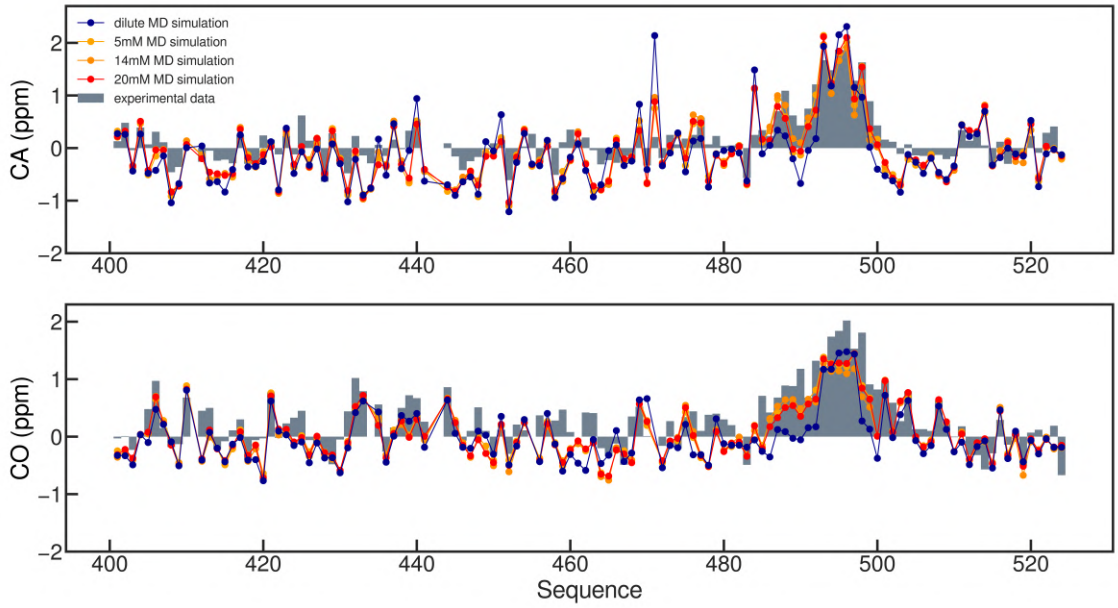


Figure 3.7: Experimental (gray bars) and simulated chemical shifts of MeV Ntail in the dilute state (blue) and in self crowded boxes of 5 mM (yellow), 14 mM (orange) and 20 mM (red).

NMR spin relaxation rates associated with the different simulation conditions were extracted, yielding highly robust data thanks to the high number of Ntail copies in each boxes. We observe a consistent increase in transverse relaxation rates and transverse cross-correlated relaxation rates with protein concentration (Figure 3.8). The highest concentration exhibits a similar base-

line for J(0)-sensitive rates compared with experimental data in MeV Ntail's dense phase. We note however the presence of specific regions were the experimental J(0)-sensitive rates exhibit some spikes that are not reproduced by simulation. Additionally, nOe values also exhibit a consistent evolution, with the highest simulated concentration exhibiting relatively consistent values with respect to experiments. However, R_1 is not well reproduced despite a consistent evolution with respect to protein concentration (assuming we are in the high τ_c regime for R_1). This suggests that the motions associated with intermediate relaxation-active timescales are less well reproduced compared to the fast and slow contributions.

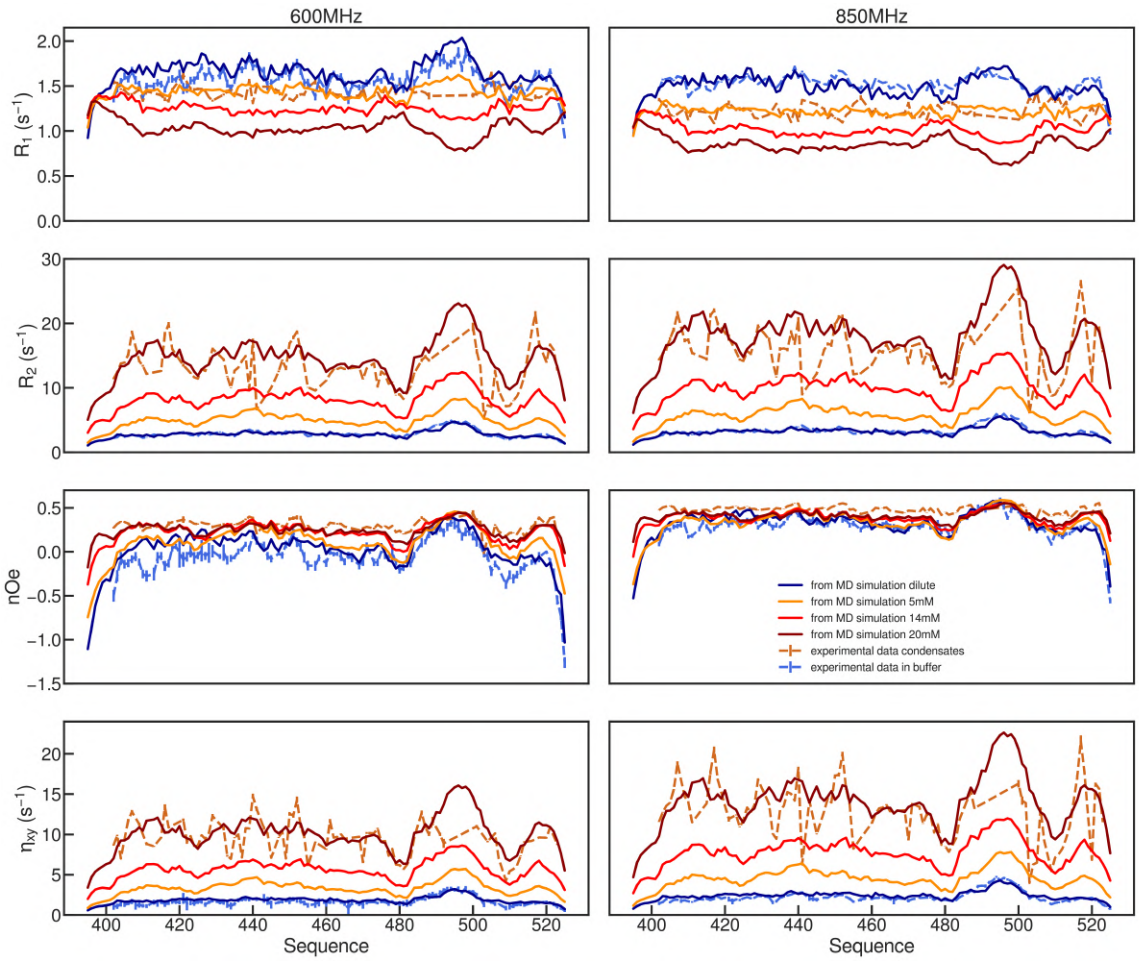


Figure 3.8: Experimental (dashed lines) and simulated (continuous lines) relaxation profiles of MeV Ntail in the dilute state (blue), at 5 mM (orange), at 14 mM (red), at 20 mM (dark red) and in the dense phase (dashed dark gold) at 600 (left) and 850 (right) MHz proton frequency.

Model-Free analysis was performed on the calculated relaxation rates. The motional timescales get higher as expected, and the amplitude modulation in the simulation analysis is akin to the observations in the experimental analysis, as figure 3.9 shows.

The MD simulations were further analyzed to better understand the origins of this amplitude redistribution. We should recall that at least qualitatively, the different amplitudes fitted from Model-Free analysis are somehow related to the amount of orientational space explored by the N-H bond vector at the given timescales. It was shown for folded proteins that the order parameter S^2 was closely related to the amount of contacts between some specific atoms and neighboring

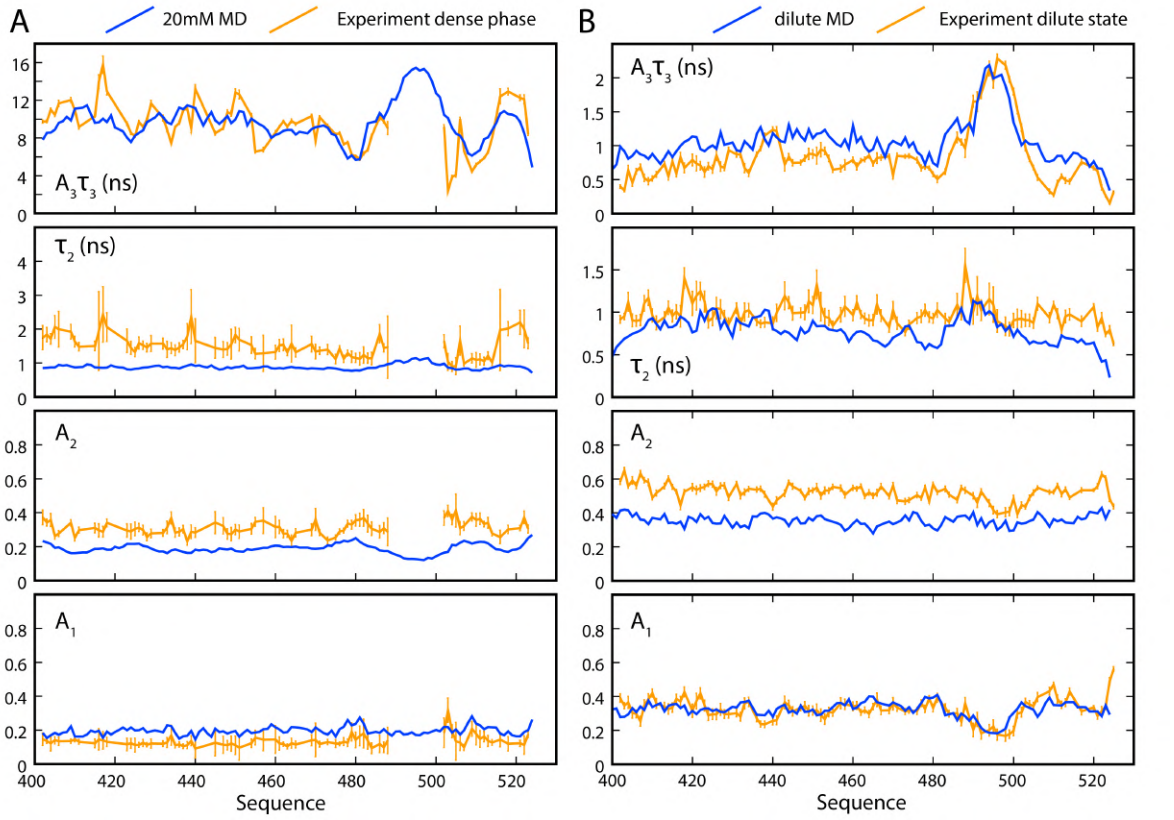


Figure 3.9: Model-Free parameters from Experiments (Orange lines) in the dense phase (left) and in the dilute state (right) and from MD simulations (Blue lines) in the 20 mM box (left) and in the dilute state (right).

atoms in space. Brüschweiler and coworkers were indeed able to predict the order parameters of folded proteins from X-ray crystal structures using a contact model [313]. A related approach was applied here on our trajectories in the different boxes: Contacts between amide protons and the neighboring atoms were calculated over the simulations using the following relationship:

$$W_i = \sum_j \left\langle exp\left(-\frac{kr_{ij}}{m_j}\right)\right\rangle$$

Where the index i corresponds to a given residue, the index j corresponds to an atom in the vicinity of the amide proton attached to the residue i, r_{ij} is the distance between these two atoms and m_j is the mass of the atom j. We noted that ignoring the neighboring protons and only taking into account heavy atoms (Carbons, Oxygens, Nitrogens) didn't significantly affect the results. The expression is averaged over the trajectories and over the ensemble for every given atom pairs. k is an arbitrary tunable parameter expressed in units of mass over distance. The parameter W_i is used as an observable of the amount of promiscuity between a given N-H bond vector and the surrounding atoms, and k tunes the sensitivity of this parameter with respect to distance. A high value of k would imply that W_i is sensitive to only very close atoms, while the lower the value of k, the more sensitive W_i is to more distant atoms (Supplementary figure 3.10A). After calculation (Supplementary figure 3.10B), the parameter W_i was then rescaled to be compared to the fitted fast motion order parameter $S_f^2 = 1 - A_1$ with the following relationship:

$$(1 - A_1)_i = \frac{1}{2} tanh\left(B \sum_j \left\langle exp\left(-\frac{k r_{ij}}{m_j} \right) \right\rangle + C \right) + \frac{1}{2}$$

Where B and C are variables and the tanh function serves only to restrain the values between 0 and 1. The results with k = 6 are shown in figure 3.10C and 3.10D and show that the distribution of intermolecular distances over the sequence exhibits a similar evolution with protein concentration. This suggests that the reduction of fast motional amplitude observed thanks to NMR is related to an increase in motional restrictions due to a higher level of proximity between the chains.

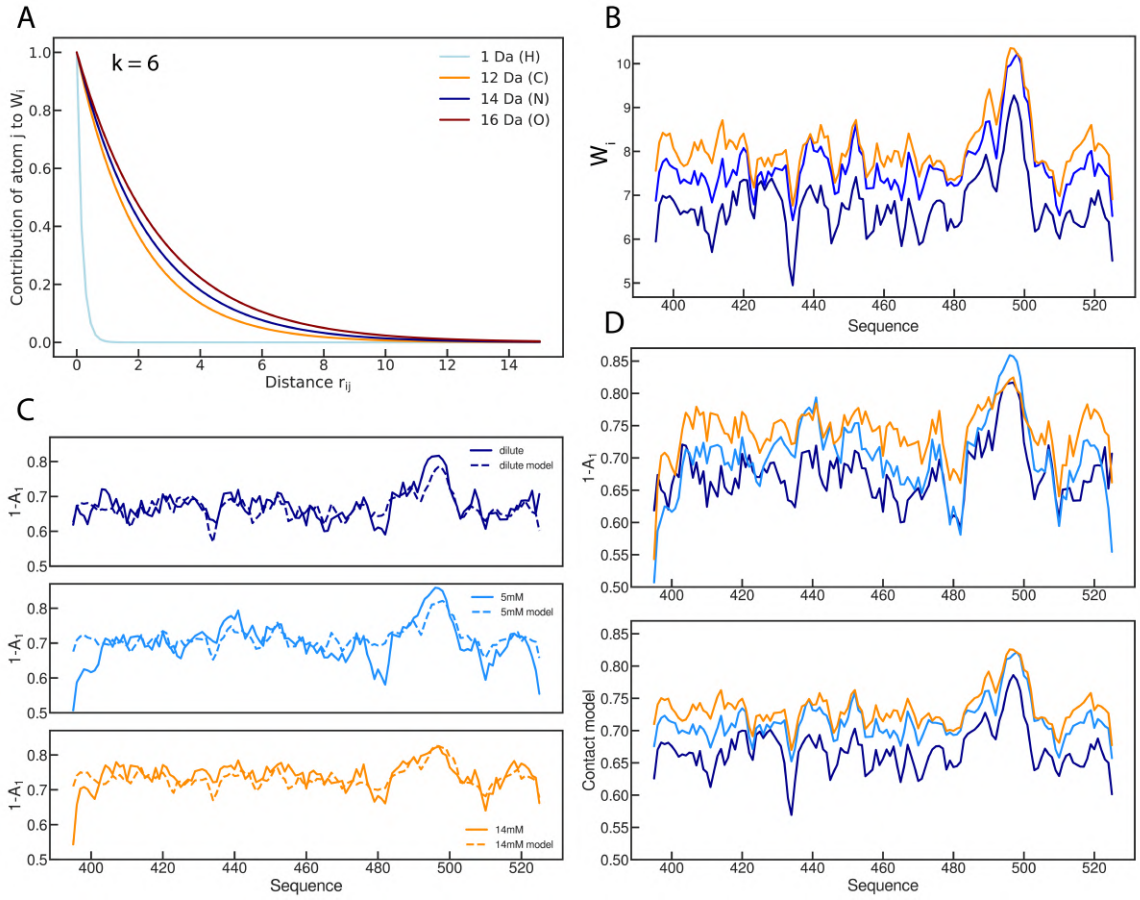


Figure 3.10: A: Distance (in Angstrom) dependence of the contribution to W_i of a single atom with a mass of 1 (light blue), 12 (orange), 14 (dark blue) or 16 (red) Daltons. B: Value of W_i for each residues i over the sequence calculated with k = 6 in MeV Ntail's simulations in the dilute state (blue), at 5 mM (green) and at 14 mM (orange). C: Superimposed fast motional mode order parameter and contact model in MeV Ntail's simulations in the dilute state (top), at 5 mM (center) and at 14 mM (bottom). D: Top: Fast motion dynamic mode order parameter extracted from the simulations in the dilute state (dark blue), at 5 mM (light blue) and at 14 mM (orange). Bottom: Contact model over the sequence.

3.4 Discussion

Understanding the dynamic properties of IDPs in biomolecular condensates is crucial to understand biological mechanisms taking place in such dynamic and nevertheless viscous and crowded environments. Here, we combined NMR and MD simulation to investigate the dynamic properties of MeV Ntail, a prototypical IDP, in a liquid-liquid phase separated condensed phase.

MeV Ntail is the C terminal domain of the Measles Virus Nucleoprotein (MeV N) and is used here as a model system. This construct is involved in the phase separation of MeV N with its partner the Phosphoprotein (MeV P) through an interaction between the helical propensity domain and the XD domain of MeV P [314]. This phase separation is believed to be important for the formation of Nucleocapsids in the MeV replication mechanisms. Studying the phase separation of MeV N with MeV P by NMR spectroscopy would be interesting but challenging given the size of these IDPs.

Using NMR, we measured NMR spin relaxation rates of MeV Ntail in dilute and crowded conditions as well as in the dense phase at two distinct magnetic fields. The relaxation rates measured in dilute states are similar to those measured in other prototypical IDPs in previous studies [51, 52]. On the other hand, the measured rates in the dense phase are significantly different, although sequence specific similarities in the relaxation profiles could be found. For instance, the transverse relaxation profile in the dense phase, which exhibits much higher rates than in the dilute state, presents some shared maxima with respect to the profile in the dilute state, including the highly charged regions $^{438}RRVK^{441}$ and $^{479}ESYRE^{453}$. However, additional maxima observed in the dense phase were not present in the dilute state, including the aromatic ring bearing regions $^{418}FLH^{420}$ and $^{517}VYN^{519}$, indicating the probable presence of dense phase-specific intermolecular interactions as observed in other biomolecular condensates [264, 265, 284]. In addition to these maxima, the R_2 profile doesn't exhibit the typical "bell shape" one would expect in classic IDPs [315], which suggests that the typical reptation motion of the tails of the protein chain is more restricted in the dense phase. While numerous research groups have already characterized intermolecular contacts by NMR, a more detailed analysis of the sequence specificity of the dynamic properties of MeV Ntail would be interesting to decipher the effect of intermolecular interactions on IDP's dynamics with respect to the more general characteristics detailed in this study.

In order to interpret these relaxation data, Model-Free analysis was performed for the dilute states and the dense phase. The extensive amount of data in the dilute state with different levels of nanoviscosity allowed a robust analysis of the dynamic properties of MeV Ntail. In addition to correlation times and amplitudes of the three dynamic modes, solvent friction coefficients could be extracted, showing consistent values with respect to previous measurements [52]. Indeed, the correlation time of the fast mode remains sequence and viscosity independent and ranges around 45 ps, while the slowest motion remains around three times more sensitive to increased viscosity compared to the indermediary mode, which is a manifestation of the length-scale dependence of the experienced viscosity in complex environments, which further illustrates the different length-scales of the motional mechanisms that are probed by NMR spin relaxation. In the dense phase, the timescales increase significantly as expected, and a significant redistribution of the amplitudes of each dynamic mode is observed. A non-negligible systematic decrease in A_1 and A_2 results in a more significant contribution of the slower motions to relaxation.

To further understand this amplitude modulation, atomistic MD simulations were performed. First, simulations of MeV Ntail in salty water solutions were performed and reweighted with

ABSURD to obtain a representative ensemble of trajectories that agree with experimental data. Then, boxes of proteins at different concentrations of MeV Ntail allowed a characterization of MeV Ntail in concentrated conditions. One should note that these boxes were not designed to reproduce liquid-liquid phase separated biomolecular condensates. Rather, it simulates highly self crowded conditions with concentrations that are analogous to a dense phase. The main difference with our simulations is that no dense phase-specific intermolecular interactions were simulated, since the boxes were generated with random conformations taken from an ensemble. Furthermore, being able to see intermolecular interactions in our simulations would have required much larger timescales that are currently unfeasible in atomistic simulations. An interesting approach for that would be to perform coarse grained simulations and then after phase separation occured, translate the coordinates into all-atom and run simulations this way [316]. There is however no guarantee that it would yield accurate results since the current available force fields cannot grasp some interactions, for example between pi orbitals bearing atom groups [298]. Nevertheless, self crowded conditions on the other hand are useful to obtain information on the dynamic response of MeV Ntail's chains to self crowding, a key characteristic that is common with biomolecular condensates.

The chemical shifts and radius of gyration distribution extracted from simulation are in agreement with experimental data, allowing further analysis. The relaxation rates associated with the calculated trajectories in the different conditions were extracted. The R_2 and nOe profiles showed similar patterns over the sequence, with the highest concentration box exhibiting similar values to the experimental dense phase. R_1 however was not well reproduced, probably because of force field and water modeling inaccuracies. Nevertheless, the Model-Free parameters in the simulations showed the same evolution than in experiments, including the amplitude distribution.

This observed amplitude redistribution is somehow related to a change in the amount of motional restriction experienced at fast and intermediary timescales. Here, a strong correlation between the sequence-specific fast motion dynamic mode amplitude and the amount of intermolecular local promiscuity between the backbone amides and other atoms is found thanks to a contact model, confirming the hypothesis that the increased intermolecular proximity is responsible for more restricted motion and therefore fast and intermediary amplitude redistribution. We should mention that the contact model presented here doesn't assume anything regarding atom groups-specific effects taking place when two atoms are close to each other. Since it only considers the mass and the distance of atoms between each others, other effects involving Van der Waals of Coulombic interactions for example are not taken into account, which can lead to discrepancies in some regions. Additionally, the model doesn't consider the increased or decreased flexibility of some regions, for example the tails of the protein chain or the serine and glycine rich regions that therefore exhibit lower order parameter values compared to the model. Nevertheless, this contact model shows a remarkable correlation between local heavy atoms density, including a contribution from intermolecular chains, and order parameters, confirming the relationship between steric effects and order parameters already observed in folded proteins.

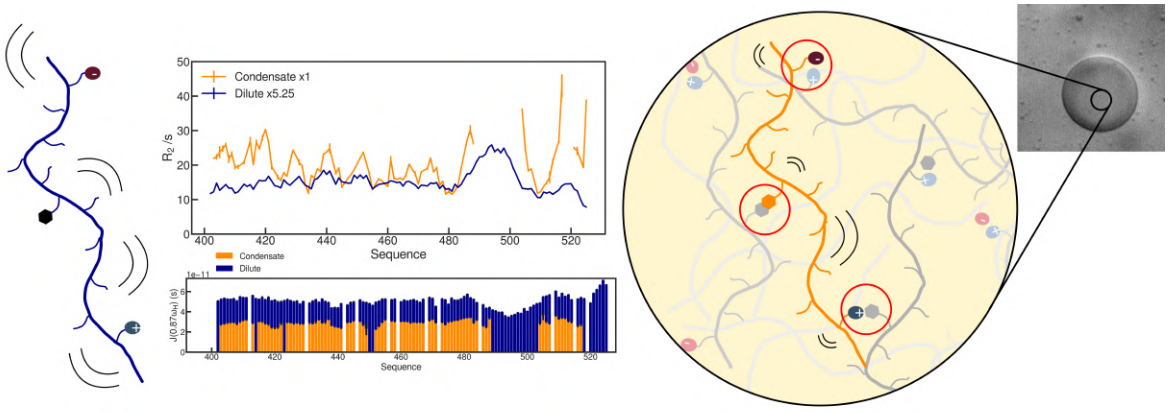
Overall, the global effect of liquid-liquid phase separation on the backbone dynamics of IDPs is now better understood. A more detailed focus on the sequence-specific contacts and their effect on backbone dynamics would be a useful addition to this work. In addition, it can be of interest to attempt a study on the dynamic properties of the IDP's side chains upon phase separation, especially the side chains that are involved in crucial intermolecular contacts like arginines and tyrosines.

3.5 Conclusion and perspective

In this work, we characterized the dynamic properties of MeV Ntail in dilute states and biomolecular condensates by NMR spectroscopy and showed that Liquid-Liquid Phase Separation induces a significant change in the dynamic properties of MeV Ntail. Notably, a redistribution of the contribution of each backbone motion timescale to the measured relaxation is observed with a greater contribution of slow motional modes in the dense phase. MD simulation reproduced this trend and exhibited a correlation between the local density of protein heavy atoms and the amplitude of the dynamic mode associated with fast motion. This study is a step forward a better understanding of the dynamic modulation of the ubiquitous phenomenon of phase separation on IDPs.

Chapter 4

Inter-molecular interactions locally slow down IDP's local backbone tumbling in biomolecular condensates of MeV Ntail



In the previous chapter, we saw that the backbone dynamics of IDPs in condensates is slowed down and that the amplitudes associated to the different dynamic modes is considerably modified due to a higher level of proximity between the atoms constituting the disordered proteins in the condensate. Many other important questions remain unanswered on how IDPs behave and how they keep the condensate stable. We saw that numerous studies have used NMR to show the presence of amino-acid specific intermolecular contacts in condensates. Whether the effect of these intermolecular interactions on IDP dynamics can be separated from the effect of pure crowding and and viscosity remains an open question. In this chapter, we further study biomolecular condensates of MeV Ntail by NMR and take a closer look on residue specific modulations of Ntail dynamics. The results of this study are in agreement with our previous work presented in the previous chapter and provides additional insight into the effect of side-chain intermolecular interactions on the backbone dynamics of IDPs.

4.1 Introduction

Numerous studies have been performed on different biomolecular condensates to characterize intermolecular interactions [264, 265, 269, 284]. The ensemble of these results show that the stabilization of a biomolecular condensate could arise from diverse and multiple factors, such as electrostatic, pi-pi, cation-pi and even hydrophobic contacts. The previous results on MeV Ntail and more specifically the salt-dependent phase diagram suggest that this MeV Ntail's phase separation is stabilized thanks to the numerous electrostatic interactions that can experience the protein [53]. Highly heterogeneous values of transverse relaxation rates over the sequence suggest the presence of specific hot spots where inter-molecular interactions take place in addition to the possible variability of the CSA tensor over the sequence. In the literature, little has been done on the characterization of the dynamics of IDPs in condensates in the light of these intermolecular interactions. Kim *et al.* observed significant increases in the ^{15}N transverse relaxation rates of CAPRIN1 in specific regions upon phase separation [284]. They could correlate R_2 in these regions, rich in arginines and tyrosines, with the presence of intermolecular interactions that are important for phase separation. Importantly, the increased transverse relaxation rates they observed could be explained by an increase in viscosity of the solution. In our previous study, we characterized a condensate of MeV Ntail and identified non-specific amplitude modulations of the backbone's dynamic modes. MeV Ntail contains a significant proportion of charged amino-acids, and only three aromatic residues.

Here, we performed further measurements of NMR spin relaxation rates on MeV Ntail in a condensed phase and further analyzed the sequence dependence of the backbone's dynamics of MeV Ntail. The obtention of robust high quality data allowed us to further characterize the sequence specific effects of biomolecular condensation on the dynamic properties of MeV Ntail. We find that the local segmental and peptidic motions in some regions of MeV Ntail are significantly slowed down compared to the rest of the sequence, while the sequence modulation of faster dynamic modes is relatively unaffected by the potential intermolecular interactions. This slow down in the segmental motion is highly correlated with the regions rich in charged residues like arginines and aromatic residues. Intermolecular NOE data suggests the presence of intermolecular contacts in the backbone while NMR exchange experiments doesn't provide any evidence for microsecond-millisecond exchange processes.

This work provides a more clear picture of the different contributions of liquid-liquid phase separation on the dynamic properties of Intrinsically Disordered Proteins and invite us to further study the dynamic properties of proteins' side chains in biomolecular condensates, where these crucial interactions take place.

4.2 Materials and Methods

Protein Preparation

MeV Ntail 402-525 The sequence of the C-terminal domain (residues 401-525) of the Measles Virus Nucleoprotein was cloned in a pET22b vector with an N-terminal 6His-tag. The gene was transformed into *Escherichia coli* Rosetta TM(IDE3)/pRARE (Novagen). The unlabeled and single labeled ^{15}N proteins were expressed in LB and M9 media respectively at 37°C until reaching a 600 nm optical density of 0.6. Then, induction of 1mM isopropyl-β-D-thiogalactopyranoside was performed before overnight incubation at 20°C. The perdeuterated double labeled $^{2}H^{13}C^{15}N$ protein was expressed in deuterated M9 medium in deuterium solution as described elsewhere [317] and induced and incubated overnight as the other labeling schemes. After cell centrifug-

ing at 5000 rpm for 20 min, the pellets were resuspended in lysis buffer (150mM NaCl, 20mM Tris-HCl, pH 8.0). Sonication was performed with 3 cycles of 11 min duration [1 sec sonication, 3 seconds delay] before centrifuging at 18000 rpm for 55min at 5°C. The supernatant was then applied in a Nickel affinity chromatography, washed with lysis buffer and lysis buffer containing 20mM imidazole and eluted with lysis buffer containing 500mM imidazole. The resulting solution was then subjected to dialysis overnight at room temperature with a 3.5 kDa dialysis membrane and finally subjected to size exclusion chromatography (SEC, Superdex75) in NMR buffer (150mM NaCl, 50mM sodium phosphate, 2mM dithiotreitol pH 6.0).

Turbidity Assay

An NMR buffer solution containing MeV Ntail was mixed with an NMR buffer solution containing PEG10000 at various concentrations in order to obtain 100 μL of NMR buffer solution containing different concentrations of MeV Ntail and PEG10000. The different mixtures were incubated 10min at room temperature. The turbidity was measured by measuring the 600 nm optical density of the different solutions with BRAND® cuvettes and a UV-VIS Spectrophotometer.

Microscopy Imaging

A 100 μL sample containing 500 μM MeV Ntail was mixed with a 100 μL sample containing 400 g/L PEG10000. The resulting mixture was imaged using DIC on an Olympus IX81 inverted microscope. An oil immersion objective 100× (Olympus UPLFLN, 1.3 NA) was used for imaging and a Hamamatsu Orca Flash4.0-V2 sCMOS (scientific Complementary Metal-Oxyde Semiconductor) 2048 × 2048 camera was used for detection.

Condensate Sample Preparation for NMR

After preparing approximately 3mL of 500 μM solutions of MeV Ntail, with a proportion of 10 or 50 % labelled protein, the protein solution is centrifuged 1 min to remove any trace of aggregation and aliquoted in several Eppendorf tubes to get 250 μL per tube. The same volume of 400 g/L PEG10000 is added to each tubes before mixing, where phase separation occurs. All the tubes are then centrifuged at room temperature and 1 rcf for 2 to 3 hours. A pellet containing the condensed phase is formed and carefully transferred to a 3 mm Shigemi tube using a glass pipette, along with a bit of dilute phase to ease the transfer and avoid having condensed phase bubbles sticking to the surface of the pipette and Shigemi tube. The Shigemi tube is then carefully manually centrifuged for a few minutes. This process is repeated until reaching an acceptable condensed phase volume in the NMR tube.

NMR spectroscopy

NMR experiments in the condensate were performed on Bruker spectrometers operating at magnetic fields corresponding to ^{1}H frequencies of 600, 700, 850 and 950 MHz equipped with cryoprobes or an Oxford-Bruker spectrometer operating at 600 MHz ^{1}H frequency equipped with a room temperature probe. All the spectra were processed using NMRPipe [318] and analyzed in Sparky.

NMR spin relaxation measurements

^{15}N longitudinal relaxation (R1), 1H-15N heteronuclear NOE (with a saturation delay of 8 sec) and R1rho (with a spin lock of 1.5kHz) were measured as described by Lakomek et al. [319] using an interscan delay of 1.2s and from 64 to 256 dummy scans. The used sets of relaxation

delays for R_1 was [0, 0.6, 0.08, 1.6, 0.4, 0.32, 0.1, 0.2, 0.6] s. For $R_{1\rho}$ in the condensed phase, the relaxation delays were [1, 15, 35, 120, 75, 22, 6, 15, 190] ms. ^{15}N transverse relaxation (R2) was determined from R1rho and R1.

Chemical exchange NMR experiments

All the chemical exchange experiments in the condensed phase were performed at a ^1H frequency of 600 MHz at 298.1 K with a room temperature probe. the CPMG relaxation dispersion experiments [87] were performed with a constant-time relaxation delay of 32 ms and CPMG frequencies between 31 and 1000 Hz. The ^{15}N DANTE-CEST experiments were performed with B_1 field strengths of 9.9, 21.5, 38.4 Hz and frequency windows of 450, 800 and 1500 Hz respectively with a constant-time relaxation delay of 450 ms. The NMR exchange data was fitted separately using the program ChemEx [320]. The off-resonance ^{15}N $R_{1\rho}$ dispersion experiments were performed on a one month old condensate with spin-lock frequencies of 800, 900, 1000, 1200 and 1500 Hz and offsets ranging between -4000 and +4000 Hz with respect to the carrier frequency, which was 117 ppm. The $R_{1\rho}$ dispersion curves were fitted to a one state model.

NOESY experiments

The ^{15}N-filtered intermolecular NOESY-HSQC experiments were performed on a 1:1 ^{15}N:^{14}N MeV Ntail condensate sample with mixing times of 100 and 250 ms at a ^1H frequency of 850 MHz at 298.1 K with a cryoprobe.

Relaxation data analysis

The relaxation data were fitted to a triple Lorentzian spectral density function expressed as:

$$J(\omega) = \sum_{k=1}^{3} \frac{A_k \tau_k}{1 + \tau_k^2 \omega^2}$$

and relaxation rates expressed as follows:

$$R_2 = \frac{1}{20}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2 (\sigma_\parallel - \sigma_\perp)^2 (4J(0) + 3J(\omega_N))$$

$$R_1 = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2 (\sigma_\parallel - \sigma_\perp)^2 J(\omega_N)$$

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N))$$

$$\eta_{xy} = \frac{1}{15} P_2(\cos\theta)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)(\sigma_\parallel - \sigma_\perp)\,\omega_N (4J(0) + 3J(\omega_N))$$

In which the CSA tensor was set to -172 ppm and the average N-H bond distance to 1.015 Angström. The reduced spectral density mapping was performed using modfree, a command-line python framework for backbone amide relaxation data analysis (Appendix). The reduced spectral density mapping technique that is used corresponds to the so-called method 3 described by Farrow *et al.* [69].

4.3 Results and discussions

4.3.1 Concentration dependence of MeV Ntail phase separation

In order to better understand the PEG and Ntail concentration dependence of phase separation, a protein-PEG concentration phase diagram was obtained by performing numerous turbidity assays at several combinations of concentrations of PEG and proteins. The result, shown in figure 4.1, defines the region in which phase separation is possible. It indicates that phase separation depends on both the PEG and the protein concentration. Phase separation occurs between concentrations of 70 to 150 g/L of PEG10000 depending on the protein's concentration. Microscopy imaging exhibits droplets of spherical shape in the conditions of the condensate NMR sample preparation (Fig. 4.1). It was previously shown by FRAP that such droplets exhibit liquid-like diffusion properties [53].

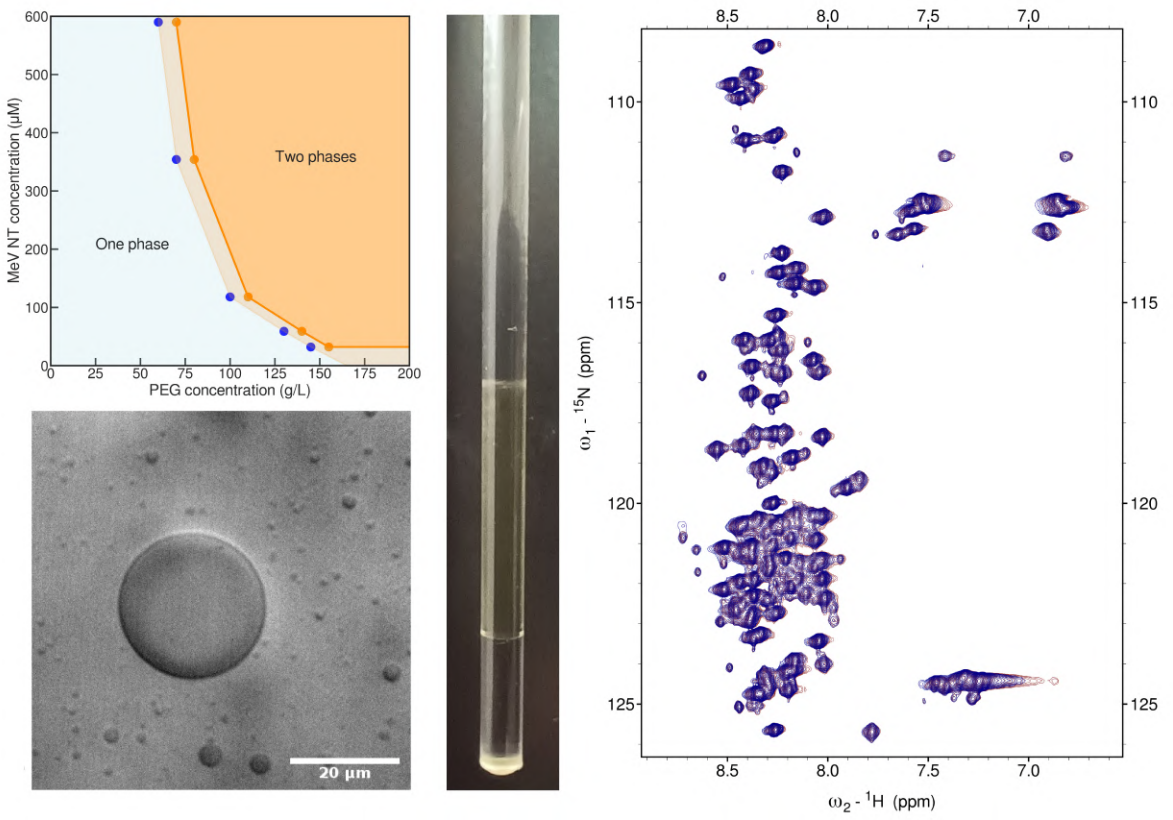


Figure 4.1: Top left: Concentration phase diagram of MeV Ntail phase separation with PEG at 298K (room temperature). Bottom left: DIC image of a MeV Ntail droplet after phase separation. Center: Image of a MeV Ntail condensed phase NMR sample. Right: Superimposed $^{1}H-^{15}N$ HSQC of MeV Ntail single labelled (red) and perdeuterated (blue) in the condensed phase (700 MHz).

Different NMR samples of MeV Ntail's dense phase were prepared with different labeling schemes. A perdeuterated sample was prepared with the hope of making the helical domain visible by quenching proton dipolar contributions to the transverse relaxation rates of amide protons. Unfortunately, the perdeuteration improved only slightly the quality of the spectra. The two overlapped HSQC spectra actually seem identical and the helical domain remains invisible (figure 4.1). NMR spin relaxation rates were measured with and without perdeuteration and showed that perdeuteration didn't significantly affect the dynamics of MeV Ntail in the condensate (figure 4.2). This repeated experiment also demonstrates the robustness of the established protocol for MeV Ntail condensate sample preparation. The present relaxation results are slightly different from the ones presented in the previous chapter. This comes from slight differences in the NMR sample preparation protocol. The physical origin of these differences is unknown so far.

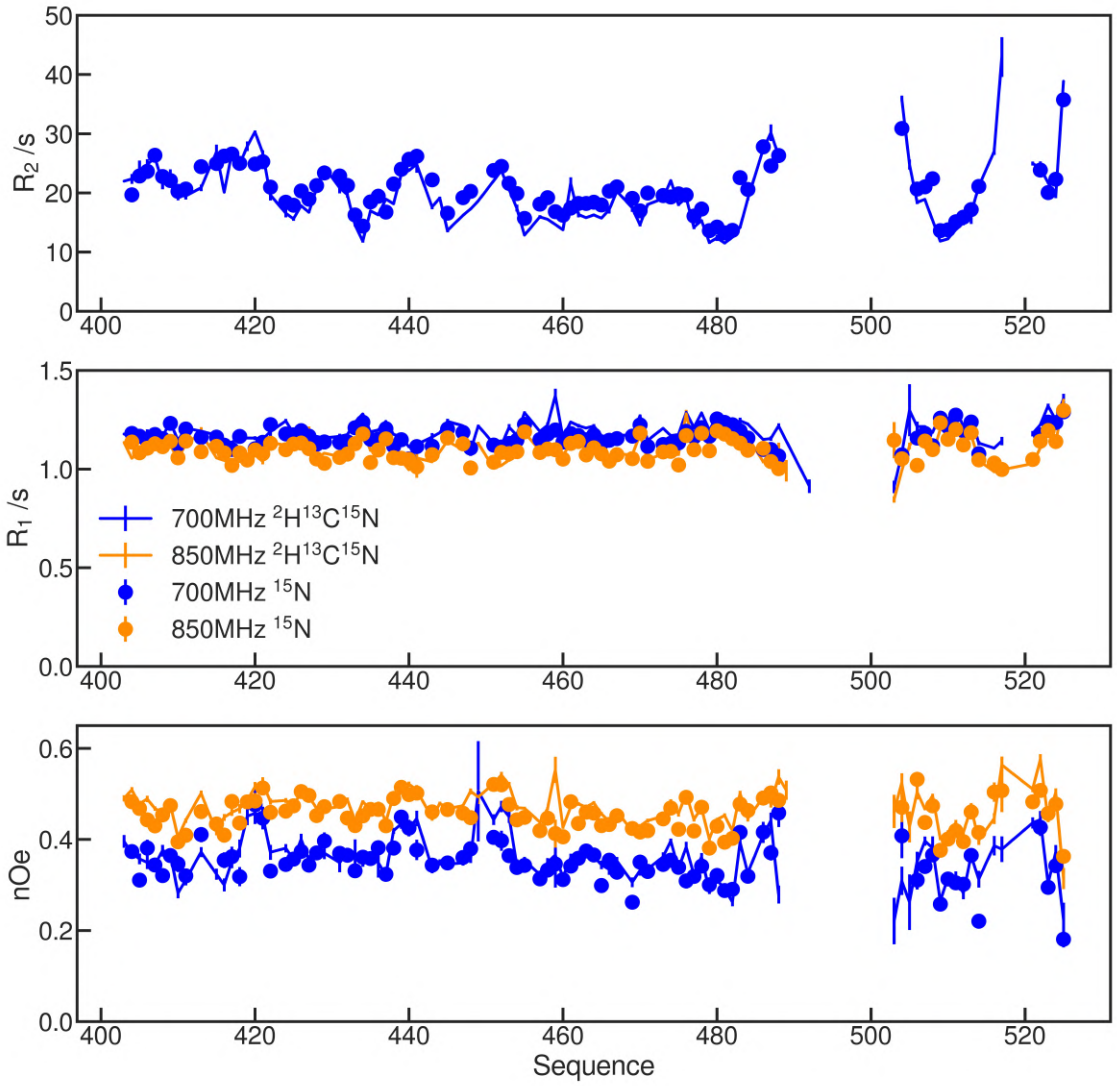


Figure 4.2: ^{15}N spin relaxation rates of MeV Ntail in the biomolecular condensate at the magnetic fields of 700MHz (Blue) and 850MHz (Orange) proton frequency. The lines correspond to the rates measured for the perdeuterated MeV Ntail condensate and the dots correspond to the single labelled MeV Ntail condensate.

4.3.2 MeV Ntail picosecond-nanosecond dynamics in the condensate

The measured longitudinal relaxation rates exhibit a relatively flat pattern over the sequence indicating a weak sequence dependence of intermediary motion around $J(\omega_N)$, except near the helical domain were R_1 seems to decrease. With the level of viscosity in biomolecular condensates, it is likely that we are in the slow timescale regime were a decrease in R_1 over the sequence indicates an increase in the rotational correlation times as the relaxation data near the helical propensity domain suggests. Heteronuclear nOe values also show a relatively flat pattern except at several regions corresponding to more rigid domains of the sequence. Overall, the protein exhibits a higher rigidity than in the dilute phase similarly to the previous observations [53]. Transverse relaxation rates oscillated around 20 per seconds with a few peaks reaching 30-40 per seconds in some regions and around the helical propensity domain. Figure 4.3 shows the R_2 profile of MeV Ntail in the condensate superimposed with the scaled R_2 profile of 300 μM MeV Ntail in NMR buffer. This figure shows that viscosity alone cannot explain the abnormally fast transverse relaxation in some regions unlike what was seen in another study [284], and that the regions where R_2 is abnormally high corresponds to regions of dense concentration of charged residues, especially arginines and lysines. It can also be noticed that the aromatic residues are systematically in these high R_2 regions. In addition, the R_2 profile of MeV Ntail condensate doesn't exhibit the typical bell shape that one would expect in such a profile for diluted proteins [315], indicating that either the condensate changes the dynamic properties of the tails of IDPs, and/or that the edges of the protein, rich in charged residues, might be involved in intermolecular contacts. An examination of the amino-acid composition of the edges of MeV Ntail suggests that this pattern comes from specific intermolecular contacts, most likely through cation-pi and electrostatic interactions, which is consistent with the observed salt dependence of phase separation of this protein, established previously [53]. Assays performed on the construct MeV Ntail 465-525 showed no phase separation, which indicates that the disordered 402-465 domain, rich in arginines and other charged residues, is important for phase separation. It is still not clear however whether the helical domain is involved in the stabilisation of the biomolecular condensate since the absence of signal in this region could indicate a very slow tumbling and/or a strong dispersion effect due to inter-molecular interactions.

Spectral density mapping confirms the global slow down of the motion of MeV Ntail in the condensate. Comparison with MeV Ntail in dilute conditions gives a quantitative comparison on the relative values of the spectral density function at different frequencies between condensed and dilute state. The spectral density function evaluated at both ω_N and $0.87\omega_H$ is lower in the condensate with respect to the dilute condition. The negligible sequence dependence in the evolution of the spectral density function at these frequencies indicates that the evolution of MeV Ntail dynamics at fast frequencies is independent of any possible sequence-specific effects, and that possible intermolecular contacts affect the slow local tumbling and chain-like motions of the protein backbone rather than faster dynamic processes in the protein's backbone. A Model-Free analysis of this data set was performed to assess the degree of reproducibility of the dynamic properties of MeV Ntail in this dense phase with respect to the previous dense phase [53]. To do so, the correlation time of the fast dynamic mode was fixed to 100 ps since this parameter is known to show limited sequence dependence [51–53], and the possible presence of exchange contribution to R_2 was neglected. Such analysis cannot provide perfectly accurate parameters if there is exchange, but it is a reasonable approximation for the purpose of a qualitative comparison. The results, shown in figure A.11, show a high degree of similarity of the Model-Free parameters between both dense phases.

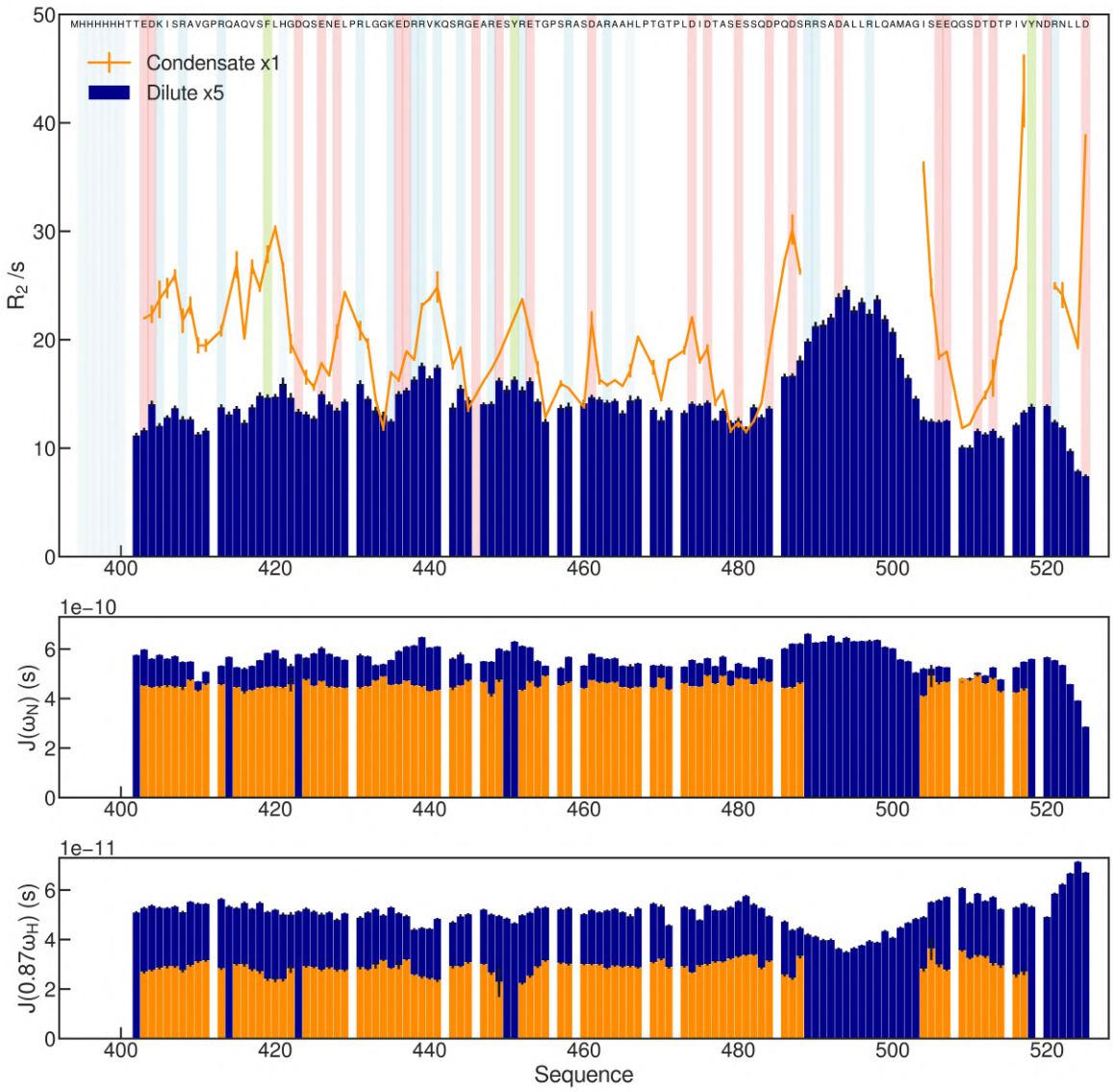


Figure 4.3: ^{15}N transverse spin relaxation rates of MeV Ntail in the biomolecular condensate (Orange lines) and in the dilute phase (Blue bars). The vertical colored bars correspond to positively charged (blue), negatively charged (red) and aromatic ring bearing (green) amino-acids. $J(\omega_N)$ and $J(0.87\omega_H)$ profiles for the condensed phase (Orange) and the dilute state (Blue) at 700MHz proton frequency.

4.3.3 MeV Ntail microsecond-millisecond dynamics in the condensate

Probing slower motion processes is of high interest in biomolecular condensates since it reports on the functional mechanisms of the studied protein in these environments. While numerous studies including this one managed to identify patterns in the sequence specific intermolecular interactions, the rate at which the different "stickers" interact is yet unknown. Here, we performed extensive chemical exchange experiments in order to explore the dynamics of MeV Ntail's backbone in the condensate at slower timescales. Carr-Purcell-Meiboom-Gill relaxation dispersion experiments as well as DANTE-CEST experiments were performed on the protein's backbone nitrogen. The results showed no significant trace of chemical exchange (Figure A.12 and A.13). These negative results are not a proof that no exchange exist, but rather that if exchange

exist, either it involves a negligible chemical shift difference and/or timescales that are inaccessible to CPMG and CEST experiments, or the exchange contribution is a relaxation exchange rather than a chemical shift driven exchange. Lewis Kay and coworkers successfully characterized relaxation exchange using off resonance $R_{1\rho}$ dispersion experiments on a condensate. It is also possible that most of the slower motion processes occur in the important side-chains including arginines and aromatics, and that the backbone only exhibits a normal liquid-like behavior where only the backbone dynamics is affected due to the global increased nanoviscosity and self-crowding. The sequence-specific slow downs in this hypothesis would be an indirect consequence of the interacting side chains attached to these regions of the protein.

Since the backbone exhibits sequence-specific modulations of its segmental motion, it can be of high interest to investigate the side chains, and characterize side chain dynamics. In the condensate of MeV Ntail, the arginine NH groups are visible by NMR. CPMG relaxation dispersion experiments were performed on the arginine NH groups region on MeV Ntail. The results in the condensate indicate a small residual chemical exchange contribution to R_2. Although some of the observed unassigned peaks could be overlapped, our observation suggest that some slow dynamics in the side chains can be probed by NMR. These preliminary results on the NH groups of the Arginine side chains invites us to study further the side chain dynamics and interactions in biomolecular condensates.

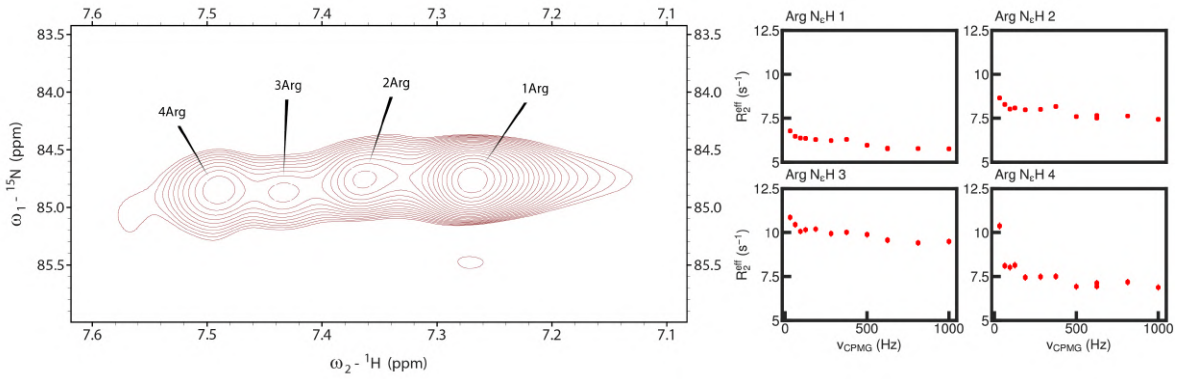


Figure 4.4: Left: Nitrogen-Proton HSQC of the Arg NH group domain. Right: CPMG relaxation dispersion profiles of the arginine NH group peaks.

4.3.4 Probing inter-molecular contacts with NOESY experiments

Numerous research groups successfully observed intermolecular contacts in biomolecular condensates using NOESY experiments. In the condensate of MeV Ntail, ^{15}N-filtered NOESY-HSQC experiments were performed with two different mixing times to probe possible intermolecular contacts. A first assay was performed in a 2D fashion and successfully showed NOE cross-peaks with a mixing time of 100 ms (figure 4.5). However, in order to obtain these cross peaks, an overweekend 2D experiment was necessary, making the extension to 3D unfeasible due to spectrometer time limitations. The inefficiency of these potential NOE transfers indicate that the dynamic-range in which MeV Ntail condensate is evolving is not optimal for intermolecular NOESYs. In addition, we cannot exclude the possibility that the observed NOE cross-peaks come from residual intra-molecular break-through. A solution to this issue would be to explore different dynamic ranges by varying temperature. However, changing the temperature of a stable macrodroplet sample can alter its stability and induce a slow transition state as the temperature-

concentration phase diagram would suggest [53]. A rigorous approach would be to separately prepare condensate samples at different temperatures, but a more economically viable approach would be to ignore this fact or try to work in a region of the phase diagram where the stable protein concentration in the droplet doesn't change significantly with temperature.

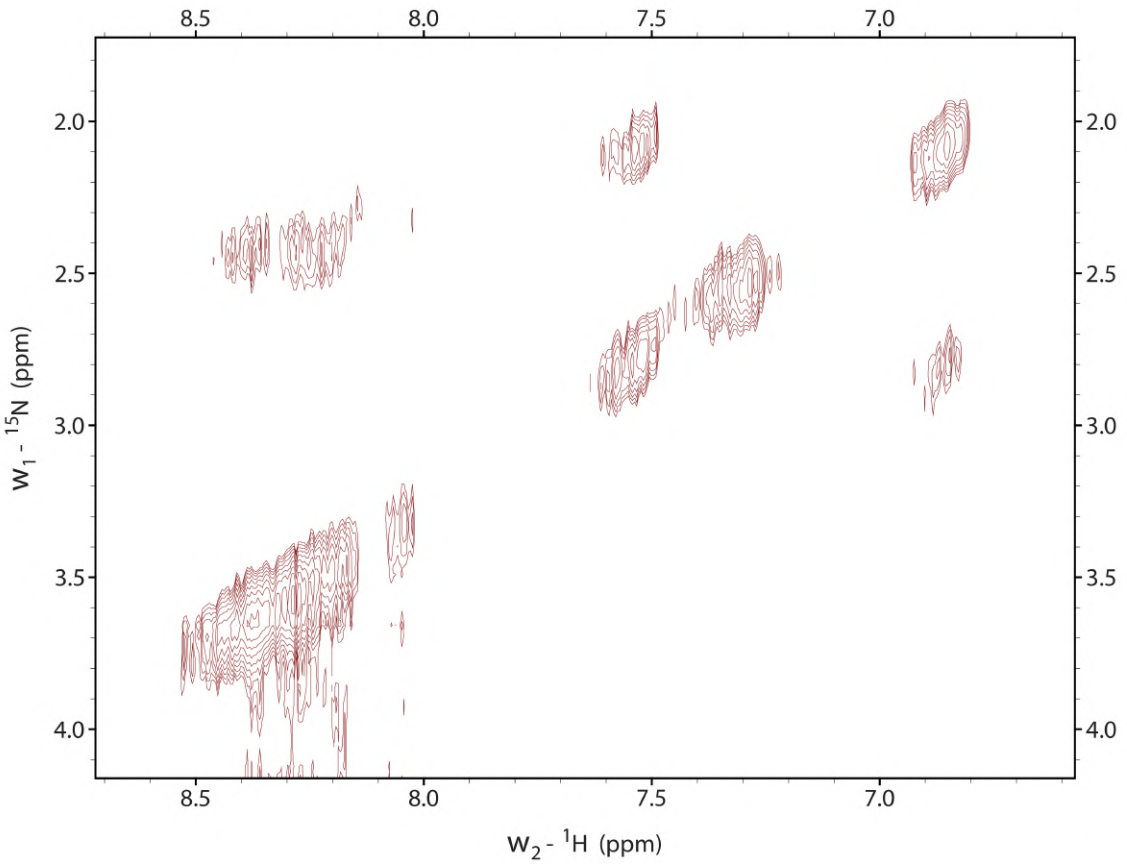


Figure 4.5: filtered NOESY-HSQC spectrum in the CH–HN intermolecular cross peak domain.

4.4 Discussion

Understanding the sequence-specific features of liquid-liquid phase separation processes in biology is important for understanding their origin and function. More specifically, understanding the sequence-dependent dynamic properties of biomolecules in such highly dynamic compartments is essential to understand how molecules behave in membrane-less organelles.

In this work, we further studied the dynamic properties of MeV Ntail upon liquid-liquid phase separation with a particular focus on sequence-specific behaviors. Robust NMR spin relaxation measurements at two magnetic fields showed a similar behavior to what was observed previously [53].

Here, the transverse relaxation profile of the dense phase show numerous regions with elevated transverse relaxation rates. The transverse relaxation rates of MeV Ntail in the dilute state were scaled up to values close to that of the dense phase. Comparison with the values corresponding

to the dense phase show that viscosity itself doesn't explain the elevated R_2 in some regions, unlike another study [284]. These abnormally high R_2 could come from direct chemical exchange contributions to R_2 and/or to local slow down of the chain segmental rotational tumbling due to intermolecular interactions taking place in the vicinity of these locations, on the side chains for example. To probe chemical exchange contributions to R_2, chemical exchange experiments were performed and no sign of significant chemical shift exchange contribution to R_2 could be observed at the probed timescales.

A reduced spectral density mapping was performed to directly look at the spectral density function evaluated at specific frequencies. The spectral density function evaluated at fast frequencies exhibited a poor sequence-specific modulation upon phase separation, suggesting that sequence-specific effects don't affect the fast motions of the backbone of MeV Ntail. Examination of the Model-Free parameters confirm that mostly the slow and more modestly the intermediary dynamic modes exhibit sequence-specific features that differ between dilute and dense phase. Such observation suggest that sequence-specific effects, probably intermolecular contacts involving Coulombic and cation-pi interactions, slow down the local backbone rotational tumbling of MeV Ntail in the dense phase.

Evidence for intermolecular interactions require extensive measurements of inter-molecular NOESY experiments. preliminary intermolecular NOESY measurements between NH and CH moieties suggests the presence of such contacts. However, further measurements are necessary at more favorable conditions in order to quantitatively characterize the sequence-specificity of these contacts. Alternatively, or in complement, Coarse-Grained simulations with a Hydrophobicity Scaled (HPS) model should provide a reasonable idea of where we could expect such contacts along the sequence, although some types of interactions are more difficult to simulate and high resolution and accuracy is often necessary to grasp some subtleties including cation-pi interactions [289].

These observations all-together allow us to propose the following model for MeV Ntail's behavior in its dense phase. First, upon phase separation, the rotational correlation times associated with the backbone chain are considerably slowed down because of the increased length-scale-dependent viscosity. Then, the contribution of the different dynamic modes corresponding to different timescales of motion are considerably redistributed, manifesting significantly more restricted fast dynamic processes because of the high level of promiscuity of the different polymer chains as shown previously [53]. Finally, the inter-molecular interactions that stabilize the dense phase, principally taking place in specific areas of the primary sequence, here in charged and aromatic bearing regions, have the effect of further slowing down the chain-like segmental motion and modestly the peptide plane fluctuations of the backbone in the vicinity of these interactions, contrasting with more dynamic Serine and Glycine-rich flexible regions. Faster motional processes are negligibly affected by sequence-specific effects.

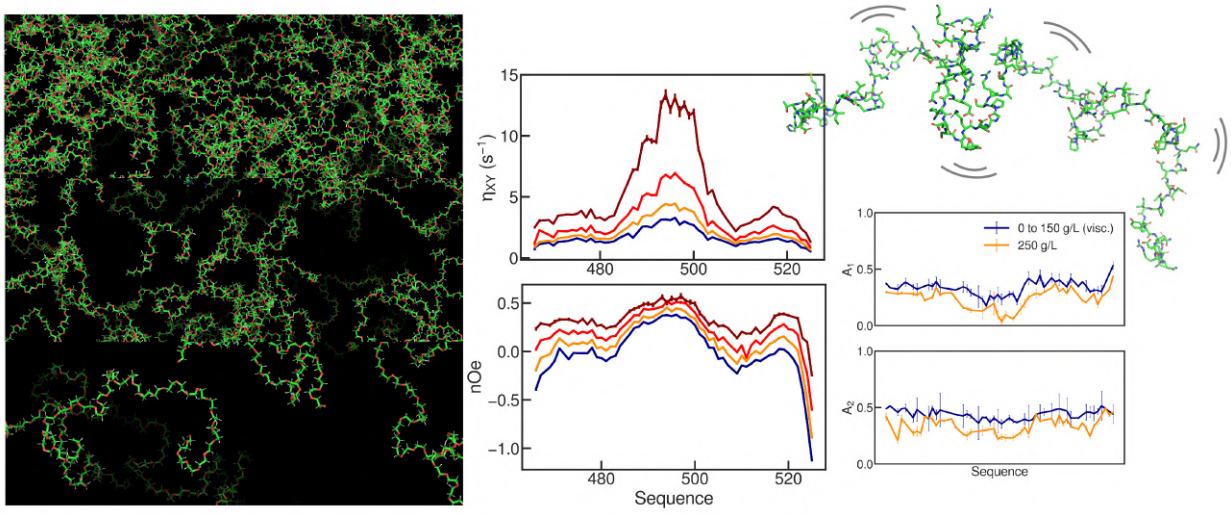
Every condensate is different, by their dynamic range, their viscosity, their amount of inter-molecular contacts, and probably many other factors including the concentration and the environment in which they formed. Therefore, further studies on droplets of different constructs are necessary to allow meta-analyses that could provide insight into possible general rules on how dynamics in the condensate is determined. Our condensate is relatively viscous and dense compared to what was observed in previous works. This system is ideal to study the effect of high levels of crowding, but lowering the viscosity would be ideal to allow the observation of more peaks, especially the helical propensity domain. Reducing the amount of inter-molecular contacts by mutations could theoretically reduce the viscosity of the condensate. However, it would also affect its stability in a way that is still difficult to predict.

4.5 Conclusion

Here, we further characterized the dynamic properties of MeV Ntail upon liquid-liquid phase separation and analyzed its sequence specificity. Our results confirm what were previously observed in our previous study, namely the global slow down in rotational correlation times and the redistribution of the different dynamic modes associated with the backbone's motion at multiple timescales. Finally, we identified the sequence-specific behavior of MeV Ntail and propose that Coulombic, cation-pi and pi-pi intermolecular interactions, responsible for the condense phase stabilization, are responsible for a local slow down in MeV Ntail's segmental and peptidic tumbling. Further studies in the side chains should enlight these intermolecular interactions, and in the best case allow to characterize the rate of inverconversion between bound and free states in the side chains important moieties, providing further insight into the biologically crucial phenomena of liquid-liquid phase separation.

Chapter 5

Polymer crowding modifies the dynamic properties of MeV Ntail 465-525



It was observed that the self-crowded condensed phase modified the dynamic properties of MeV Ntail, first because of the increased proximity of the polypeptides, then because of sequence-specific contacts. Is the non-specific dynamic change observed in the previous chapter specific to liquid-liquid phase separation ? The effect of viscosity induced by polymer crowders on the backbone dynamics of IDP has been already well established by previous studies. In this chapter, the effect of extreme crowding is studied with the use of a polymer crowder: PEG10000. Extensive NMR spin relaxation measurements showed that an amplitude modulation can be observed even outside of a biomolecular condensate if the level of crowding is sufficiently high, suggesting that polymer crowding modifies the dynamics of MeV Ntail 465-525 regardless of phase-separation. We propose several possible explanations to this observation in the light of polymer theory.

5.1 Introduction

In the previous chapters, we saw that liquid-liquid phase separation strongly affected the dynamic properties of Intrinsically Disordered Proteins, with both uniform and sequence dependent contributions. It was shown that an increased level of inter-chain proximity manifested by high levels of interatomic contact in MD simulations was correlated with a uniform redistribution of the dynamic modes of MeV Ntail's backbone in addition to a global increase of the model-free rotational correlation times. In the last chapter, it was shown that sequence-specific slow down in the segmental dynamics of the backbone of MeV Ntail was correlated with regions rich in residues known to be responsible for promoting and stabilizing liquid-liquid phase separation.

Biomolecular condensates are highly concentrated environments of proteins and can be categorized as a crowded environment for this reason. The non specific redistribution of MeV Ntail's dynamic modes were shown to be correlated with the reduced intermolecular distances between the different protein atoms. It is not difficult to make a parallel with other crowded environments like what we can find in cells. Crowded environments are all characterized by the presence of large amounts of macromolecules with volumic proportions reaching relatively high values. More specifically, an IDP in a biomolecular condensates can be considered to be in a complex heteropolymer crowder with both crowding and soft interaction effects that are specific to the condensate.

The theory of polymer crowding is already well understood thanks to pioneer work from numerous scientists including P.J. Flory and P.G. de Gennes [228, 321] and has been discussed in chapter 2. Briefly, considering a polymer crowding species, there are three different regimes in which the solution can exist. The dilute regime corresponds to low concentrations of crowding in which each polymer chain is free to diffuse. The semi-dilute regime corresponds to relatively higher concentration of crowder in which the hydrodynamic spheres of each polymer coil starts to overlap. Finally, the concentrated regime corresponds to a high concentration condition in which a high level of entanglement and proximity among the polymer chains is present and the concept of hydrodynamic sphere becomes irrelevant due to the high polymer density in the mixture.

Crowded environments were extensively studied with fluorescence methods, providing insightful low-resolution information on the diffusion and rearrangement of the studied IDPs [238]. It was shown that increased viscosity in crowded solutions resulted in length-scale dependent slow down of both the diffusion and rearrangement of the IDP's global conformations. NMR also provided extremely detailed insight into the backbone dynamics of IDPs at atomic resolution, as a relationship between nanoviscosity induced by crowding and the rotational correlation times of the backbone could be empirically established in semi-dilute regimes [52], as discussed in chapter 1 and 2.

Despite all this progress, it must be recalled that crowding theoretically doesn't induce only increased viscosity. We propose the hypothesis that at sufficiently high amount of polymer crowding agents, the high level of chain promiscuity should also result in more restricted motions which should change the dynamic properties of IDPs as seen by NMR spin relaxation in biomolecular condensates. Additionally, the possibility to reproduce the crowding conditions of biomolecular condensates with polymer crowding would allow a more extensive study of the protein's behavior in these conditions including temperature dependence since no thermodynamic phase equilibrium is perturbed upon temperature change in such conditions.

Here, to further improve our understanding of the dynamics of IDPs in polymer crowded solutions, we perform extensive NMR spin relaxation on MeV Ntail 465-525 (short Ntail or sNtail), a construct of MeV Ntail 402-525 that doesn't phase separate upon addition of PEG. We measure NMR spin relaxation at different concentrations of PEG10000 up to 250 g/L, which is the highest crowding concentration measured so far, comparable to what can be found in some cells in terms of mass concentration of macromolecules as well as some biomolecular condensates. The results show that the relaxation rates of MeV Ntail 465-525 can be predicted using the previously established viscosity-dependent model-free analysis in the semi-dilute regime up to 150 g/L of PEG. Then, extreme crowding shows a uniform non negligible redistribution of the dynamic modes of the backbone of MeV Ntail 465-525 across the sequence, akin to the non specific modulations in the biomolecular condensate of MeV Ntail. We rationalize this finding by simulating boxes of concentrated PEG solutions. We propose that 250 g/L of PEG corresponds to a concentrated regime where the non specific effects of chain contacts affect the internal dynamic properties of MeV Ntail 465-525.

We observe for the first time these effects on the internal dynamics of an intrinsically disordered protein, and show that IDP's behavior can be highly modulated by the macromolecular concentration in concentrated environments, even outside liquid-liquid phase separated systems.

5.2 Materials and Methods

Protein Preparation

MeV Ntail 465-525 The sequence of the short C-terminal domain (residues 465-525) of the Measles Virus Nucleoprotein was cloned in a pET22b vector with an N-terminal 6His-tag and a TEV cleavage site. The gene was transformed into *Escherichia coli* BL21 (Novagen). The unlabelled, single labelled ^{15}N and double labelled $^{13}C^{15}N$ proteins were expressed in LB and M9 media respectively at 37°C until reaching a 600 nm optical density of 0.6. Then, induction of 1mM isopropyl-β-D-thiogalactopyranoside was performed before overnight incubation at 20°C. After cell centrifuging at 5000 rpm for 20 min, the pellets were resuspended in lysis buffer (150mM NaCl, 20mM Tris-HCl, pH 8.0). Sonication was performed with 3 cycles of 11 min duration [1 sec sonication, 3 seconds delay] before centrifuging at 18000 rpm for 55min at 5°C. The supernatant was then applied in a Nickel affinity chromatography, washed with lysis buffer and lysis buffer containing 20mM imidazole and eluted with lysis buffer containing 500mM imidazole. The resulting solution was then subjected to dialysis overnight at room temperature along with TEV protease with a 3.5 kDa dialysis membrane and finally subjected to size exclusion chromatography (SEC, Superdex75) in NMR buffer (150mM NaCl, 50mM sodium phosphate, 2mM dithiotreitol pH 6.0).

NMR spectroscopy

NMR experiments on Ntail 465-525 were performed on Bruker spectrometers operating at magnetic fields corresponding to ^{1}H frequencies of 600, 700, 850 and 950 MHz. The protein NMR buffer contained 0, 75, 150 or 250 g/L PEG10000 and the measurements were performed at 298.1K, 288.1K or 278.1K. The protein concentration was 300μM unless stated otherwise. All the spectra were processed using NMRPipe [318] and analyzed with Sparky.

NMR assignment and chemical shifts

The assignment of the chemical shifts of MeV Ntail 465-525 was performed with state of the art BEST-HNCO, BEST-HN(CA)CO, BEST-HNCA and BEST-HN(CO)CA experiments in the dilute state. BEST-HNCA and BEST-HNCO experiments were also performed at 75, 150 and 250 g/L of PEG to probe carbon chemical shifts as a function of PEG concentration. Secondary structure propensities were extracted with the program SSP as described by the authors [111].

NMR spin relaxation measurements

^{15}N longitudinal relaxation (R1), 1H-15N heteronuclear NOE (with a saturation delay of 8 sec) and R1rho (with a spin lock of 1.5kHz) were measured as described by Lakomek et al. [319] using an interscan delay of 1.2s and from 64 to 256 dummy scans. The used sets of relaxation delays for R_1 was [0, 0.6, 0.08, 1.6, 0.4, 0.32, 0.1, 0.2, 0.6] s. For $R_{1\rho}$, the delay list varied depending on the level of crowding and on temperature and was systematically composed of 9 to 10 delays between 0 and up to 230 ms. ^{15}N transverse relaxation (R2) was determined from R1rho and R1.

Relaxation data analysis

The model-free analyses were performed with modfree (Annex A). The relaxation data were fitted to a triple Lorentzian spectral density function expressed as:

$$J(\omega) = \sum_{k=1}^{3} \frac{A_k \tau_k}{1 + \tau_k^2 \omega^2}$$

and relaxation rates expressed as follows:

$$R_2 = \frac{1}{20}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2 (\sigma_\parallel - \sigma_\perp)^2 (4J(0) + 3J(\omega_N))$$

$$R_1 = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2 (\sigma_\parallel - \sigma_\perp)^2 J(\omega_N)$$

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N))$$

$$\eta_{xy} = \frac{1}{15} P_2(\cos\theta)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)(\sigma_\parallel - \sigma_\perp)\,\omega_N (4J(0) + 3J(\omega_N))$$

In which the CSA tensor was set to -172 ppm and the average N-H bond distance to 1.015 Angström. The DD/CSA angle was fitted simultaneously with the model-free parameters.

Temperature-dependent model-free analyses were performed on the relaxation datasets at each PEG concentrations with the following formula for the correlation time:

$$\tau_k(T) = \tau_{k,\infty} exp\left(\frac{E_{a,k}}{RT}\right)$$

Where $E_{a,k}$ is a residue-specific activation energy expressed in J/mol, R is the ideal gas constant and T is temperature in K. $\tau_{k,\infty}$ is the correlation time at infinite temperature. The fast correlation time τ_1 was fixed to 45 ps and $E_{a,1}$ was fixed to 0.

Viscosity-dependent model-free analyses were performed on the relaxation datasets at 298.1 K from 0 up to 150 or 250 g/L PEG with the following formula for the correlation time:

$$\tau_k = \tau_{k,0}(1 + \varepsilon_k \rho)$$

Were ε_k is a residue-specific solvent friction coefficient, ρ is the solvent friction and $\tau_{k,0}$ is the reference correlation time at 0 g/L PEG.

Modeling MeV Ntail 465-525 in PEG solutions

A statistical coil ensemble of PEG molecules was generated with an home-made PEG statistical-coil generator. Briefly, each backbone atom of the PEG molecule is generated, followed by eventual hydrogen atoms. The geometry of the atoms with respect to each other was taken from the CHARMM36 force field equilibrium parameters. The dihedral angles were sampled from the dihedral angle potentials taken from the CHARMM36 force field. Clashes were avoided by applying a distance cutoff with the surrounding atoms. An ensemble of thousands of conformations was generated this way.

Boxes of MeV Ntail 465-525 were generated using GROMACS. a given number of PEG molecules was then inserted in the box so that the final concentration is 75, 150 or 250 g/L. The generation was performed in a similar way to what was described previously [53].

5.3 Results and discussions

5.3.1 Impact of crowding on the local conformational sampling

To evaluate the effect of crowding on the local conformational sampling of MeV Ntail 465-525, chemical shift titrations were performed between 0 g/L and 250 g/L of PEG. $^{1}H\ -^{15}\ N$ HSQCs exhibit a non specific shift in the proton and nitrogen chemical shifts indicating either a change in the properties of the surrounding solvent, non specific interactions, or a change in the conformational sampling. To probe such possible changes, carbon chemical shifts were measured with 3D backbone correlation experiments and the chemical shift indexes and secondary structure propensities were calculated. The obtained carbon chemical shifts negligibly change with PEG concentration, indicating a stable secondary structure propensity (Fig. 5.1). A slight but almost negligible increase in the helical propensity can be observed, indicating a possible stabilization

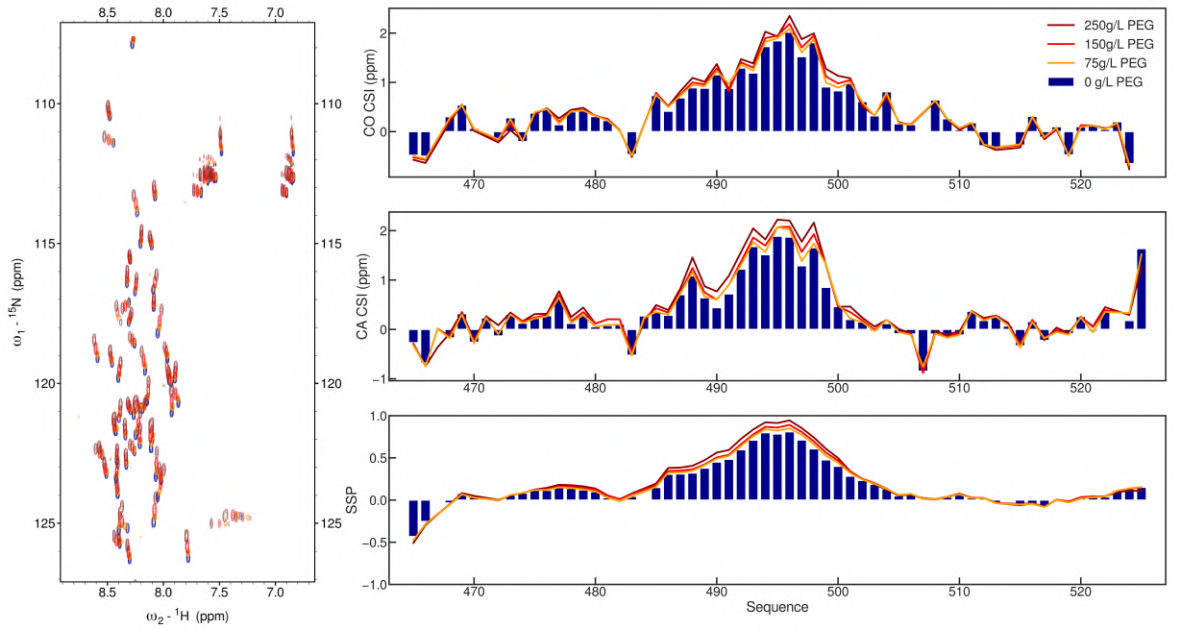


Figure 5.1: Impact of crowding on MeV Ntail 465-525 chemical shifts. Left: ^{1}N-^{15}N HSQC spectra at different concentrations of PEG: 0 (Blue), 75 (Orange), 150 (Red) and 250 g/L (Darkred). Right: Chemical shift index of carbon CO (Top) and $C\alpha$ (Middle) and resulting secondary structure propensity (Bottom) at 0 (Blue bars), 75 (Orange lines), 150 (Res lines) and 250 (Dark red lines) g/L of PEG.

of the helix upon addition of PEG, however it doesn't significantly alter the local conformational sampling of the IDP.

5.3.2 R_1 of water protons and solvent friction in crowded samples

The knowledge of viscosity in the solution can provide interesting insight in the IDP dynamics from 10 ps to 10 ns. Adamski and coworkers studied the evolution of rotational correlation times extracted from model free analysis and found that rotational correlation times could be modeled with a linear relationship with respect to solvent friction, defined as:

$$\rho = \frac{\eta - \eta_0}{\eta_0} = \frac{R_1^{water} - R_{1,0}^{water}}{R_{1,0}^{water}}$$

Where η and η_0 are the nanoviscosity of the crowded and dilute solutions respectively and R_1^{water} and $R_{1,0}^{water}$ the respective longitudinal relaxation rates of water protons in the solutions. To probe solvent nanoviscosity as a function of the concentration of our crowder, the longitudinal relaxation of water protons was measured at different magnetic fields, concentrations of PEG and temperatures. The results, shown in figure 5.2, show that R_1 of water protons is field independent.

The data at both 298.1 and 278.1 K at 950 MHz were used to calculate the corresponding solvent frictions for each PEG concentrations. The calculations at the two temperatures yielded similar results (Figure 5.3), confirming that the evolution of solvent friction with respect to PEG concentration is temperature independent, and that the data are consistent.

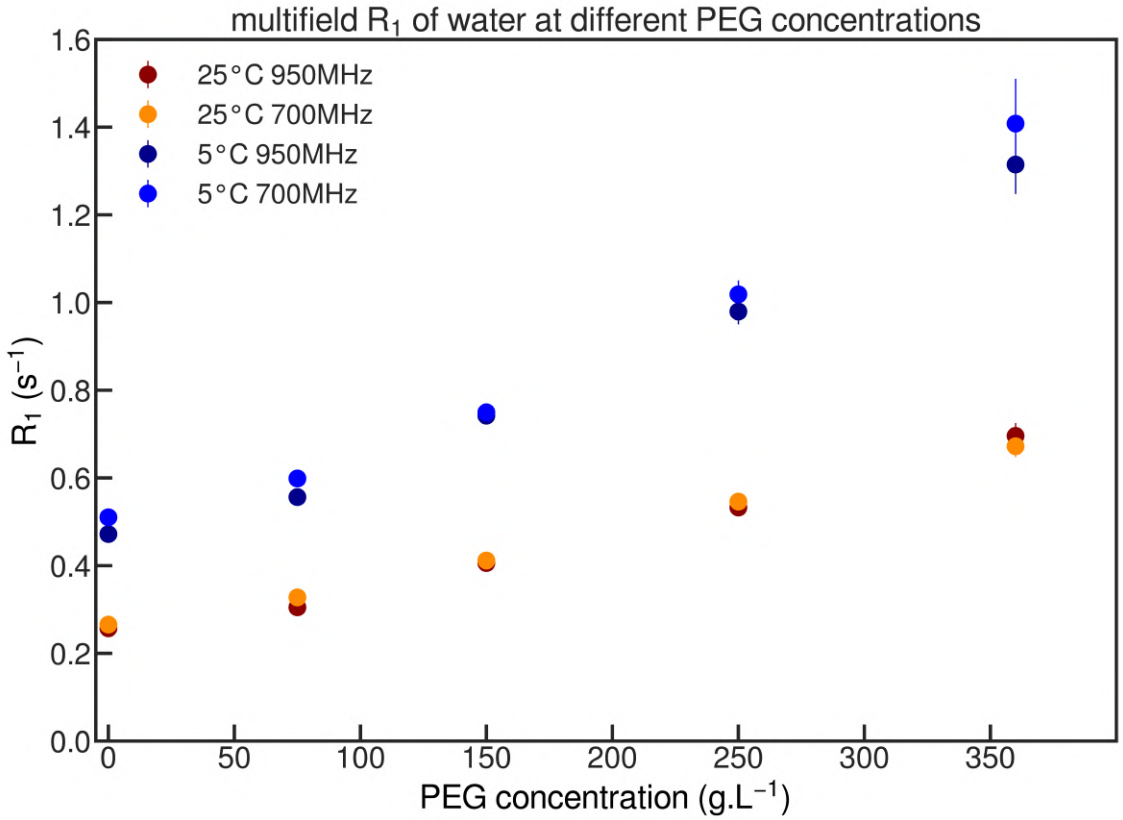


Figure 5.2: Multifield longitudinal relaxation rates of water protons in 300 μM MeV Ntail 465-525 solutions at 298.1 K (Orange: 700 MHz, Red 950 Mhz) and 278.1 K (Blue: 700 MHz, Darkblue: 950 MHz). The measures performed at 360 g/L PEG were made in the absence of protein.

5.3.3 Backbone ^{15}N Relaxation rates of MeV Ntail 465-525 at different concentrations of PEG10000

To characterize the dynamics of MeV Ntail 465-525, extensive NMR spin relaxation rates were measured at several concentrations of PEG from 0 to 250 g/L, and at different temperatures at 278.1, 278.1 and 298.1 K. This extensive amount of data will allow us to analyze the effect of polymer crowding on the dynamics of our protein and eventually obtain insight into the contribution of crowding on the protein's dynamic compared with what is observed in biomolecular condensates. Regarding Ntail 465-525 in the dilute state at 298.1 K, the R_2 and η_{XY} profiles exhibit a characteristic bell shape with a visible increase in the helical propensity domain indicating an increase in rotational correlation time in this area (Figure 5.4). The nOe profile gives a similar pattern with a visible bump in the helical domain, manifesting a smaller contribution of fast motions to relaxation in the helix.

Relaxation rates were measured in very demanding conditions, up to 250 g/L PEG and down to 278.1 K, where the observed relaxation rates are comparable to what is observed in biomolecular condensates. Some transverse relaxation rates were measured up to more than 60 s^{-1} (Figure 5.4). The viscosity dependence at 298.1 K shows the obvious increase in the rotational correlation times as manifested by an increase in R_2 and η_{XY} with respect to PEG concentration. In addition, the protein's rigidity seems to increase as well as shown by the evolution of the nOe values. Interestingly, the evolution regime of R_1 with respect to correlation times varies over the sequence. Indeed, the tails of the protein seem to be in the "low τ_c" regime where R_1 increases

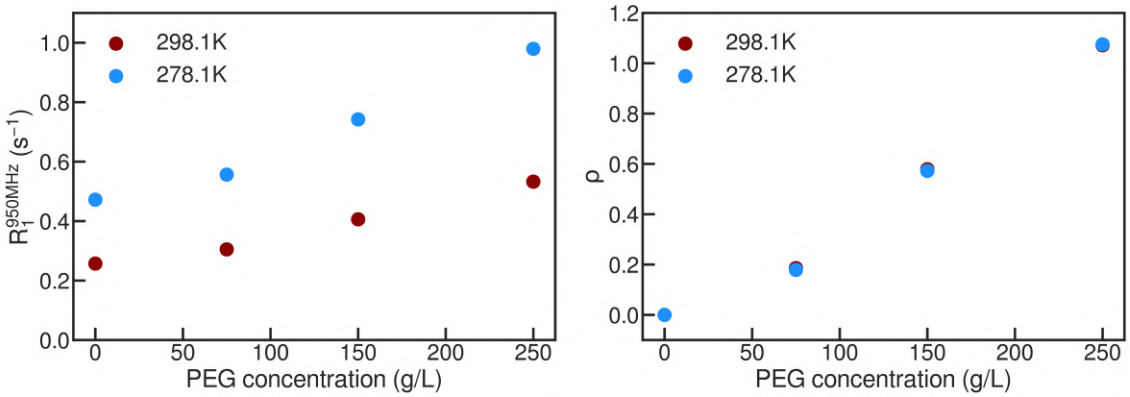


Figure 5.3: Left: Longitudinal relaxation of water protons measured at 950 MHz proton frequency at 278.1 K (Blue) and 298.1 K (Red). Right: Associated solvent frictions for 278.1 K (Blue) and 298.1 K (Red).

600MHz	0 g/L	75 g/L	150 g/L	250 g/L
298.1 K	R_1, R_2, nOe	R_1	R_1, R_2, η_{XY}	
288.1 K			R_1, R_2	R_1, nOe
278.1 K	nOe		R_1, R_2, η_{XY}	
700MHz	0 g/L	75 g/L	150 g/L	250 g/L
298.1 K	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe, η_{XY}
288.1 K	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe	R_1, R_2, nOe, η_{XY}
278.1 K	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe, η_{XY}	R_1, R_2, η_{XY}	R_1, R_2, nOe
850MHz	0 g/L	75 g/L	150 g/L	250 g/L
298.1 K			R_1, η_{XY}	
950MHz	0 g/L	75 g/L	150 g/L	250 g/L
298.1 K	R_1, R_2, nOe	R_1, R_2, nOe, η_{XY}	R_1, R_2, nOe	R_1, R_2, nOe, η_{XY}
288.1 K			R_1, R_2, nOe	R_1, R_2
278.1 K			R_1, R_2, nOe	R_1, R_2, nOe

Table 5.1: Measured and analyzed NMR ^{15}N spin relaxation rates of MeV Ntail 465-525 in different conditions

with the slow down of rotational tumbling, while towards the helix, this regime is inverted since the rotational tumbling is intrinsically slower there because of the alpha helix. at 278.1 K however, R_1 seems to decrease with the slow down of tumbling indicating a much slower tumbling regime. In these conditions at high PEG concentration, nOe values are close to saturation, from 0.4 at 700 MHz proton frequency to more than 0.6 at 950 MHz on average (Figure 5.4).

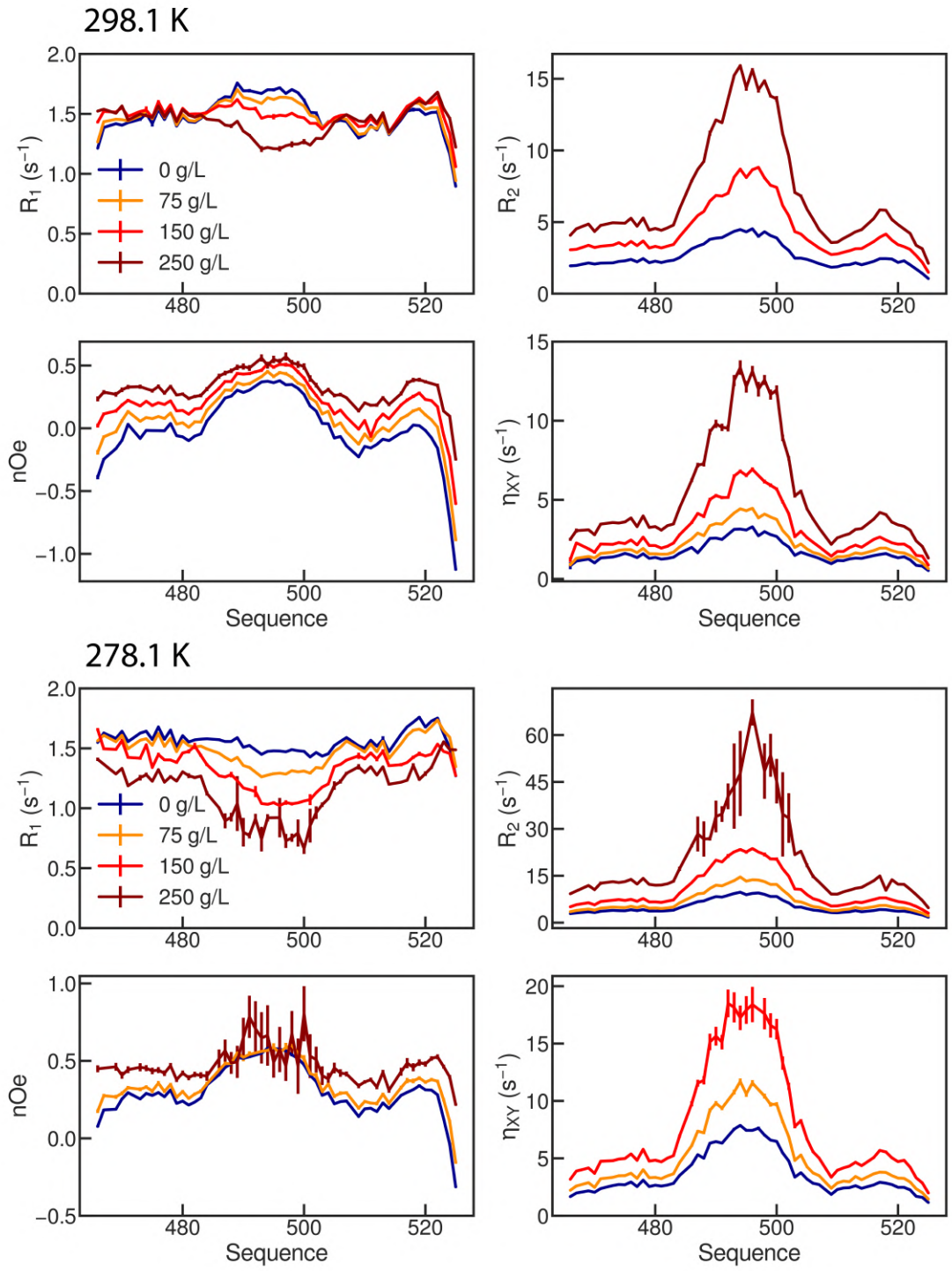


Figure 5.4: Relaxation rates of MeV Ntail 465-525 at 298.1 K (Top four panels) and 278.1 K (Down) in the dilute state (blue), with 75 g/L PEG (orange), with 150 g/L PEG (red) and with 250 g/L PEG (dark red) at 700 MHz proton frequency.

5.3.4 Model-free analysis of MeV sNtail relaxation rates

It was previously shown that with the knowledge of the crowder concentration-dependent solvent friction values, it was possible to model the correlation time with a linear relationship with respect to solvent friction using the following relationship [52]:

$$\tau_k = \tau_{k,0}(1 + \varepsilon_k \rho)$$

All the relaxation data at 298.1 K was analyzed with a model-free analysis using this relationship. The relaxation data from 0 to 150 g/L PEG fit perfectly together (Figure A.14). The extracted sequence-dependent solvent friction coefficients exhibit the expected profile, with a more important viscosity dependence of the slow dynamic mode with respect to the intermediary timescale motion (Figure A.15). However, including the highest crowder concentration condition results in a dramatic reduction of the quality of the fit, as shown in figure 5.5.

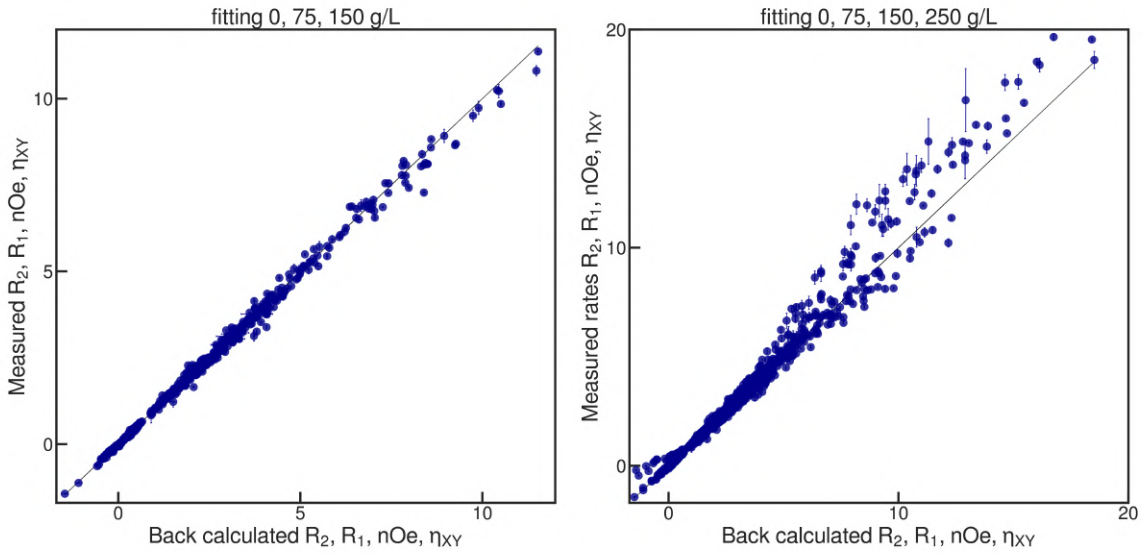


Figure 5.5: Correlation plot of the experimental relaxation rates against the calculated rates from the model-free analysis including relaxation data from 0 to 150 g/L PEG (left) and from 0 to 250 g/L PEG (right).

This observation suggests that above a certain threshold of crowding, the high field relaxation rates cannot be accurately modeled by taking only the change in the nanoviscosity into account [52]. This suggests that another contribution is significantly important to invalidate this model at 250 g/L PEG. To make sure that this doesn't come from an incorrect solvent friction estimation, the model-free analysis was performed with different values for the solvent friction associated with 250 g/L PEG. The resulting evolution of the reduced χ^2 as a function of this solvent friction value is plotted in figure 5.6 for randomly selected residues. The results show no significant improvement in the reduced χ^2 with a different solvent friction value, confirming that another parameter than nanoviscosity modifies the dynamics of MeV Ntail 465-525 at high crowding concentration.

In order to better understand how the dynamic properties of MeV Nail 465-525 evolve with

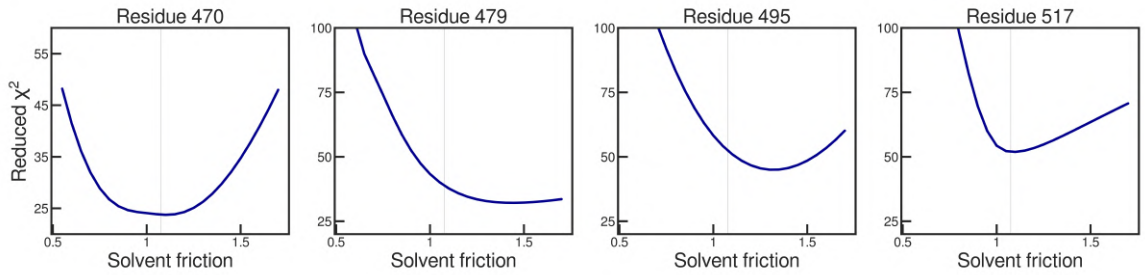


Figure 5.6: Reduced χ^2 from extended model-free analysis with viscosity model from 0 to 250 g/L PEG as a function of the solvent friction set for the 250 g/L PEG condition data set for four randomly selected residues.

crowding, the relaxation data in each PEG concentrations was separately analyzed. As shown in other studies [51, 52], the complementary relaxation data measured at the three temperatures allow a more robust analysis and reduces possible solution degeneracy arising in conventional model-free analyses because of R_1's double solutions as a function of the rotational correlation times. The degeneracy isn't a problem anymore with multiple temperature data because the evolution of R_1 as a function of temperature is providing enough information to decipher the correct R_1 evolution regime, allowing a more robust fit [51]. The fit for each conditions separately yields good agreement with the experimental rates as shown in figure 5.7 (See also figures A.16, A.17, A.18 and A.19).

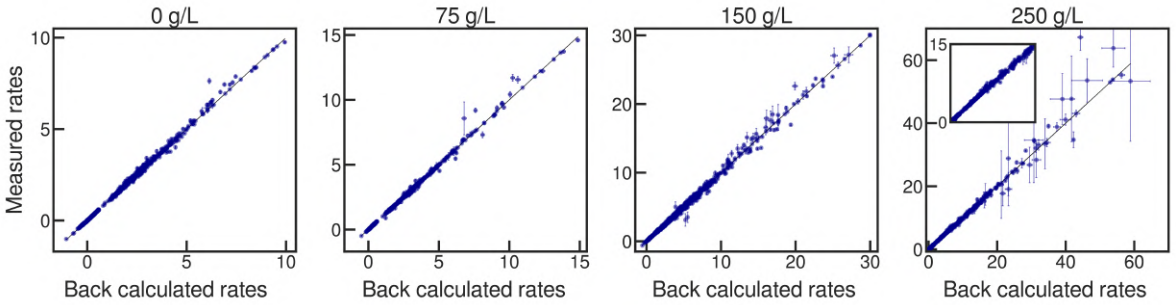


Figure 5.7: Correlation plot of the experimental relaxation rates against the calculated rates from the model-free analysis including relaxation data at three temperatures at 0, 75, 150 and 250 g/L PEG respectively from left to right.

The model-free analysis parameters as a function of PEG concentration are shown in figure 5.8.

The results show the expected increase in the rotational correlation times as a function of crowding, with a more significant dependence for the slowest dynamic mode, associated with local segmental tumbling [52]. In addition, we observe a significant evolution of the amplitudes of each dynamic mode at the highest crowding condition. While from 0 to 150 g/L, assuming constant amplitudes for each dynamic mode is a good approximation, the amplitudes at 250 g/L are significantly different with respect to 0 g/L of PEG. In fact, the model-free amplitudes evolve in the same direction as the amplitudes extracted in biomolecular condensates in chapter 3 and 4. We observe a net decrease in the fast dynamic mode amplitude at 250 g/L of PEG, which almost tends to zero (Figure 5.8 and A.20). Such observation implies that the contribution of fast motions to relaxation becomes almost negligible in such high level of crowding. This also suggests that a simpler model with two dynamic modes could be used for the analysis at this concentration. To

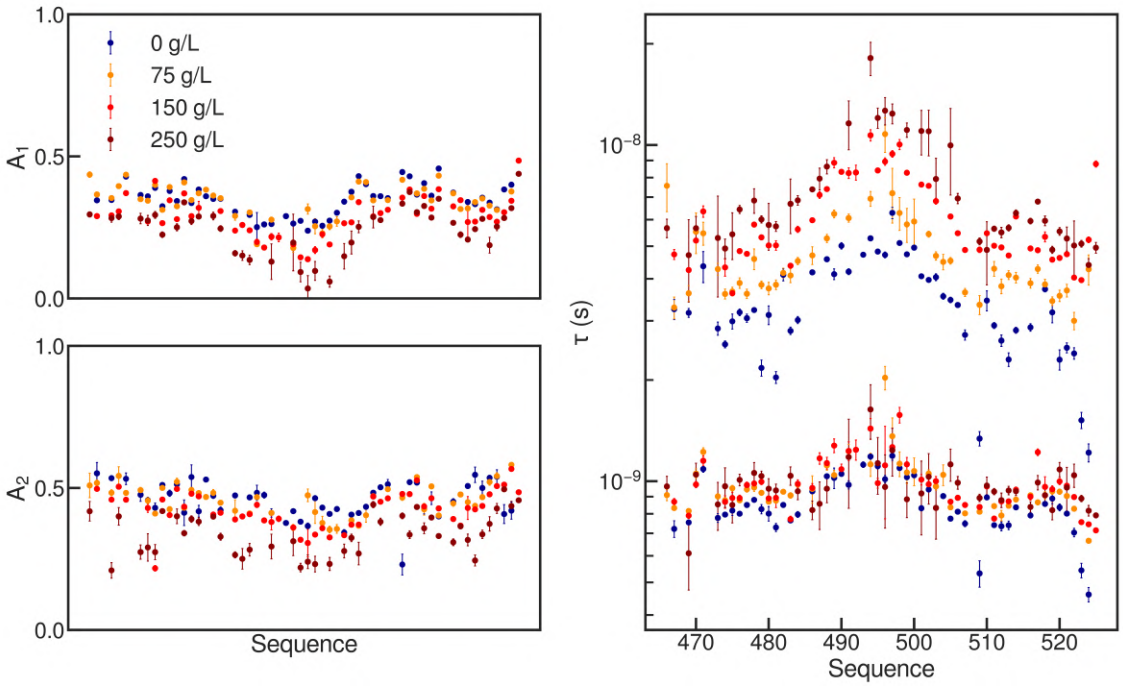


Figure 5.8: crowding dependence of the model-free parameters: Amplitudes of the first and second dynamic modes at 298.1 K (left) and intermediary and slow rotational correlation times at 298.1 K (right).

test this, a statistical analysis of different versions of the Model-Free analysis was performed for each conditions at 298.1 K (Supplementary figure A.21). The results show that from 0 to 150 g/L of PEG, a three dynamic modes model is still relevant given the experimental data. At 250 g/L of PEG, the statistical analysis suggests that a two dynamic modes fitting does nearly as well, with equal reduces chi-square values in the helical propensity domain, which confirms our previous observations. The activation energy parameters are also fitted in the Arrhenius fit, but they are poorly defined due to the relatively limited amount of experimental data. We note however that the distribution of their value tend to increase at 250 g/L of PEG, suggesting an increased average energy barrier between different conformations inter-converting at the relaxation-active timescales (Figure A.22). This is consistent with the hypothesis of restricted motional processes which was also given in the study of the dynamic properties of MeV Ntail upon liquid-liquid phase separation [53].

The two proteins that were previously characterized [52] did not show strong evidence of amplitude modulations upon crowding. This can be explained by the fact that the maximum mass concentration of polymer crowder used for the analysis was 195 g/L for SeV Ntail and 95 g/L for MKK4, values that are close or within the range where the model is valid in our analysis. However, Adamski and coworkers also measured two relaxation rates in oocyte cells, where the solvent friction was measured to be around 1, which corresponds to the solvent friction at 250 g/L of PEG. The measured relaxation rates were in reasonable agreement with the predictions from the model suggesting that the model predicts NMR spin relaxation rates in oocyte cells. The main difference between 250 g/L of PEG and oocyte cells is the type of crowder. In cells, a significantly high proportion of the crowding proteins are globular, and unlike polymer coils and IDPs that can exhibit high levels of chain proximity, we can hypothesize that globular proteins, because of their rigid and compact nature, give more space for IDPs to move between the inter-macromolecular space.

To visualize what 250 g/L PEG represents in a protein solution, we generated simulation boxes containing one copy of MeV Ntail 465 with a conformation taken from an experimental data-based ensemble and filled this box with multiple copies of PEG10000 molecules with random conformations such that the concentration in the box reached 75, 150 and 250 g/L PEG respectively. An image of MeV Ntail 465 surrounded by PEG inside these boxes is shown in figure 5.9. In the 75 g/L PEG box, a reasonable amount of space is present between each polymer coils such that we could define a hydrodynamic sphere. At 150 g/L PEG, the coils start to overlap, while 250 g/L PEG exhibit a highly concentrated and overlapped box of polymer coils. By examining these boxes, it appears that 75 g/L of PEG is in the limit between the dilute and semi-dilute regime, while the 150 g/L condition is in the semi-dilute regime. In the 250 g/L box, the PEG molecules are highly close to each other and a significant proximity between the polymer chains is observed, akin to the situation observed in our simulations of a biomolecular condensate, suggesting that 250 g/L of PEG is in the concentrated regime as defined by polymer theory. Our qualitative observations does not rely on assumptions concerning the average radius of gyration of PEG as a function of PEG concentration, which in principle plays a role in the determination of the polymer solution regime.

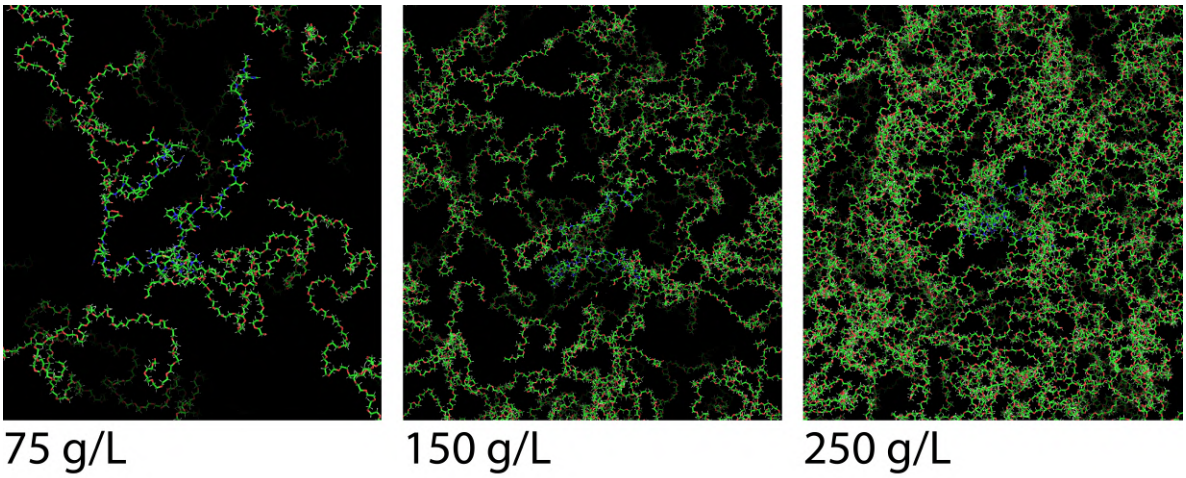


Figure 5.9: All-atom simulation boxes of MeV Ntail 465-525 in an environment including PEG10000 molecules at a concentration of 75, 150 and 250 g/L PEG respectively from left to right.

5.4 Discussion

In this study, extensive NMR spin relaxation measurements were performed on MeV Ntail 465-525 in different levels of PEG concentration. We showed that viscosity alone could not explain the dynamic properties of IDPs in a highly concentrated regime of polymer-crowded solution. Further analysis of the measured relaxation rates showed that a non negligible redistribution of the contribution of each dynamic mode on NMR spin relaxation was present at the highest crowding concentration, akin to the observations made in biomolecular condensates in chapter 3. Previous measurements in lower concentrations of polymer crowding did not reveal this effect on internal backbone dynamics and are in agreement with our relaxation measurements at concentrations up to 150 g/L of PEG. Few measurements performed in oocyte cells in the previous study didn't provide any evidence for a deviation from the nanoviscosity model. The pos-

sibility of a protein-specific mechanism that would change the dynamic properties of MeV Ntail 465-525 cannot be excluded, and intramolecular PRE measurements would be helpful to determine whether more long-range contacts are affected at higher PEG concentrations. However, the carbon chemical shift data along with the uniformity of this amplitude modulation strongly suggests that the observations are of the same physical origins as the observations of MeV Ntail in biomolecular condensates. A clear difference between polymer and globular crowding can be hypothesized from these considerations. While a high mass concentration of polymers, synthetic or peptidic, would result in a highly entangled solution, with high levels of promiscuity between the chains akin to a spaghetti plate, a globular crowded environment, with rigid crowding agents can provide more inter-molecular space for IDPs to move freely independently of the presence of the crowders. MD simulations in two of such conditions would confirm this hypothesis.

This result has important implications, since it means that the amplitude of each dynamic modes in Intrinsically Disordered Proteins can be highly affected in other physiological environments where polymer crowding is important. While it was not observed in the two measured rates in oocyte celles, a very recent paper observed related results on a loop region of a folded protein [322]. They showed by model-free analysis of backbone amide NMR spin relaxation rates that the order parameter of such a loop domain increases *in cellulo* when compared with *in vitro* results, suggesting a restriction of the motion in the cellular environment. The difference between this partially folded protein and the IDP studied in Adamski *et al.* [52] is that it contains two folded parts that can be expected to be more sensitive to globular crowders than flexible backbone chains. This difference could increase the sensitivity of the disordered region to the globular crowding, compared to a purely flexible chain.

Cellular environments are highly different depending on the organism and the location in the cell. In *E.coli*, the cellular environment is known to be highly crowded with between 200 and 320 g/L of proteins, 75 to 120 g/L of RNA and 11 to 18 g/L of DNA whereas mammalian cells contain rather 50 to 250 g/L of proteins and 20 to 50 g/L of nucleic acids [323]. We can therefore expect a different modulation of the dynamic properties of proteins in these different environments.

Our study provides an extensive data set of relaxation rates in different levels of crowding concentrations with PEG as the crowder. It can be of interest in the future to look for different kind of crowding agents, since each crowding agents vary in terms of shape and possible soft interactions. It can therefore be expected that different results might be encountered with different crowding agents as the previous results in oocyte suggest. However, it seems likely that these observations are not specific to PEG as a polymer crowder, since similar results are found in biomolecular condensates, another kind of crowded condition. We suggest that these observations are of the same physical origins than what is observed in biomolecular condensates. Furthermore, PEG is an extensively used crowding agent in the literature, and it was shown to provide relevant results on crowding effects despite the known existence of weak soft interactions [238, 250].

In addition, it might be of high interest to perform NMR spin relaxation measurements of such systems in a wider range of magnetic fields to grasp the full details of the backbone dynamics. It is clear that high resolution relaxometry might provide valuable complementary information on slower motion processes.

Understanding the behavior of proteins in crowded and cellular environments is crucial to understand biological processes. The development of *in-cell* methods is therefore of high importance. Progress is being made in both NMR [324] and single fluorescence methods [239] to tackle the

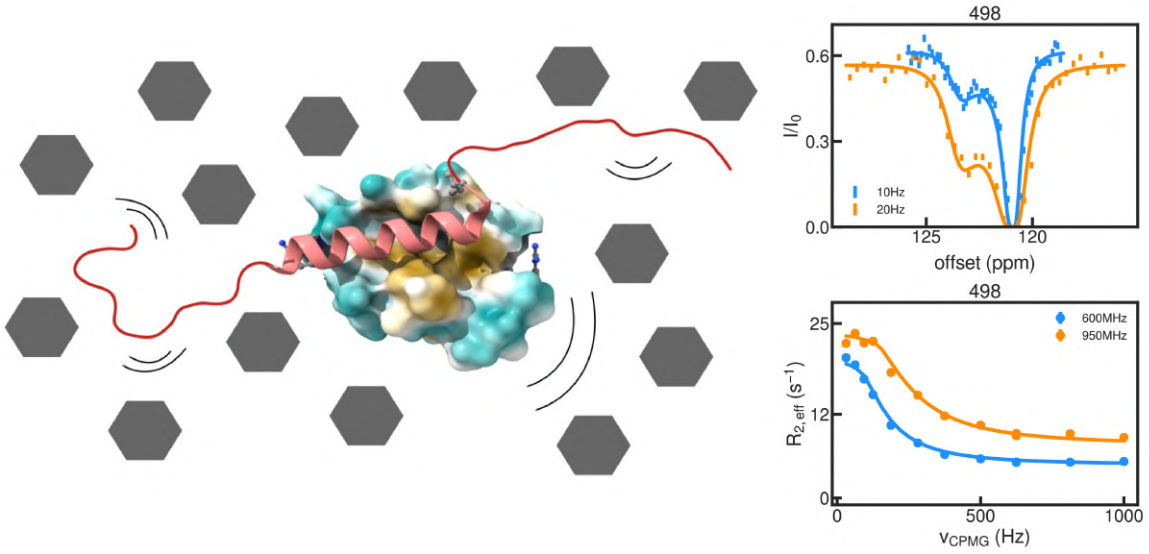
experimental challenges that these complex environments represent, which opens new perspectives towards understanding the dynamics of proteins in living systems. In addition, effort also needs to be made towards the assessment of the impact of crowding on the function of IDPs. since NMR provides atomic resolution information, there is no doubt that it will contribute to a better understanding of how the dynamic properties of IDPs are related to their function in the physiological environment.

5.5 Conclusion

In this work, we characterized the dynamic properties of MeV Ntail 465-525 in polymer crowders and show that its behavior can be explained only by nanoviscosity up to a given threshold of polymer concentration. In the more concentrated regime, a non negligible redistribution of the dynamic modes of Ntail's backbone is observed, akin to the observations in biomolecular condensates. Long-known polymer physicochemistry principles allow a sound explanation to these findings. These observations show how extreme polymer crowding can affect the internal dynamic properties of IDPs, suggesting that highly crowded environments in some cellular systems might modulate protein dynamics.

Chapter 6

Exploring protein interactions in crowded environments by NMR



At the nanoscale, a majority of the biological studies can be reduced to the study of structures, biomolecular assemblies, slow conformational dynamics and interactions. A massive proportion of such studies are performed *in vitro* in buffer solution even though it is widely accepted that biomolecular interactions are highly dependent on the solvent we are working on: pH, viscosity, salt composition and concentration, temperature and crowding. Obviously, the physiological environments in which the real biological mechanisms take place is very different from the usual NMR test tube and even if test tube studies represent a reasonable approximation of what is interacting with what and with which affinity, it remains very important to understand the above mentioned environmental parameters and how they affect protein-protein interactions. In this chapter, we study the interaction of the C-terminal domain of the Measles Virus Nucleoprotein with its partner the XD domain of the Measles Virus Phosphoprotein upon crowding to understand, as a first iteration, the effect of crowding on such interaction. This specific interaction is involved in the replication machinery of the Measles Virus and it is hypothesized that a phase separation of these two proteins in the cell might be important. This study is a first step towards a more ambitious extensive study of how a specific environmental condition affect a specific type of interactions.

6.1 Introduction

Nearly all biological mechanisms are modulated by protein interactions. Each interaction has a multiplicity, an affinity and kinetic properties that define the function of the interacting proteins in the context of a biological mechanism. Most of the interaction studies are performed *in vitro*, in the hope that the results obtained in buffer solutions are transferable in the living cell. To a good extent, the interactions that are detected *in vitro* usually exist *in vivo* as well, and this assumption is thus valid. However, it is now well known that an interaction in physiological environments, because of its complex nature, exhibit different kinetic and thermodynamic properties than what is studied *in vitro*. This can be important for drug development, when the target interaction between a protein and its ligand is affected by the physiological environment, a different response to a drug might be expected compared to what was studied *in vitro*. This can also affect the conclusion of *in vitro* protein interaction studies. For these reasons, it is extremely important to study the effect of the physiological environments on the kinetic and thermodynamic properties of protein interactions.

A physiological environment differs from the test tube in many ways. The composition of salt can be modulated in the buffer solution to approximate the salt concentration in the cell. Macromolecular crowding on the other hand is difficult to take into account without prior knowledge of its effects on the studied interaction. In addition, the macromolecular crowding in the cell arises from an extremely complex mixture of proteins, RNA, DNA and metabolites. Numerous studies attempted to understand how we could approximate the effects observed *in cellulo* with simple crowders. Recent results with PEG found that long chain polymers were appropriate to simulate the viscosity of physiological environments [250].

The recent single molecule fluorescence studies of protein-protein interactions in cellular and crowded environments rationalize their results in the light of polymer theory [243, 250]. They find that the modulation of the kinetic and affinity properties of their interactions could be modeled taking into account the increased viscosity as well as the so-called depletion interactions that takes place between two species close together in a crowded mixture of polymers [246–248]. In a dilute and semi-dilute regime, the association rate k_{on} found by fluorescence increases with the polymer concentration because of the depletion interaction, and after a certain threshold, it is believed to dramatically decrease because of the high viscosity and the quenching of the depletion effect due to the high density of polymer coils.

So far, such results were only observed with low resolution methods like single molecule fluorescence, where the accessible concentration is limited by technical limitations [250]. NMR can provide atomic-resolution information on protein-protein interactions through the process of chemical exchange and give valuable information on how crowding can affect protein-protein interactions.

Here, we use NMR spectroscopy to investigate the effect of polymer crowding on the interaction between the intrinsically disordered C-terminal domain 465-525 of the MeV Nucleoprotein and its folded partner the XD domain of the MeV phosphoprotein (XD). This interaction is known to be important for the replication machinery of the Measles virus [314], and it was recently shown that this interaction was necessary for phase separation of the Nucleoprotein with the Phosphoprotein. It was hypothesized that the resulting highly concentrated biomolecular condensate is important for the formation of Nucleocapsids in the context of the Measles virus replication [325].

We show by CPMG dispersion and DANTE-CEST experiments on both MeV Ntail 465-525 and XD that the interaction is significantly slowed down upon crowding, while the chemical shift difference isn't strongly affected by the concentration of PEG. Further future measurements on different admixtures should give information on the affinity and association rate of the interaction, in order to understand how polymer crowding affects this important interaction.

6.2 Materials and Methods

Protein Preparation

The sequence of the short C-terminal domain (residues 465-525) of the Measles Virus Nucleoprotein was cloned in a pET22b vector with an N-terminal 6His-tag and a TEV cleavage site. The gene was transformed into *Escherichia coli* BL21 (Novagen). The unlabelled, single labelled ^{15}N and double labelled $^{13}C^{15}N$ proteins were expressed in LB and M9 media respectively at 37°C until reaching a 600 nm optical density of 0.6. Then, induction of 1mM isopropyl-β-D-thiogalactopyranoside was performed before overnight incubation at 20°C. After cell centrifuging at 5000 rpm for 20 min, the pellets were resuspended in lysis buffer (150mM NaCl, 20mM Tris-HCl, pH 8.0). Sonication was performed with 3 cycles of 11 min duration [1 sec sonication, 3 seconds delay] before centrifuging at 18000 rpm for 55min at 5°C. The supernatant was then applied in a Nickel affinity chromatography, washed with lysis buffer and lysis buffer containing 20mM imidazole and eluted with lysis buffer containing 500mM imidazole. The resulting solution was then subjected to dialysis overnight at room temperature along with TEV protease with a 3.5 kDa dialysis membrane and finally subjected to size exclusion chromatography (SEC, Superdex75) in NMR buffer (150mM NaCl, 50mM sodium phosphate, 2mM dithiotreitol pH 6.0).

The sequence of the C-terminal XD domain of the Measles Virus Phosphoprotein was cloned in a pET22b vector with an C-terminal 6His-tag. The XD domain of the phosphoprotein was purified as Ntail 465-525 except that the dialysis step was performed without TEV enzyme and at 5°C instead of room temperature.

NMR spectroscopy

NMR experiments were performed on Bruker spectrometers operating at magnetic fields corresponding to ^{1}H frequencies of 600 and 950 MHz. The protein NMR buffer contained 0, 75, 150 or 250 g/L PEG10000 and the measurements were performed at 298.1K, 288.1K or 278.1K. The protein concentration was 300μM unless stated otherwise. The spectra were processed using NMRPipe [318] and analyzed in Sparky.

NMR Chemical exchange experiments

All the chemical exchange experiments were performed at a ^{1}H frequency of 600 MHz and 950 MHz at 298.1 K with a cryogenic probe. the CPMG relaxation dispersion experiments [87] were performed with a constant-time relaxation delay of 32 ms and CPMG frequencies between 31 and 1000 Hz. The ^{15}N DANTE-CEST experiments were performed with B_1 field strengths of 10 and 20 Hz and frequency windows of 450 and 800 Hz respectively with a constant-time relaxation delay of X ms. All the NMR exchange data was fitted using Chemex [320]. Error bars were determined with Monte-Carlo analyses.

6.3 Result and discussion

6.3.1 XD and Ntail 465-525 in crowding conditions

The behavior of both MeV Ntail 465-525 and XD was assessed by NMR spectroscopy at different concentration of PEG. The ^{15}N-^{1}H HSQC shows very little variation of the chemical shifts in XD upon crowding apart from a few residues (figure 6.1), while, as seen in the previous chapter, the chemical shifts of MeV Ntail 465-525 exhibits more visible perturbations despite no variations of the carbon chemical shifts. These HSQCs show that the integrity of XD is not affected upon crowding with PEG.

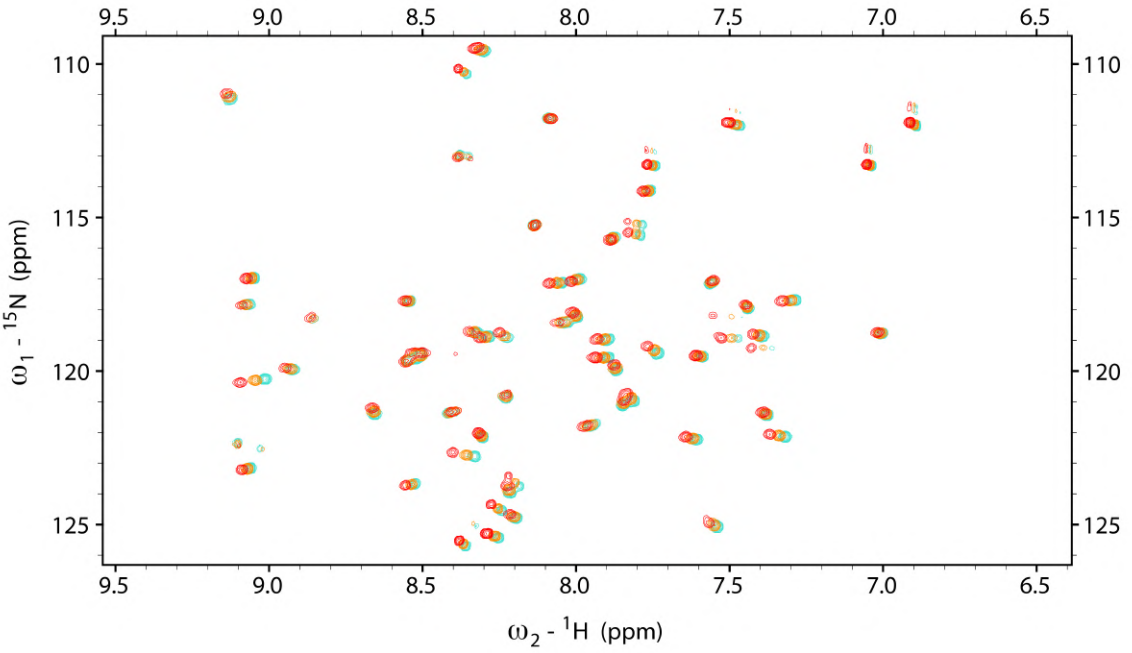


Figure 6.1: Impact of crowding on MeV XD chemical shifts. ^{1}N-^{15}N HSQC spectra at different concentrations of PEG: 0 (Blue), 75 (Orange) and 150 g/L (Red).

6.3.2 Interaction from the side of Ntail 465-525

The interaction of MeV Ntail 465-525 with XD was first studied on the side of Ntail. Different admixtures were prepared with different populations. As expected from previous work [325, 326], strong dispersion is observed at low proportions of XD, in both dilute and crowded conditions (Figure 6.2) in the helical propensity domain of the IDP. Above 20% of partner, the signals coming from the helical propensity domain of Ntail 465-525 approach 0 or start to disappear.

To further characterize this interaction, CPMG relaxation dispersion experiments were performed with 5% of XD at concentrations of 0, 75 and 150 g/L of PEG10000 and at two magnetic fields: 600 and 950 MHz. For each condition, a global fit of all the residues was performed and yielded a relatively consistent agreement with the experimental data (Supplementary figures A.24, A.25 and A.26) with reduced chi-squares of 4.45, 7.00 and 5.86 for 0, 75 and 150 g/L of PEG respectively. These chi2 values are high in terms of confidence limits. We are not sure yet whether the origin of this discrepancy comes from the over-simplicity of the 2-state model or from an underestimation of the experimental uncertainty. The global results show that most of the dispersion is found as expected in the helical propensity domain, with chemical shift differences that vary

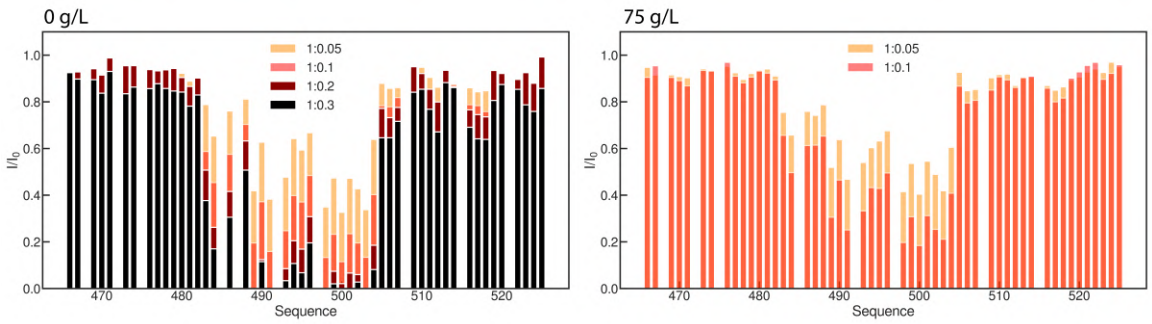


Figure 6.2: Signal intensity ratio of ^{1}N-^{15}N HSQC spectra between MeV Ntail 465-525 alone and mixed with 5 (Orange), 10 (red), 20 (dark red) or 30 % (black) of XD with 0 g/L (left) or 75 g/L (right) of PEG10000.

along the sequence and reach 2 ppm is some residues, especially residues L498, A500, A502 and G503 where the chemical shift difference is the highest between the free and the bound state (Figure 6.3). This region is a relatively hydrophobic part of the helical propensity domain known to bind the hydrophobic patch of XD [327]. The exchange rate were fitted around 291 ± 15, 223 ± 17 and $147 \pm 39\ s^{-1}$ with PEG concentrations of 0, 75 and 150 g/L respectively, again bearing in mind a possible underestimation of the uncertainty that here takes only into account the error from the NMR measurements. The bound state populations were fitted to be around 0.05, which corresponds to the proportion of XD that we have in the solution. The affinity of the interaction, at least in the dilute state, is known to be relatively low and within the micro-molar range, which explains why the present fitting gives p_B values close to the maximal value of 0.05 (which would correspond to an infinitely strong interaction). Further measurements with different admixtures should be performed for each conditions with a careful estimation of the XD concentration to obtain information on the K_D of the interaction.

The dynamic range of this interaction is favorable to the measurement of CEST experiments, since some residues exhibit relatively high chemical shift differences and since the extracted k_{ex} were between 100 and 300 s^{-1}. DANTE-CEST experiments were performed with an admixture of Ntail 465-525 with 5% XD with 75 g/L PEG to obtain further information on the interaction, including the sign of the chemical shift difference. The DANTE-CEST data in the high chemical shift difference region were fitted simultaneously with the CPMG data in this condition and yielded a relatively good overall agreement with experiments although the reduced chi-square of 5.02 indicates that the fit does not fully capture the data, again either because of error underestimation, or because of the model's simplicity. Assuming that the difference in $\Delta\delta$ is negligible between the different conditions, the data at 0 and 150 g/L of PEG were fitted with the fixed chemical shift differences obtained from the DANTE-CEST experiments performed at 75 g/L of PEG. A residue-specific fit of the exchange parameters in the helical propensity domain is shown for each conditions in supplementary figure A.27 and further proves that the chemical shift difference doesn't significantly evolves with PEG concentration between residues 498 and 503. The fit of each condition with the $\Delta\delta$ from the DANTE-CEST gives good agreement with experiments with reduced chi-squares of 1.76 and 2.95 for 0 and 150 g/L of PEG respectively. The DANTE-CEST and CPMG profiles are shown in figure 6.4A and the fitted parameters are shown in figure 6.4B. The fitted exchange rates remain in the range that was obtained from the CPMG data alone, and the sign of the chemical shift difference is now accessible, giving more accurate structural information on the bound state. The real values of the bound state population remains close to 0.05 in each condition and is difficult to obtain precisely due to the strength of the interaction and the uncertainty of the protein concentration in the sample.

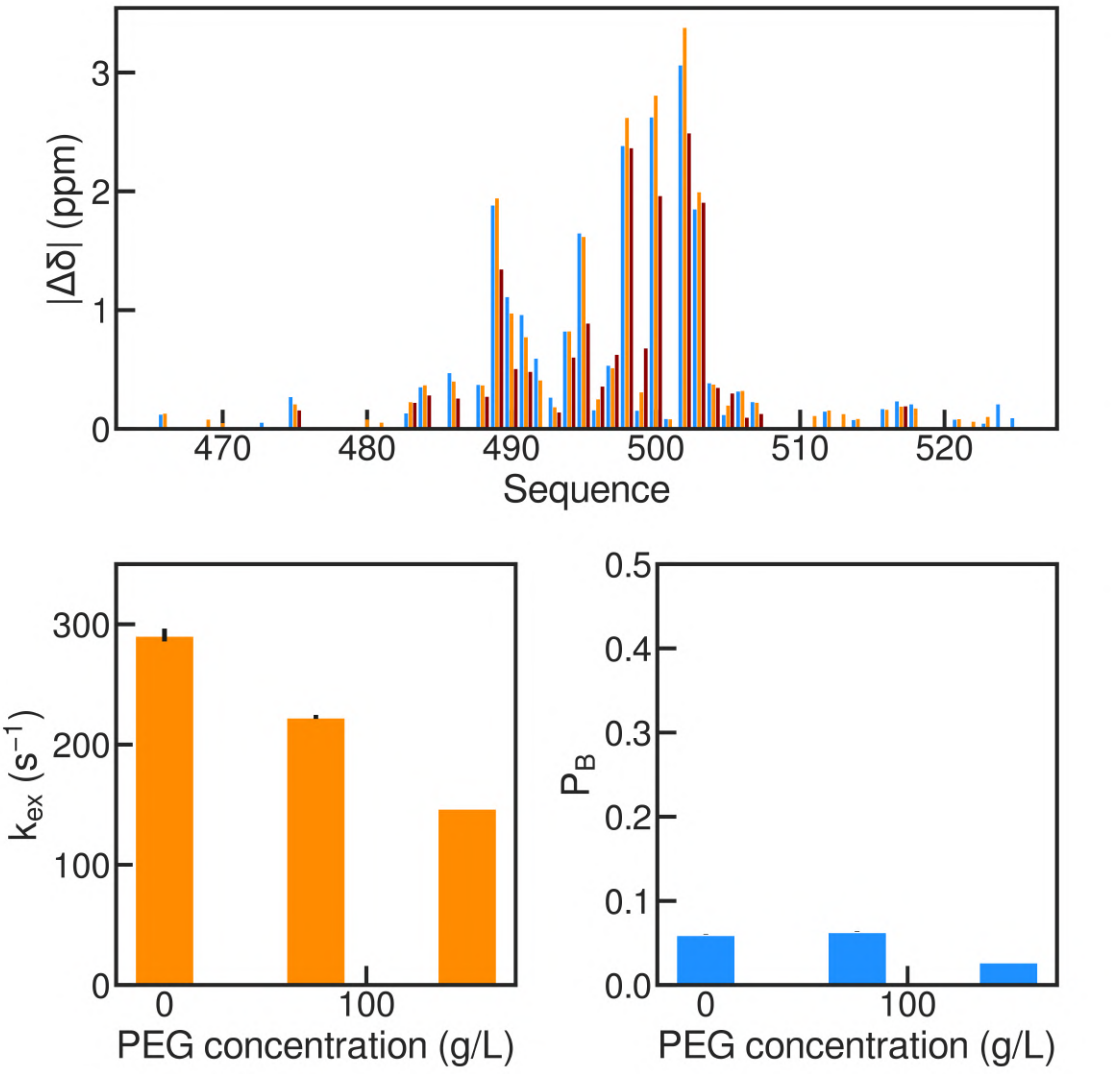


Figure 6.3: Chemical shift differences (top), exchange rates (bottom left) and bound state populations (bottom right) fitted to a 2 states model on all residues from CPMG data at two magnetic fields on Ntail 465-525 with 5% of XD with 0 (blue $\Delta\delta$), 75 (orange $\Delta\delta$) and 150 (red $\Delta\delta$) g/L of PEG.

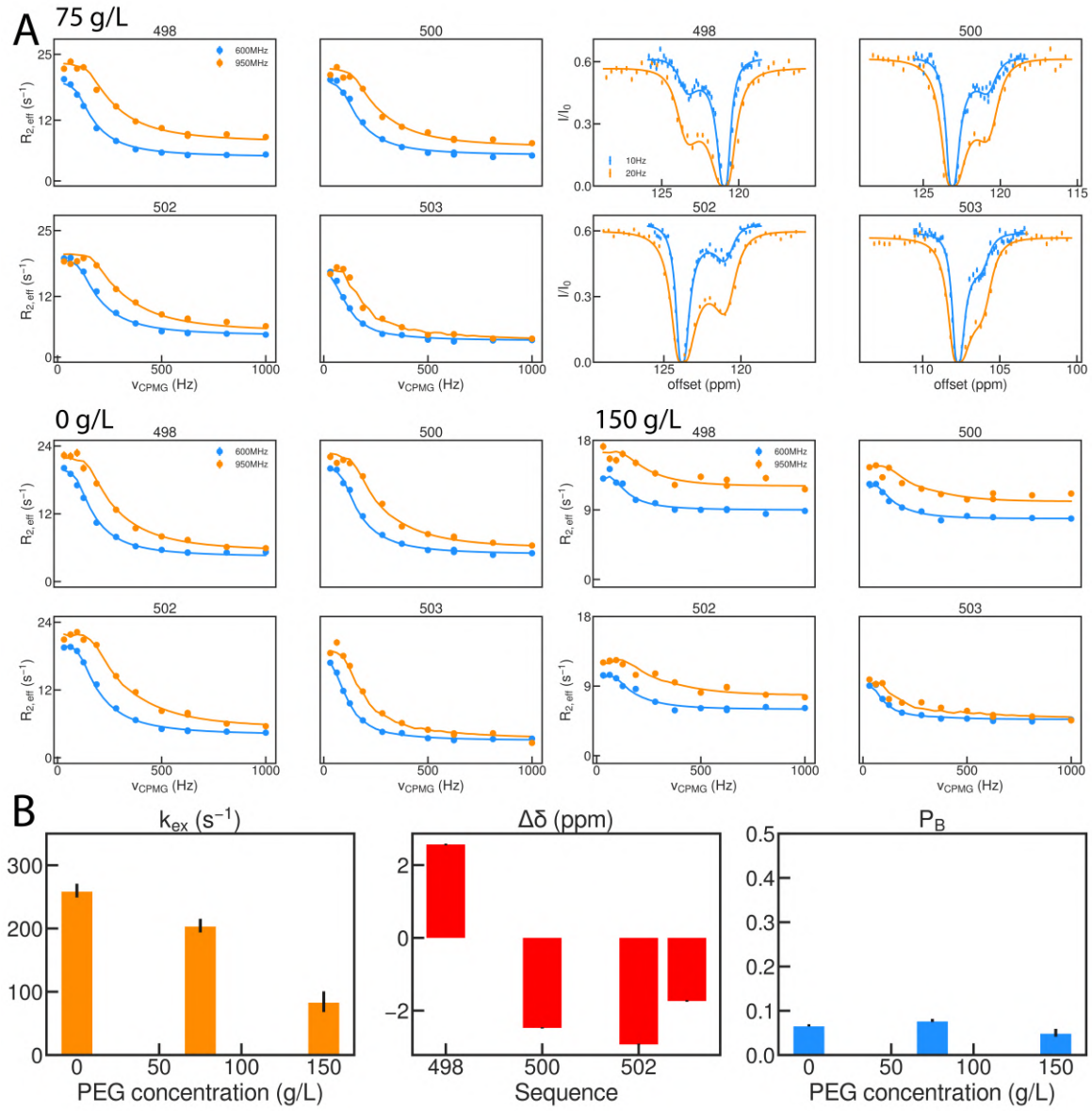


Figure 6.4: Panel A: experimental (circles) and fitted (lines) CPMG and DANTE-CEST profiles of four residues of Ntail with 5% XD. Top left: CPMG profiles at 75 g/L PEG at 600 (blue) and 950 (orange) MHz proton frequency. Top right: DANTE-CEST profiles at 75 g/L PEG with b1 fields around 10 (blue, 11.058 Hz) and 20 Hz (orange, 20.796 Hz). Panel B: Exchange rates (bottom left), chemical shift differences (bottom center) and bound state populations (bottom right) fitted to a 2 states model on residues 498, 500, 502 and 503 with CPMG and DANTE-CEST data on Ntail 465-525 with 5% of XD.

6.3.3 Interaction from the side of XD

To further confirm the observations made on the side of Ntail, the interaction was also studied on the side of XD at the conditions of 0 and 75 g/L of PEG. CPMG relaxation dispersion experiments were performed at two magnetic fields with labeled XD in presence of 15% Ntail 465-525. The dispersion profiles with a global fit over all residues are shown in supplementary figures A.28 and A.29. In this case fitting all the residues together yields acceptable agreement with experiments with a reduced chi-square of 1.25 and 1.09 for the conditions with 0 and 75 g/L of PEG respectively. The fitted parameters are shown in figure 6.5. Again, the chemical shift differences are consistently similar between the buffer and crowded condition. The extracted k_{ex} values were fitted around 290 and 230 s^{-1}, which falls within the range that was obtained from the side of Ntail 465-525. Additionally, the extracted populations remained close to 0.15 with values of 0.14 and 0.11. If the protein concentrations were accurately estimated, this means that there is less binding upon crowding with 75 g/L of PEG. However, more measurements are necessary to confirm such observation. Globally, these measurements further confirm that crowding slows down the exchange rate between Ntail and XD.

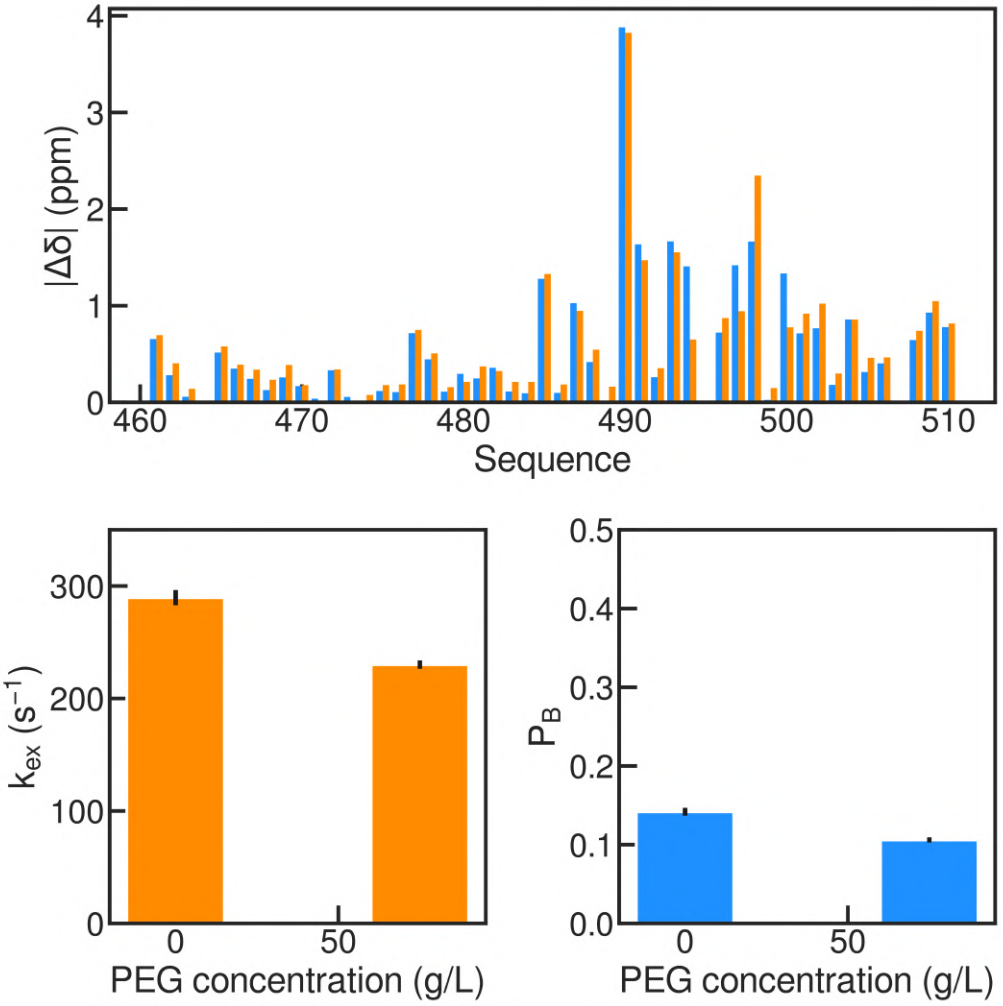


Figure 6.5: Chemical shift differences (top), exchange rates (bottom left) and bound state populations (bottom right) fitted to a 2 states model on all residues from CPMG data at two magnetic fields on the XD domain of the MeV Phosphoprotein with 15% of Ntail 465-525 with 0 (blue $\Delta\delta$) and 75 (orange $\Delta\delta$) g/L of PEG.

6.4 Discussion and future work

The interaction between the XD domain of the MeV Phosphoprotein and MeV Ntail 465-525 has been studied from both sides by NMR with one admixture for each sides. All the results show that crowding slows down the exchange rate between the free and bound state, as expected from an increased viscosity. The NMR data also shows that the nature of the interaction and the binding site is not strongly affected by crowding, as shown by the similarity in chemical shift differences between the free and bound state in each condition.

Further measurements with different proportions of the proteins will be carried out to obtain insight on the relative affinity of the interaction in the different crowding levels as well as on the on k_{on} and off-rate k_{off}. monitoring these parameters will allow us to interpret the effect of PEG on this interaction in terms of depletion interaction and protein diffusion slow-down upon crowding. If enough data are obtained, a fit of the parameters with the Berezhkovskii-Szabo model [249] is possible, provided that this model can explain the data, allowing to quantitatively describe the phenomena of depletion interaction and diffusion modulation.

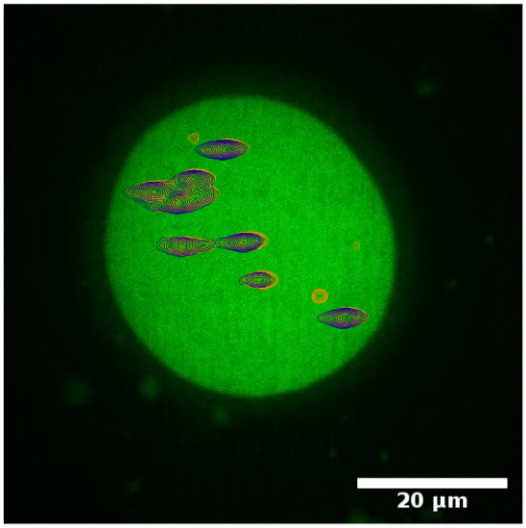
NMR shines at providing atomic-resolution information on protein interactions. It is known that the helical propensity domain of Ntail is exhibiting a perpetual exchange between different conformations. A similar interaction system, SeV Ntail with SeV XD, exhibited a three-state exchange mechanism if fitted with carbon CPMG data and described a folding upon binding mechanism of the helical propensity domain of Ntail with XD [328]. It can be interesting to observe how this mechanism is affected upon crowding. To do this, we could perform CPMG relaxation dispersion experiments on carbon atoms of the backbone of MeV Ntail and analyze the data along with the nitrogen CPMG data.

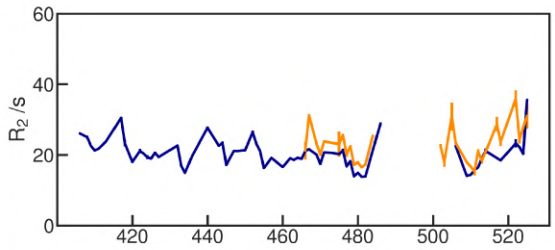
So far, we characterized the global kinetic properties of the interaction of MeV Ntail with MeV XD. We showed that the nitrogen chemical shift changed only slightly along the sequence of both interacting proteins, and that the exchange rate between the proteins was slowed down by the crowding agent. Further work should provide information on the thermodynamic properties of this interaction and quantitatively explain it in terms of the already known crowding effects that are depletion interactions and diffusion slow down. Furthermore, a multinuclear exchange study could allow us to study the effect of crowding on conformational funneling mechanisms as observed in another related system [328].

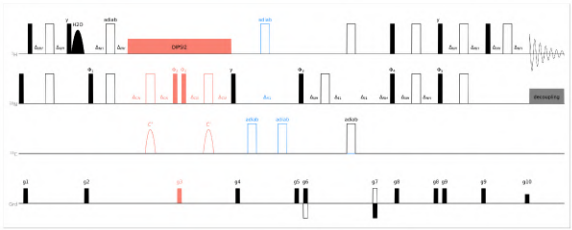
Understanding the modulation of molecular mechanisms at atomic resolution in complex environments is of high importance to confidently transfer *in vitro* studies *in vivo*, and understand how the complex cellular environment can modulate biological mechanisms.

Chapter 7

Optimizing NMR methodologies to study IDPs in condensed phases and complex environments







Performing NMR spectroscopy in environments like biomolecular condensates is a challenging task, first because of the rather challenging sample preparation that requires a large amount of protein, but also for technical reasons because of the intrinsic complexity and high viscosity of the system. In this chapter, we address a few issues encountered during our studies of liquid-liquid phase separation and propose some solutions. In a first part, we address the variability of the condensate's dynamic properties as a function of the sample preparation conditions and time. Then, we will address the difficulty of measuring cross-correlated transverse relaxation rates and propose a sensitivity enhancement version of the pulse sequence that successfully solved our issues. Finally, we address the complexity of studying mixtures in a biomolecular condensates and other complex systems and the possible occurrence of natural abundance signal that can affect intensity-dependent measurements. We propose isotope edited/filtered relaxation experiments to allow precise measurements of relaxation rates in complex environments.

7.1 Introduction

7.2 Materials and Methods

Protein preparation

The expression and purification of MeV Ntail 402-525 was performed as described in chapter 4. The sequence of the short C-terminal domain (residues 465-525) of the Measles Virus Nucleoprotein was cloned in a pET22b vector with an N-terminal 6His-tag and a TEV cleavage site. The gene was transformed into *Escherichia coli* BL21 (Novagen). The unlabelled, single labelled ^{15}N and double labelled $^{13}C^{15}N$ proteins were expressed in LB and M9 media respectively at 37°C until reaching a 600 nm optical density of 0.6. Then, induction of 1mM isopropyl-β-D-thiogalactopyranoside was performed before overnight incubation at 20°C. After cell centrifuging at 5000 rpm for 20 min, the pellets were resuspended in lysis buffer (150mM NaCl, 20mM Tris-HCl, pH 8.0). Sonication was performed with 3 cycles of 11 min duration [1 sec sonication, 3 seconds delay] before centrifuging at 18000 rpm for 55min at 5°C. The supernatant was then applied in a Nickel affinity chromatography, washed with lysis buffer and lysis buffer containing 20mM imidazole and eluted with lysis buffer containing 500mM imidazole. The resulting solution was then subjected to dialysis overnight at room temperature along with TEV protease with a 3.5 kDa dialysis membrane and finally subjected to size exclusion chromatography (SEC, Superdex75) in NMR buffer (150mM NaCl, 50mM sodium phosphate, 2mM dithiotreitol pH 6.0).

Condensate Sample Preparation for NMR

The condensate samples were prepared as described previously in chapter 4.

NMR spectroscopy

The NMR experiments were performed on Bruker spectrometers operating at magnetic fields corresponding to ^{1}H frequencies of 700 MHz equipped with a cryogenic probe or an Oxford-Bruker spectrometer operating at 600 MHz ^{1}H frequency equipped with a room temperature probe. All the spectra were processed using NMRPipe [318] and analyzed in Sparky.

NMR spin relaxation measurements

Transverse cross-correlated relaxation rates were measured with the symmetrical-reconversion method of Pelupessy *el al.* [37] using either the conventional pulse sequence presented in chapter 1 or the sensitivity enhancement method implemented and presented in this chapter.

Conventional ^{15}N longitudinal relaxation (R1), and R1rho (with a spin lock of 1.5kHz) were measured as described by Lakomek et al. [319] using an interscan delay of 1.2s and from 64 to 256 dummy scans. The used sets of relaxation delays for R_1 was [0, 0.6, 0.08, 1.6, 0.4, 0.32, 0.1, 0.2, 0.6] s. For $R_{1\rho}$ in the condensed phase, the relaxation delays were [1, 15, 35, 120, 75, 22, 6, 15, 190] ms. ^{15}N transverse relaxation (R2) was determined from R1rho and R1. Isotope filtered-edited ^{15}N longitudinal relaxation (R1), and R1rho experiments were performed with modified pulse sequences based on the conventional sequences. These pulse programs are described in this chapter.

7.3 Result and discussion

7.3.1 Impact of the phase separation conditions on MeV Ntail dynamics in the condensate

A few MeV Ntail condensate sample preparations resulted in a very viscous and opaque sample where the NMR signals are barely visible in the NMR spectra. After several trials, it was found that the initial protein concentration prior to phase separation was critical for obtaining a sample of good quality. The initial protein concentration was set to either 2 mM or 0.5 mM. The obtained samples are presented in supplementary figure A.30. With too high initial concentration of MeV Ntail, the condensate appears opaque suggesting the presence of protein aggregation in the droplet. A low initial concentration phase separation however resulted in a clean and transparent sample. NMR spectroscopy of both sample was performed and transverse relaxation was obtained. The tumbling of MeV Ntail in both samples is highly different. In the high concentration preparation, the ^{15}N transverse relaxation rates are found around 60 /s while in the clean sample, tansverse relaxation rates are around 20 /s, resulting in very clean NMR spectra. This observation highlights the importance of the initial conditions for obtaining an exploitable condensate sample. It is also believed that many other initial parameters can affect the final condensed phase droplet. We will see in the next section that the presence of degradation or impurities can also affect the dynamics of IDPs in a condensate.

7.3.2 Impact of aging on the dynamics of MeV Ntail in the condensate

All the experiments presented previously were performed within a one week time window where the biomolecular condensate is spectroscopically stable. In order to asses the effect of aging on the dynamics of IDPs in the condensates, a MeV Ntail condensate sample was kept at room temperature for one month and examined again by NMR spectroscopy. The HSQC spectra exhibit the apparition of numerous C-terminal peaks among others, indicating protein degradation (Figure 7.1). The dynamics of this aged condensate was probed using off-resonance $R_{1\rho}$ dispersion experiments (Figure 7.1 and A.31). The results show that the backbone tumbling of MeV Ntail is globally more rapid after degradation. Again, no exchange was found in the condensate. This result suggests that the presence of degradation or impurities in the sample can modify the dynamics of IDPs in our condensate. Because of high sensitivity of the droplet properties to the initial protein sample conditions, it is of high importance to standardize the protein and sample preparation, both in terms of sample purity and initial concentrations in order to obtain reproducible results.

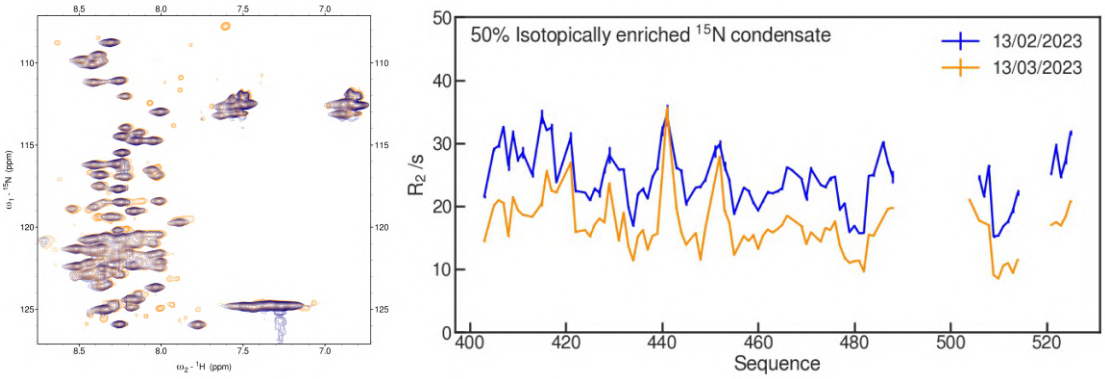


Figure 7.1: HSQC spectra (left) and ^{15}N R_2 relaxation rates (right) in MeV Ntail condense phase before (blue) and after (orange) a one month long room temperature incubation.

7.3.3 Measuring cross correlated relaxation rates in biomolecular condensates

Cross correlated relaxation rates are difficult to measure. As discussed in Chapter 1, the rate of coherence exchange between in-phase magnetization and anti-phase magnetization is probed with a symmetrical reconversion scheme [37]. In MeV Ntail biomolecular condensates, it appears that the implemented pulse programs for measuring longitudinal and transverse DD/CSA cross-correlated relaxation rates in backbone amides yield spectra that exhibit t1-noise artifacts (Figure 7.2). We inferred that these artifacts come from unwanted first order coherence during the acquisition that are not encoded in the indirect dimension. In order to overcome this artifact, an idea would be to include a first order coherence-selection between indirect dimension evolution and acquisition using gradients. The absence of t1-noise artifacts in sensitivity enhanced experiments with Echo-Antiecho frequency discrimination performed on our condensate supports our hypothesis, since such first-order coherence selection is performed in these experiments. To overcome this issue, a sensitivity enhanced transverse cross correlated relaxation rate pulse program was written, using a sensitivity enhancement sequence with preservation of equivalent pathways (Figure A.32). The resulting spectra using the new pulse program yielded no t1 noise artifact (Figure 7.2) and measurements on a conventional diluted Ntail 465-525 sample performed with both pulse programs give the same resulting cross correlated relaxation rates (Figure 7.3), confirming the accuracy of the coded sequence.

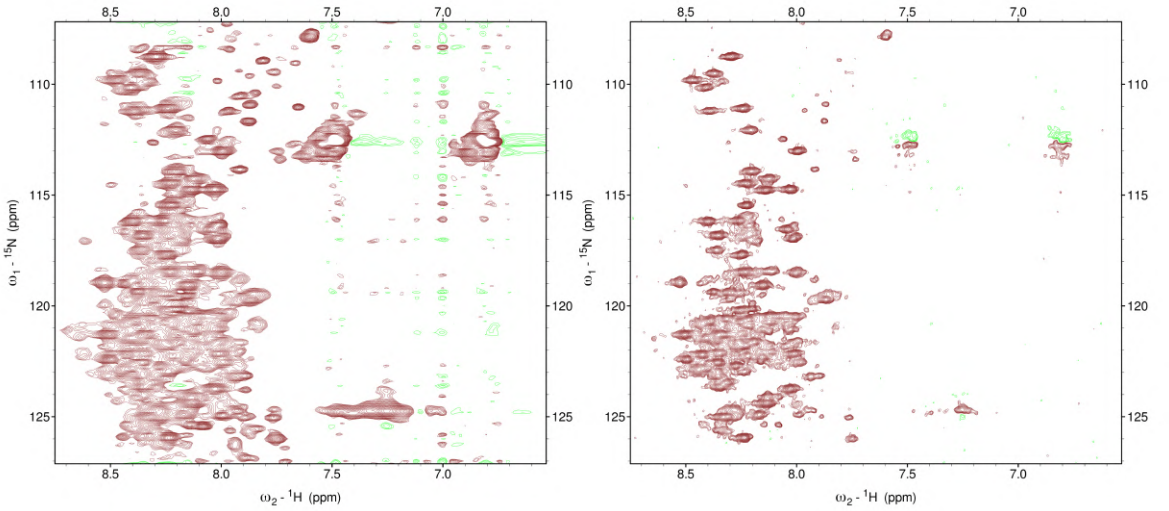


Figure 7.2: $^{15}N-^{1}H$ 2D correlation spectra taken from cross correlated transverse relaxation rate measurements in a one month old MeV Ntail condensate with a states-TPPI (left) or a sensitivity enhancement encoding (right).

7.3.4 NMR of client proteins in a biomolecular condensate

Biomolecular condensates in cells are expected to involve numerous biological processes. In addition, membraneless organelles are known to involve several interacting proteins and RNAs. It is thus of interest to be able to study different proteins inside these condensates. As a first attempt towards this direction, the short construct of MeV Ntail was incorporated as a client protein inside biomolecular condensates of MeV Ntail. Fluorescence studies showed that MeV Ntail 465 enters the condensed phase of MeV Ntail, and that the dense phase still has liquid properties (Figure 7.4 top). An NMR sample was prepared and NMR spectra were recorded. The obtained HSQC spectrum is of poor quality and exhibits few observable peaks, with unexpected natural-abundance signal coming from the non enriched MeV Ntail full length protein (Figure

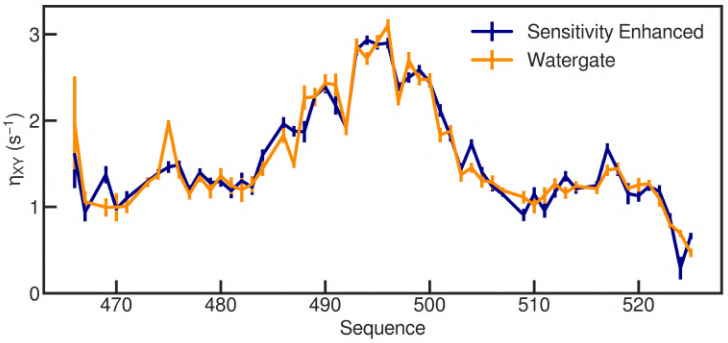


Figure 7.3: $^{15}N-^{1}H$ 2D correlation spectra taken from cross correlated transverse relaxation rate measurements in a one month old MeV Ntail condensate with a states-TPPI (left) or a sensitivity enhancement encoding (right).

7.4). NMR spin relaxation measurements were performed in this sample. Despite a small proportion of natural abundance signal coming from the dynamic residues (essentially small residues like glycines and serines) of the long form, monoexponential decays were observed in the relaxation data analysis. Transverse relaxation rates of short Ntail in the condensate appear slightly higher than the full length Ntail in the pure full length condensate, indicating a slightly increased rotational tumbling of the molecules in the condensate mixture compared to the pure MeV full length Ntail version, which means that the incorporation of another construct changed the dynamic properties of the condensate. This observation supports numerous studies on the impact of the condensate composition on the condensate's viscosity. The poor quality of the spectrum could indicate the presence of nanoscale homogeneity in the sample, which was not detected in fluorescence microscopy. It was shown that a small change in composition could drastically change the condensate's properties, making the study of condensate mixtures highly challenging.

7.3.5 Design of isotope-filtered relaxation experiments for complex protein mixtures

The observation of natural abundance signals in biomolecular condensates invited the design of isotope-filtered experiments for backbone amides, since overlap between the peaks of interest and contamination is possible in complex systems and can alter the peak intensities and therefore the measurement of observables like NMR spin relaxation. Isotope editing/filtering methods were already established and are widely used for NOESY and TOCSY experiments [329]. The typical method to remove unwanted coherence is to take advantage of the J-coupling evolution to separate the magnetization coming from our isotope of interest from the magnetization from contaminant or undesired species. This requires a different labeling scheme between the wanted and the undesired species. Therefore, in our case, double labeling our protein of interest is necessary since natural abundance signal will usually be single labeled. Some of the established methods use a double filtering scheme to take into account the broad distribution of J-couplings in the studied system. In our system, we will take advantage of the coupling between nitrogen and carbons, and the J-coupling between carbonyl and nitrogen is typically around -15 Hz with a relatively small standard deviation so that we can consider that only one scheme is sufficient. Such a filter scheme was implemented in relaxation experiment to assess its validity. The implementation on longitudinal relaxation measurements is presented in figure 7.5. Starting from the end of the refocused INEPT, our magnetization can be written as N_x. Then, the first evolution period with 180° pulses on nitrogen and carbonyl is chosen so that it transforms the double labeled systems N_x into antiphase coherence N_yC_z. At this point, single labeled proteins have

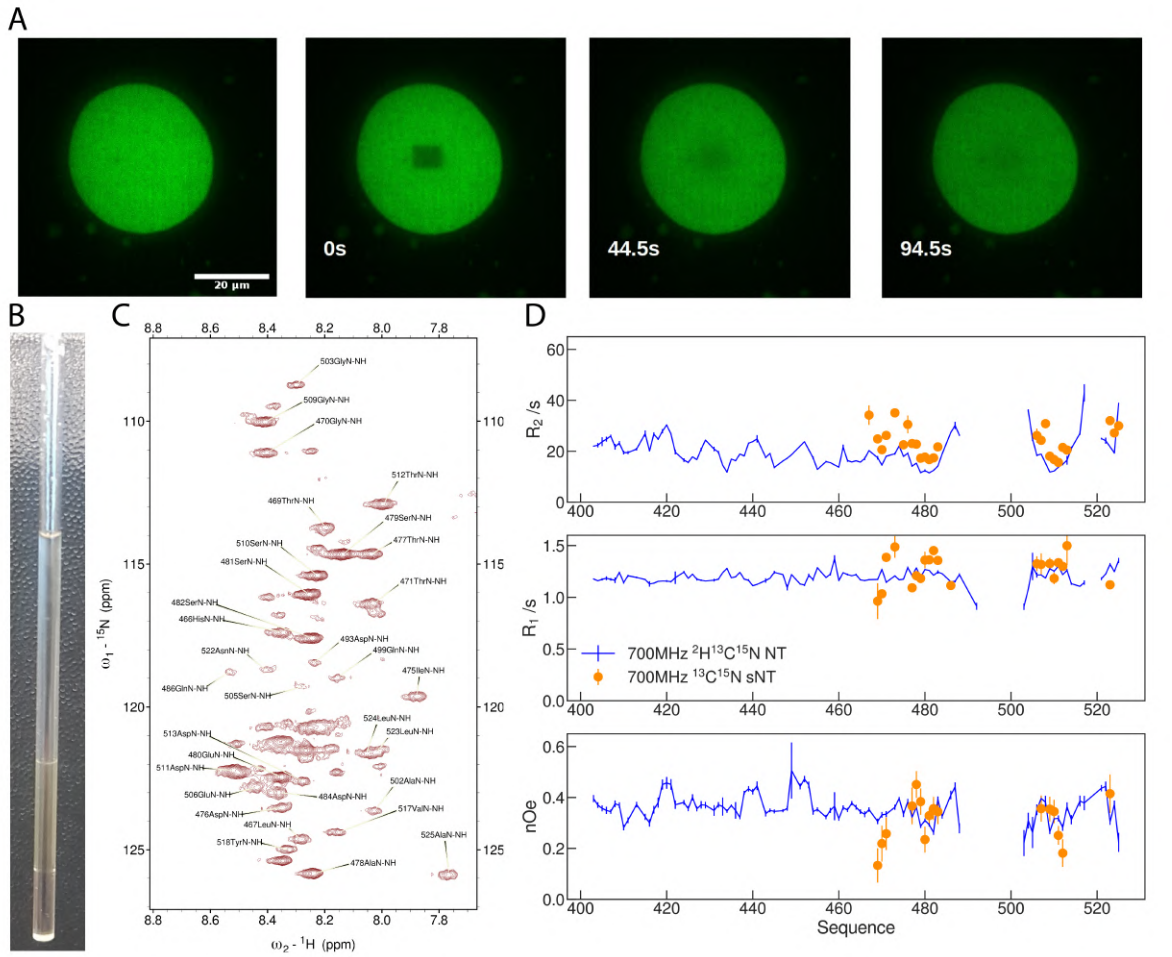


Figure 7.4: A: Fluorescence spectroscopy and FRAP images of fluorescein-labelled MeV Ntail 465-525 inside condense phases of MeV Ntail. B: NMR sample of 10% MeV Ntail 465 and 90% MeV Ntail biomolecular condensate. C: Associated HSQC spectrum. D: Relaxation rates of MeV Ntail 465 (Orange) inside MeV Ntail biomolecular condensate. In blue is the reference rate measured on the full length on a pure MeV Ntail condensate.

in-phase coherence along the x-axis, so that depending on the phase of the following 90° pulse, either the in-phase coherence is transferred along z, or the antiphase coherence is turned into ZQ coherence $N_z C_z$. If we put the magnetization or interest along z, we can then purge the unwanted coherence with the following gradient and transfer back our signal along the transverse plane for conversion back to N_x if the signal from double-labeled proteins is required.

To assess the accuracy of this filtering technique, isotope-filtered experiments were performed on a 1:1 mixture of double-labelled $^{13}C^{15}N$ MeV Ntail 465-525 and single labelled ^{15}N MeV Ntail 402-525. The 2D ^{15}N-^1H correlation spectra of the double label-selected and single label-selected experiments were compared with the unfiltered HSQC plane. The result show no trace of contamination in both selected experiments (Figure 7.7).

The filter scheme was implemented in both R_1 and $R_{1\rho}$ relaxation experiments. These relaxation experiments were performed on our mixture sample with the conventional relaxation pulse sequence as control, and with our isotope selection strategy to assess the accuracy of these experiments. The conventional and the filtered relaxation experiments give the same relaxation rate,

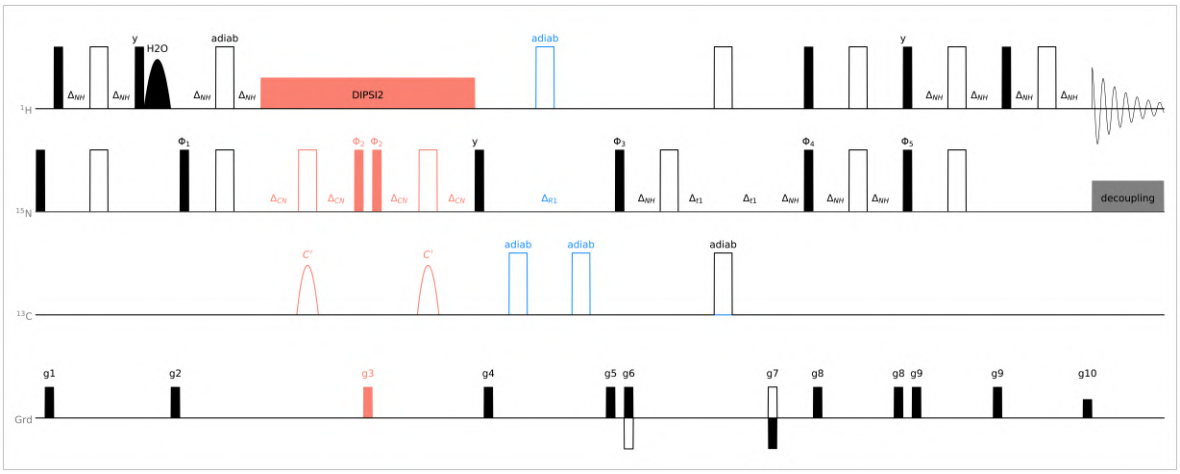


Figure 7.5: Pulse sequence for protein backbone ^{15}N filtered/edited longitudinal spin relaxation rates measurement. The filter scheme, incorporated between the refocused INEPT and the relaxation delay is highlighted in red. In blue is highlighted the relaxation delay. In principle, the delay Δ_{t1} is small and this part is looped several times to quench the longitudinal cross-correlated relaxation contribution to relaxation (With the help of the adiabatic pulse on the proton). Gradients 1, 2, 3, 4, 5, 8 and 9 are purging and cleaning gradients. Gradient 3 quenches the unwanted isotope signal. Gradients 6, 7 and 10 are the echo-anti echo encoded coherence selection gradients. Φ_2 can be chosen for single label selection (y) or double label selection (x).

confirming the accuracy of the filtered experiments. By examining the data, we observe the expected R_1 and R_2 profiles of MeV Ntail and its short form. We also retrieve the expected length dependence of the disordered chain tumbling described by Abyzov and coworkers for different constructs of the related Sendai virus Ntail protein [51].

The downside of this filter is the increased length of the resulting pulse sequence. Indeed, the J-coupling between backbone nitrogen and carbon being relatively small (15 Hz), the duration of the J-coupling evolution is long enough for signals in residues with slow rotational tumbling and therefore high transverse relaxation rates to quench almost entirely. Coherence evolution during a J-coupling transfer were simulated and showed that, while with a J-coupling of -95 Hz, we obtain a good magnetization yield with a conventional transfer length, with a J-coupling of -15 Hz and a transverse relaxation rate around 20 s^{-1}, which is the range found in our biomolecular condensate, almost half of the signal is lost, even with the shortened optimal half-transfer delay of 12.41 ms instead of 16.22 ms (for R_2 around 2 s^{-1}).

The length of the isotope selected experiments can be further optimized by performing combination experiments. To illustrate this, we take our initial in-phase nitrogen coherence N_x. In case *A*, letting the J-coupling evolve once yields $N_y C_z$ and N_x coherence for double and single labeled proteins respectively. If instead, in case *B*, the J-coupling is refocused, N_x coherences are obtained from both single labeled and double labeled proteins. Selecting N_x in both cases and calculating the difference between the signal from case *B* and the signal from case *A*, only signal from the double labeled protein is obtained. With this idea, the length of the experiment can be halved. Such a pulse sequence can also be implemented in relaxation experiment as shown in figure 7.9. With this method, at the cost of recording two 2D spectra plans per delay instead of only one, we save 50% of the signal in a condensed phase sample according to the previous simulations, which is a significant save. Unfortunately, this pulse sequence could not be tested so far, but it is a useful step for the study of complex mixtures of proteins in a biomolecular condensate.

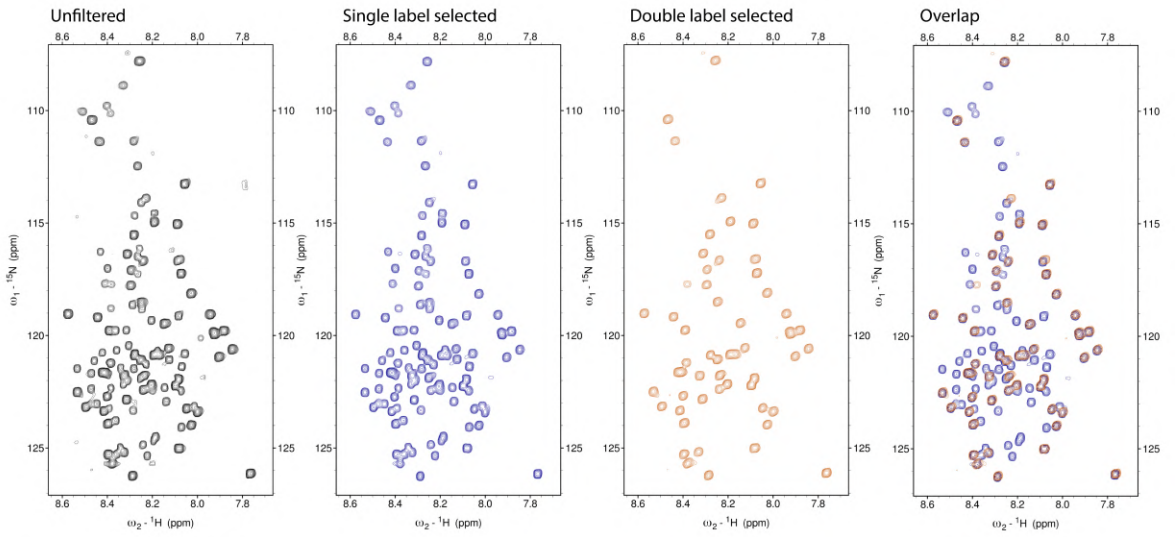


Figure 7.6: 2D ^{15}N-^{1}H correlation spectra of a 1:1 mixture of double-labelled $^{13}C^{15}N$ MeV Ntail 465-525 and single labelled ^{15}N MeV Ntail 402-525 using a conventional correlation experiment (left, black), a single label selection experiment (blue), a double label selection experiment (orange) and the overlap of the isotope selection experiments (right). The same pulse sequence is used for both selection experiment. The only variable changing is the phase Φ_2 presented in the pulse sequence in figure 7.5.

7.4 Conclusion and perspective

In this chapter, a number of challenging experiments were discussed, and some solutions were developed and proposed.

Our biomolecular condensate is relatively stable as long as no degradation is observed. However, the dynamic properties are extremely dependent on the initial preparation condition as our results show, and it is crucial to be consistent in the sample preparation to obtain reproducible results. Upon protein degradation (and probably aggregation), an increase in the protein chain's rotational dynamics is observed, possibly because of less entanglement restrictions because of the presence of shorter chains and possible concentration decrease upon protein aggregation. The preparation of biomolecular condensates is a challenging task, and environmental-dependent effects are not yet perfectly understood. In addition, every condensate is different and might have different properties. Prior extensive physicochemical analyses of them is necessary to understand the features of a given system.

The measurement of cross-correlated relaxation rates were adapted to our biomolecular condensate. While the length of the pulse sequence and therefore the sensitivity could still be improved, the new pulse sequence effectively allows us to perform reliable measurements in our system, were sensitivity is usually not a problem given the protein concentration. It should be mentioned that such t1 noise artifacts could also be encountered in other experiments, where we believe that a first order coherence selection on the t1 evolution can effectively suppress the unwanted signal.

Incorporating client proteins in a biomolecular condensate to study biological processes is a challenging task, since it can alter the dynamic properties of the studied condensate. Incorporating a short construct into the dense phase of the long form of Ntail altered the dynamic properties of the condensate and probably its physicochemical properties. In order to study more complex

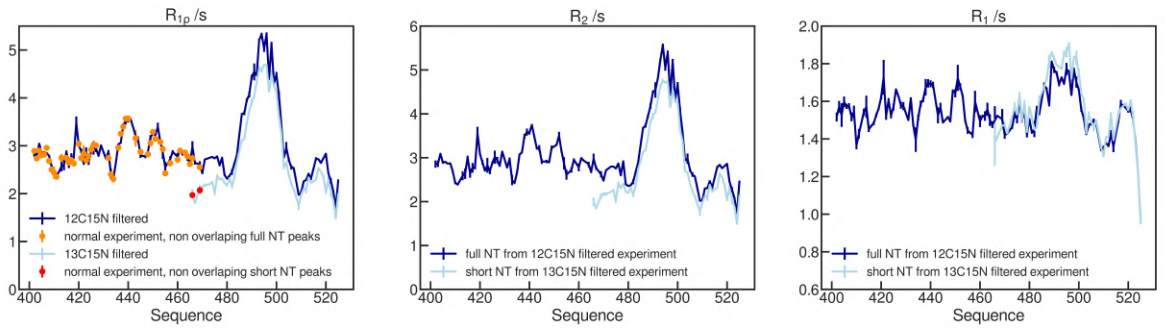


Figure 7.7: NMR spin relaxation rates of a 1:1 mixture of double-labelled $^{13}C^{15}N$ MeV Ntail 465-525 and single labelled ^{15}N MeV Ntail 402-525. Left: rotating frame ^{15}N relaxation rates of MeV Ntail 402-525 with the conventional pulse sequence (orange dots) and the isotope filtered pulse sequence (dark blue) and of MeV Ntail 465-525 with the conventional pulse sequence (red dots) and the isotope filtered pulse sequence (light blue). Center: Transverse relaxation rates of MeV Ntail 402-525 (dark blue) and MeV Ntail 465-525 (light blue) using the isotope filtered $R_{1\rho}$ and R_1 experiments. Right: Longitudinal relaxation rates of MeV Ntail 402-525 (dark blue) and MeV Ntail 465-525 (light blue) using the isotope filtered R_1 experiment

systems by NMR, one of the most exciting and challenging perspectives of the method for the future of integrated structural biology, the benchmark of the spectroscopic characteristics of diverse conditions and temperatures will be essential.

Finally, we proposed isotope-filtered relaxation experiments as a means to study complex systems in which either natural abundance contamination is present or multiple labeling schemes are necessary. These pulse sequences also allow us to study the dynamic properties of two different proteins in the same NMR sample, which can be useful in some cases. We used these pulse sequences to study the dynamics of MeV Ntail and its short construct Ntail 465-525. Studying them in the same sample ensures that both proteins experience the same experimental conditions including nanoviscosity. This allowed us to observe the subtle difference of tumbling between the long and short disordered protein chains and confirm this phenotypical observation that was already observed previously for another protein [51].

Studying challenging systems sometimes require technical solutions. In this chapter, we observed the challenges of studying biomolecular condensates by NMR and we proposed some useful spectroscopic solutions that could help the scientific community in the challenging task of exploring biological mechanisms in these highly complex systems.

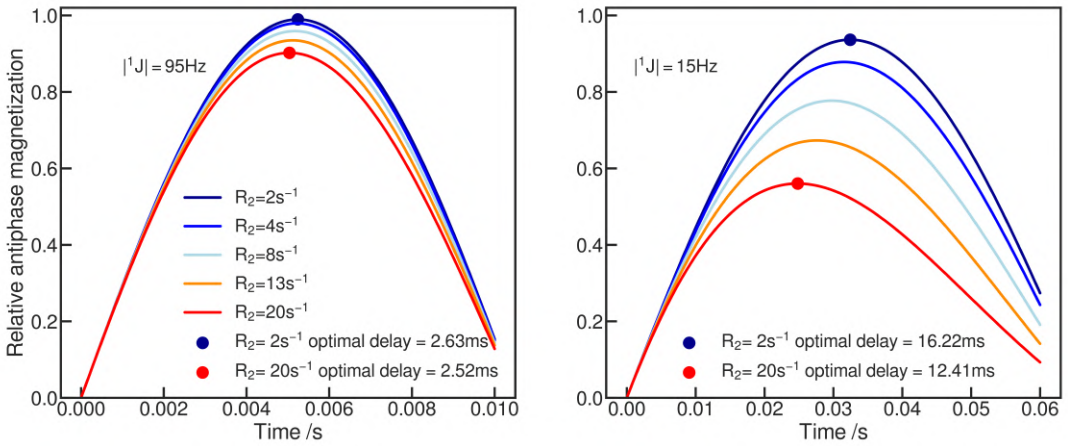


Figure 7.8: Simulation of anti-phase coherence build-up of a one half spin scalar coupled with another one half spin undergoing J-coupling evolution assuming a transverse relaxation rate of the in-phase and anti-phase coherence of 2 (dark blue), 4 (blue), 8 (light blue), 13 (orange) and 20 (red) s^{-1} for J-coupling constants of 95 (left) and 15 Hz (right). The fitted optimal half-transfer delays (1/4J if R_2=0) are indicated by blue dots for R_2=2 s^{-1} and red dots for R_2 = 20 s^{-1}.

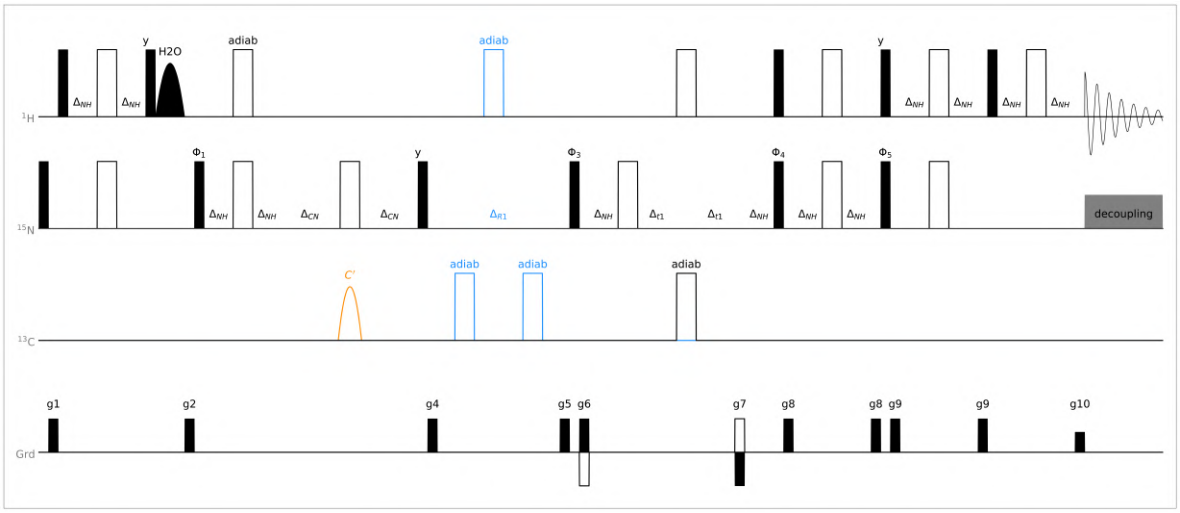


Figure 7.9: Shorter pulse sequence for protein backbone ^{15}N filtered/edited longitudinal spin relaxation rates measurement. The features are similar to the previous pulse sequence. The orange pulse on carbonyl is present in one plan and absent in the other.

Conclusion and future outlook

The different chapters of this thesis show that exploring the properties of IDPs in complex environments remains a challenging task. The two first chapters summed up the necessary theoretical background and the current topics underlying the study of IDPs. Then, we addressed the question of how are the dynamic properties of IDPs evolving upon liquid-liquid phase separation into biomolecular condensates with chapter 3 and 4, followed by a focus on the dynamic and interaction properties upon crowding in chapter 5 and 6. Finally, we addressed more methodological questions on studying complex environments with chapter 7 focused on experimental methodologies.

In the first experimental part of this thesis represented by chapters 3 and 4, we focused on studying the dynamic properties of IDPs in phase separated liquid biomolecular condensates. The C-terminal domain of the Measles virus Nucleoprotein proved to be a suitable system for this study. In chapter 3, we used NMR spin relaxation to characterize the dynamic properties of MeV Ntail in the dilute state and in the dense phase, which allowed us to perform a direct comparison and assess the effect of such a condensation on the dynamic properties of our prototypical IDP. We found that, in addition to a slow down of the rotational diffusion along the chain, a significant redistribution of the contribution of each dynamic modes to relaxation occurs upon phase separation. MD simulations of MeV Ntail at different concentrations from 0 to 20 mM were performed to interpret the experimental data. Simulations successfully reproduced the Model-Free amplitude redistribution, and a careful examination of the sequence-dependent proximity between NH vectors and neighboring atoms revealed a correlation between the fast motions order parameter and the amount of non specific self crowding in the simulation boxes. We can interpret our observations as follows: Fast motions associated with vibrations, librations and peptide plane fluctuations are found to be more restricted in the condensed phase due to the high amount of crowding, induced by the increased promiscuity in the dense phase, which as a result quenches the contribution of these motions to the spectral density function. Therefore, the dominant dynamic mode in the biomolecular condensate is the slowest, associated with chain-like segmental motions.

These MD simulations were performed on concentrated boxes of MeV Ntail with random conformations in agreement with experimental data. However, the sequence-specific intermolecular contacts were not simulated since this level of detail would have required longer simulation times exceeding the microsecond timescale. Chapter 3 focused on the global dynamic properties of MeV Ntail. Although some sequence specific features were already observed in the NMR relaxation profiles, we studied further the dense phase of MeV Ntail in chapter 4 and extensively studied a robust MeV Ntail condensed phase with a standardized protocol. In this chapter, we focused rather on the sequence-specific dynamic modulations of MeV Ntail upon phase separation in the light of the well known intermolecular interactions that stabilize biomolecular condensates. We found that the sequence dependence of the transverse relaxation rates profile, characterized by several abnormally elevated values, could not be interpreted by only a systematic effect such as

viscosity unlike previous studies [284]. A local slow down of segmental and peptide plane motion is observed in the vicinity of these so-called hot spots, correlated with the presence of highly charged and aromatic residues. On the other hand, no sequence specificity was observed in the modulation of fast motion upon liquid-liquid phase separation.

In the light of our observations on the dynamic properties of MeV Ntail in the dense phase, we propose that the backbone dynamics of our protein was modified in three key aspects. First, a significant increase of the viscosity of the solution in the dense phase leads to a significant slow down of the rotational correlation times at multiple timescales. We believe this slow down to be length-scale dependent in analogy with previous findings using polymer crowders [52, 53]. However, this effect couldn't be quantified because of the unreliable measurements of longitudinal relaxation rates of water protons. To obtain reliable solvent frictions in the dense phase, another strategy should be used. An idea would be for example to use a small rigid inert molecule with a convenient spin system as a proxy for solvent friction. The second aspect is the significant and non negligible self crowding experienced in the dense phase. A high level of promiscuity in the solution induces more restricted motional processes all along the sequence. As a result, an increase in the effective order parameters is observed upon Model-Free analysis of NMR spin relaxation rates in the condensed phase. We should mention that here, only two magnetic fields were used. It can be of interest to measure a larger relaxation data set with a broader range of magnetic field to obtain more insight into these features. Finally, the last aspect is the sequence-specific modulation of the local backbone tumbling due to intermolecular interactions. Our results indeed suggest that intermolecular interactions, occurring mostly in the side chains (but not only) induce a local slow down of the chain-like motion in the vicinity of the important hot spots for phase separation. Although these intermolecular interactions were not entirely characterized here, another explanation to these observations is difficult to find given the high amount of correlation between the residues known to be involved in such interactions and the regions with elevated transverse relaxation rates. A more detailed investigation could be performed in our dense phase via NOESY experiments provided that an appropriate dynamic range is found for efficient NOE transfers. Alternatively, MD simulations with an appropriate model can support the hypothetical presence of these contacts.

Liquid-liquid phase separation, although ignored for decades, is a fundamental process in biology. While it adds a new complexity layer to the already complex physiological environments, the study of the dynamic properties of IDPs in such environments remained accessible with experimental techniques such as NMR spectroscopy [53] and single molecule fluorescence microscopy [330], with the help of MD simulation as a complementary technique. Most recent studies, including this one, focused on model systems. However, it might be interesting to attempt to study molecular mechanisms in biologically relevant biomolecular condensates. In our case, in the context of the Measles Virus replication machinery, this would first imply the overproduction of MeV N and MeV P to form a macrodroplet, which is already a challenge in itself. Then, it would require the obtention of observable signal in such mixture of high molecular weight IDPs. Even in the dilute state, obtaining a highly resolved spectrum of MeV N is challenging. Therefore, such study would probably imply the use of expensive methods such as selective labeling, labeling only one region of the IDP and perdeuteration, which is not a guarantee of success since we also need our IDP to tumble sufficiently rapidly by itself to yield reasonable transverse relaxation rates. Nevertheless, NMR already proved to be a unique method for characterizing the very important intermolecular interactions that stabilize these condensates, and it is not excluded that it will continue to be useful for the study of important biological mechanisms involving protein-protein or protein-ligand interactions. It is likely that in the future, membrane-less organelles will be a valuable target for pharmaceutical studies given the ubiquitous presence of these com-

partments in eukaryotic cells and viral replication systems. Appropriate experimental methods will be necessary to perform assays in these systems, and it is likely that NMR will play a key role in these studies, in addition to computational methods.

Additionally, it can be interesting to combine different methods to obtain structural and dynamical information in biomolecular condensates to derive accurate structural and MD trajectory ensembles in addition to obtain more complete information, to adjust force field parameters for instance, and more generally to better understand the behavior of proteins in condensates across multiple time- and length-scales. A lot of work is still necessary in this direction, and it is likely that current and future studies will go in this direction.

In the second experimental part of our thesis, we focused on the dynamic and interaction properties of IDPs in polymer crowders, in order to better understand how crowding affects the internal dynamics of IDPs, as well as protein-protein interactions. In chapter 5, we studied the backbone dynamics of MeV Ntail 465-525, a short construct of the MeV Nucleoprotein C-terminal domain that doesn't phase separate upon addition of PEG. We used this construct as a model protein for further studies in semi-dilute and concentrated crowding regimes. Our analysis of the NMR spin relaxation rates revealed that at very high concentrations of our polymer, the dynamic properties of MeV Ntail 465-525 change in a way akin to what we observed with MeV Ntail upon liquid-liquid phase separation. More specifically, we observed a redistribution of the contributions of each dynamic modes in the Model-Free analysis, suggesting more restricted motion in extreme crowding. These findings show that LLPS is not necessary for these amplitude modulations to be observed. Rather, they are possible provided that a sufficiently significant amount of polymer crowding is present in the system. While a lot of cellular environments don't present sufficiently high proportions of disordered proteins in their environments to exhibit such behavior, it is possible that some highly crowded environments such as the cytoplasm of bacteria like *E. coli* might be adapted to this specificity. In any case, this study confirms the previous observation that a crowding mechanism was responsible for this redistribution, and show how a concentrated regime of polymer solution affects the dynamic properties of Intrinsically Disordered Proteins.

In chapter 6, we focused on the interaction properties of IDPs in crowded environments. Interactions are key to all biological processes, and crowding might play a role in the properties of these interactions. To better understand the role of crowding, we used the interaction between MeV Ntail 465-525 and PXD, the XD domain of MeV P as a model. The advantage of NMR over other techniques like fluorescence microscopy here is the obtention of atomic-resolution information. We used NMR chemical exchange techniques to characterize the interaction between these two proteins. Experiments were performed on both sides of the interaction and showed consistent results. The results show that the interaction is slowed down upon crowding, which can be expected from a higher viscosity. Additionally, the chemical shift difference between free and bound state is barely affected by the concentration of crowder, suggesting that the topology of the complex between the two proteins is not strongly affected by crowding. More experimental data should be obtained at different proportions of the proteins in order to obtain more detailed insight into the modulation of the affinity of the interaction, as well as the on and off rate. A comparison with fluorescence studies of other interactions might be interesting, in the light of recent theories considering the coexistence of depletion interactions and increased viscosity in the mechanisms that modulate protein-protein interactions in crowded environments [250]. Multi-nuclear chemical exchange experiments should also provide a more detailed insight into the molecular trajectory MeV Ntail follows upon interaction with PXD, and how this mechanism is affected by macro-molecular crowding.

Since protein interactions are involved in almost all biological processes, it is crucial to understand all the environmental sources that can modulate their kinetic and thermodynamic properties. An incorrect estimation of these parameters can have an important effect, notably in the context of pharmaceutical studies or in the study of processes where the dynamic properties of the interactions have an impact on the underlying biological mechanism. A lot of experimental studies are still necessary to understand all the environmental factors and their effect on protein dynamic and interaction properties. It is possible that the system is so complex that a mathematical model becomes necessary to compute the modulations of the properties of a given interaction in specific environmental conditions. The current rise of machine learning techniques should provide the tools for such prediction framework, but this requires the obtention of a massive amount of experimental data that so far don't exist. Even with the availability of such method, it remains important to understand the physical processes underlying these environmental modulations.

The last part of the thesis explored more methodological aspects of the subject. In chapter 7, we addressed some experimental aspects of studying IDPs in biomolecular condensates and other complex environments. First, we explored the properties of MeV Ntail to form various biomolecular condensates with different properties depending on the initial conditions. Then, we attempted to study the properties of client proteins in biomolecular condensates before implementing an NMR pulse sequence for transverse cross-correlated relaxation rates that doesn't produce t1-noise artifacts. Finally, we proposed pulse sequence schemes for isotope-filtered/edited NMR spin relaxation rate measurements in complex environments were natural abundance contamination is present. This chapter illustrates the necessity of developing innovative methods for studying proteins in more complex environments. While nothing revolutionary was invented here, the study of large and complex systems will require the use of sophisticated labeling schemes and smart NMR experiments with appropriate encoding. It is therefore important to carefully identify which method is the most appropriate to study complex systems.

To conclude this thesis, we saw how a prototypical IDP is affected by liquid-liquid phase separation, as well as various levels of polymer crowding. We also addressed the question of the effect of crowding on protein-protein interactions. Finally, we addressed different methodological challenges in studying complex systems. In the light of this work, we further showed that NMR and Molecular Dynamics simulations are unique tools for characterizing IDPs at atomic resolutions, from the dynamic properties to their interactions with other partner. There is no doubt that these two techniques will keep providing unique insight into the properties of IDPs in complex environments, from the cellular cytosolic environment to the ubiquitous liquid-liquid phase separated membraneless organelles.

Appendix A

Supplementary figures

A.1 Chapter 3

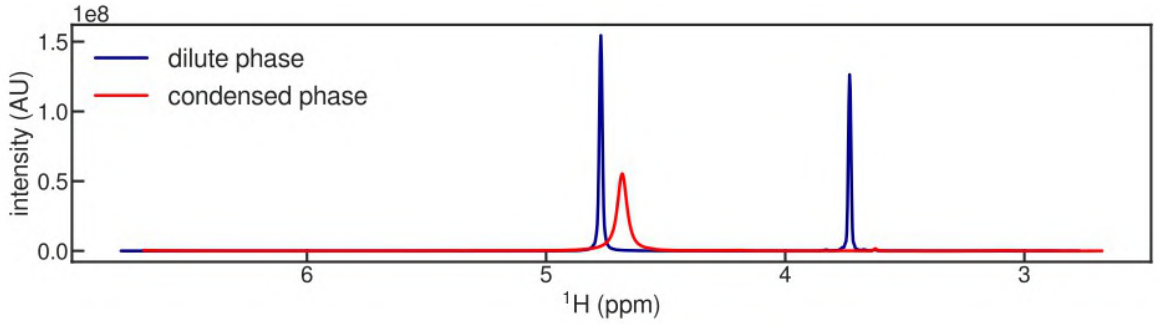


Figure A.1: 1D proton NMR spectra of MeV Ntail condense phase (red) and dilute phase (blue). The signal around 4.7 ppm comes from water, the signal around 3.7 ppm comes from PEG.

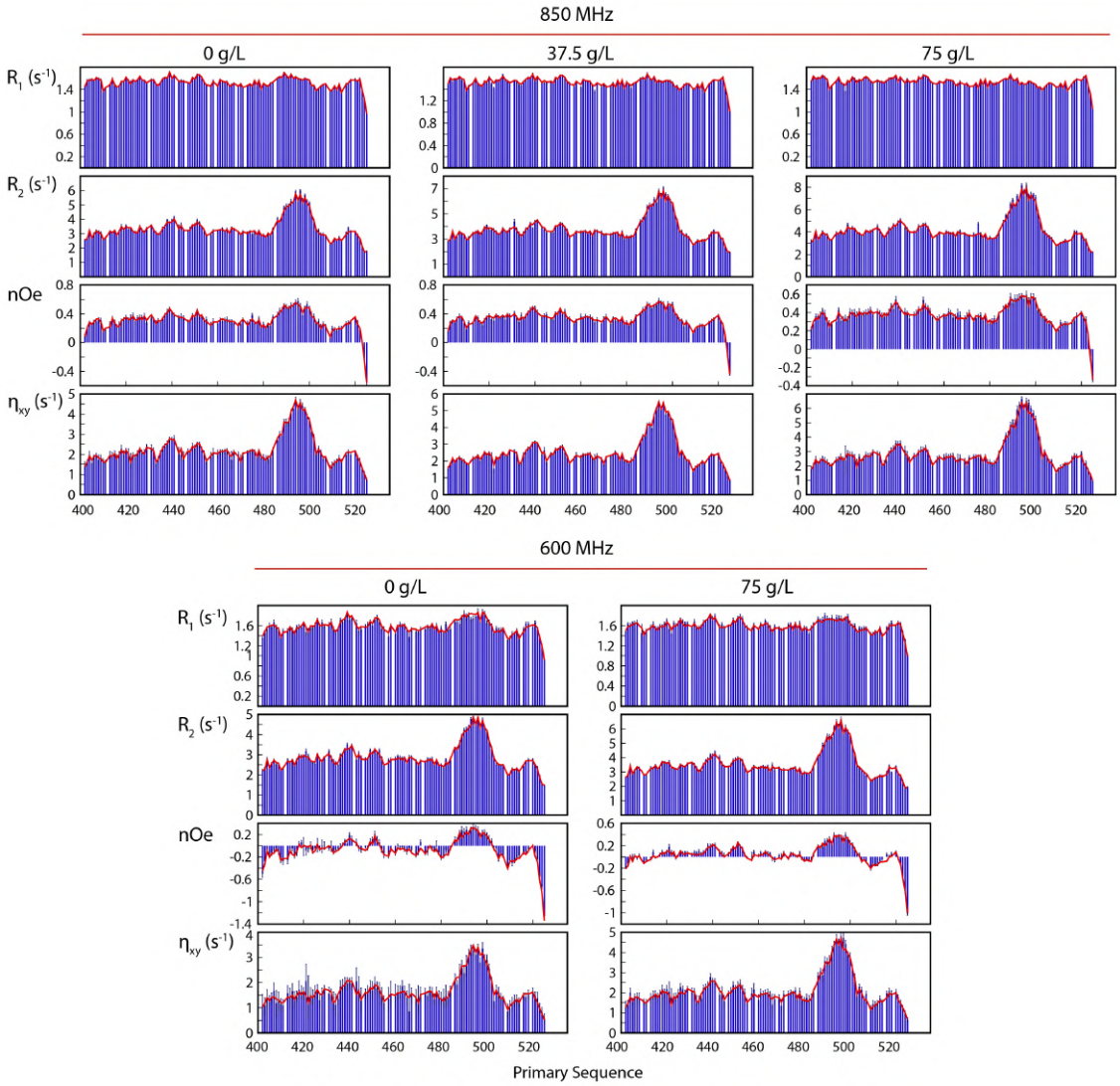


Figure A.2: Experimental (Blue bars) and calculated relaxation rates from Model-Free analysis (red lines) of MeV Ntail in the dilute state with 0, 37 and 75 g/L of PEG at 600 and 850 MHz proton frequency.

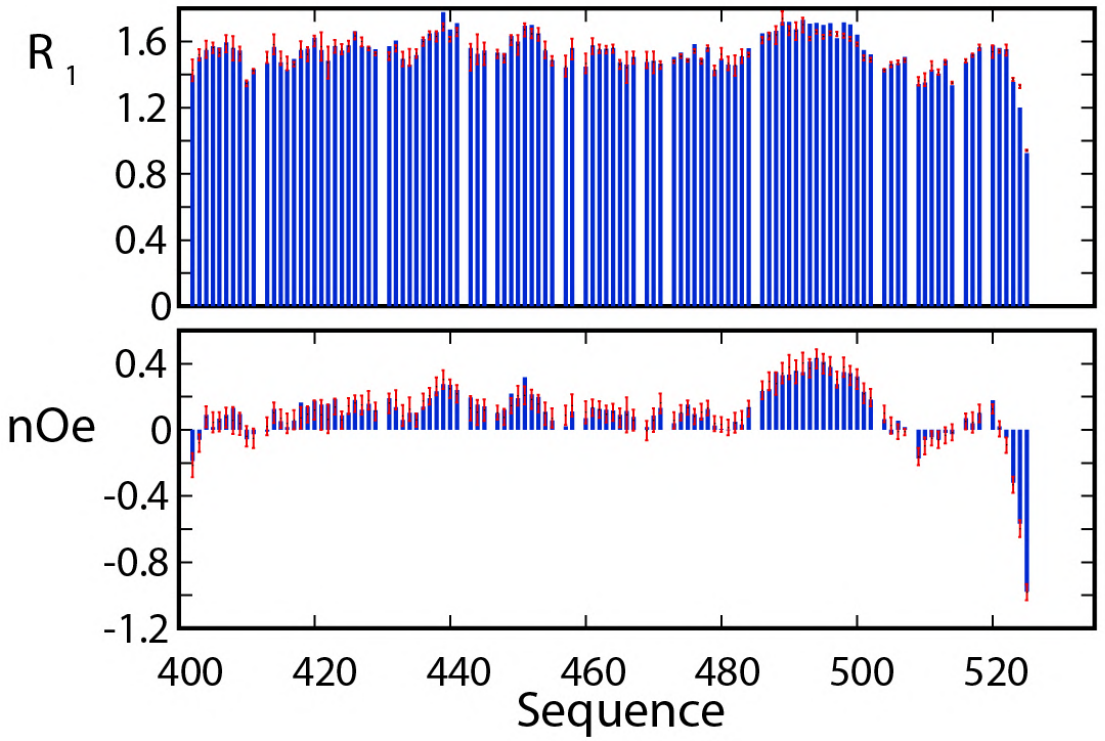


Figure A.3: Experimental (Blue bars) and calculated relaxation rates from Model-Free analysis without including these rates (red lines) of MeV Ntail in the dilute state at 700 MHz proton frequency.

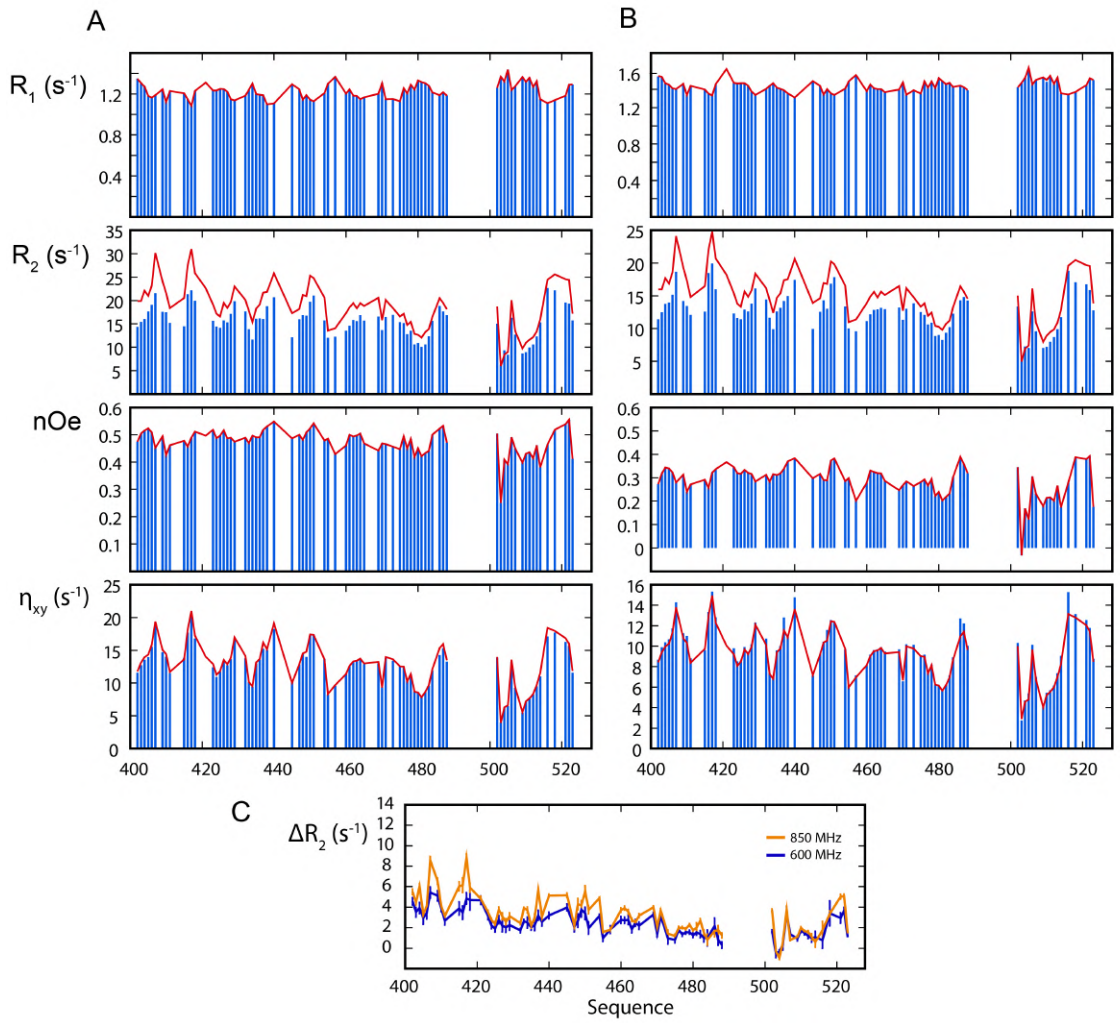


Figure A.4: A and B: Fitted (blue bars) and experimental (red lines) relaxation rates at 850 MHz (A) and 600 MHz (B) proton frequency. The fit was performed with the R_1, nOe and η_{xy} values at both fields. C: difference between experimental and back calculated transverse relaxation rate at 850 MHz (orange) and 600 MHz (blue)

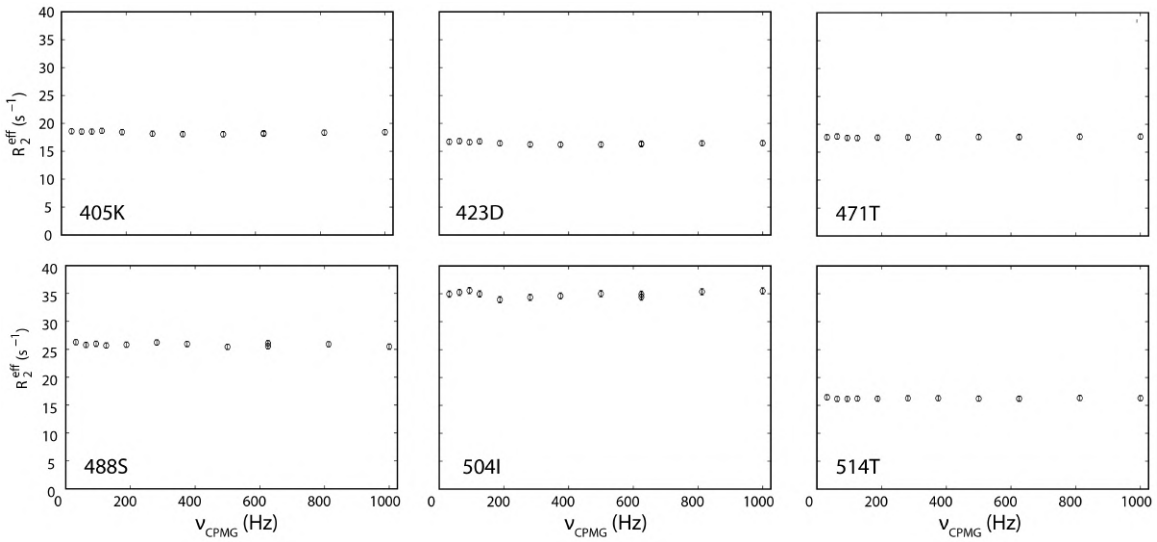


Figure A.5: A and B: Fitted (blue bars) and experimental (red lines) relaxation rates at 850 MHz (A) and 600 MHz (B) proton frequency. The fit was performed with the R_1, nOe and η_{xy} values at both fields. C: difference between experimental and back calculated transverse relaxation rate at 850 MHz (orange) and 600 MHz (blue)

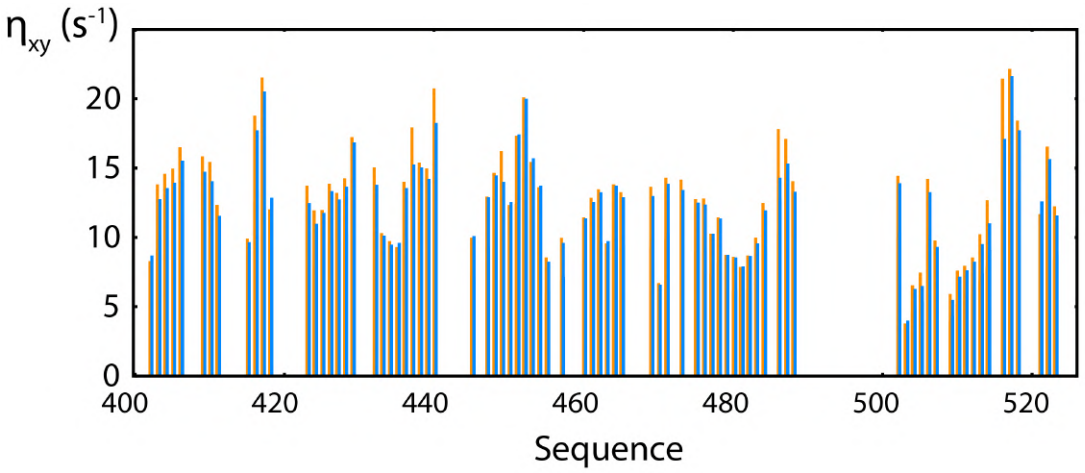


Figure A.6: Experimental (Blue bars) and calculated (Orange bars) transverse cross correlated relaxation rate at 600 MHz proton frequency. The rate was calculated from a Model-Free analysis of the data set without including η_{XY} at 600 MHz proton frequency in the fit for cross validation.

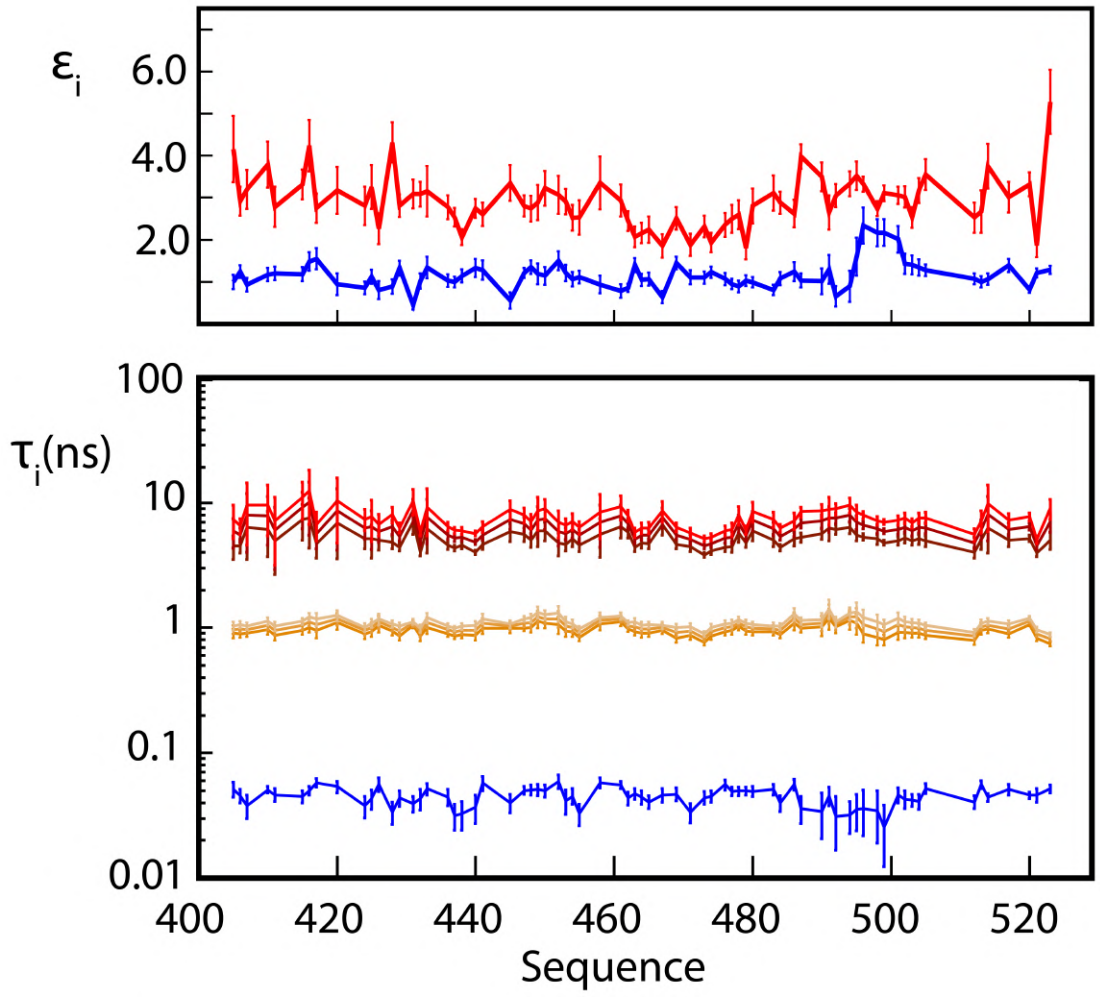


Figure A.7: Model-Free parameters of MeV Ntail in dilute and crowded conditions. Top: Residue specific solvent frictions of the slow (red) and intermediate (blue) dunamic mode. Bottom: Correlation times of each dynamic modes

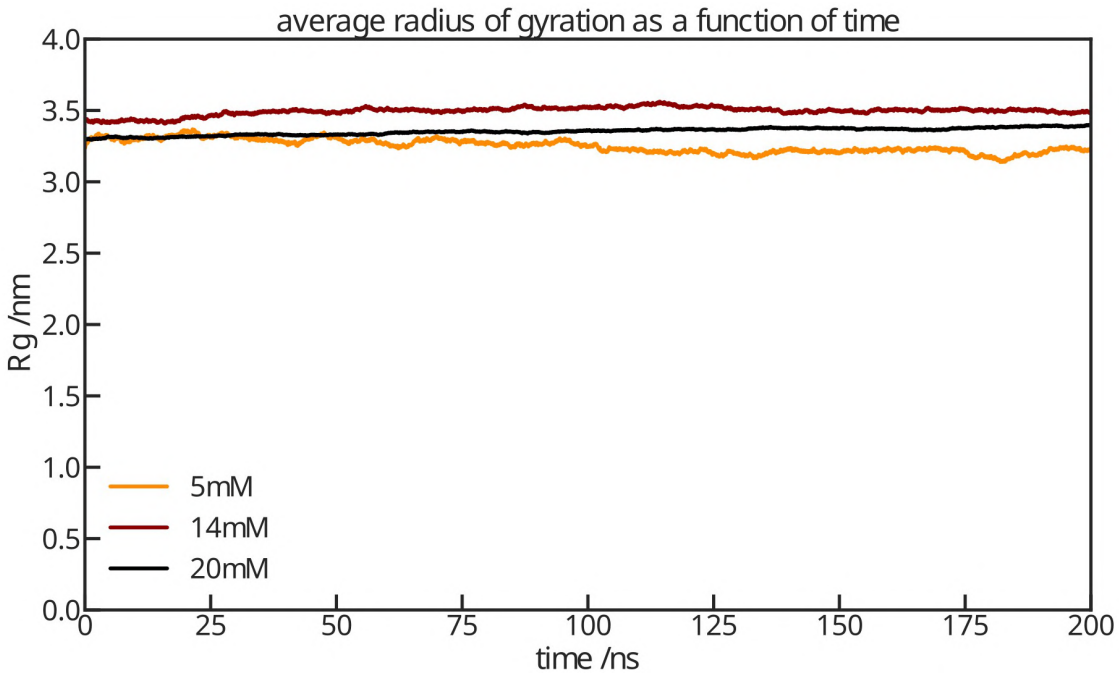


Figure A.8: Evolution of the average radius of gyration of MeV Ntail over time in the simulations at 5 (orange), 14 (red) and 20 mM (black).

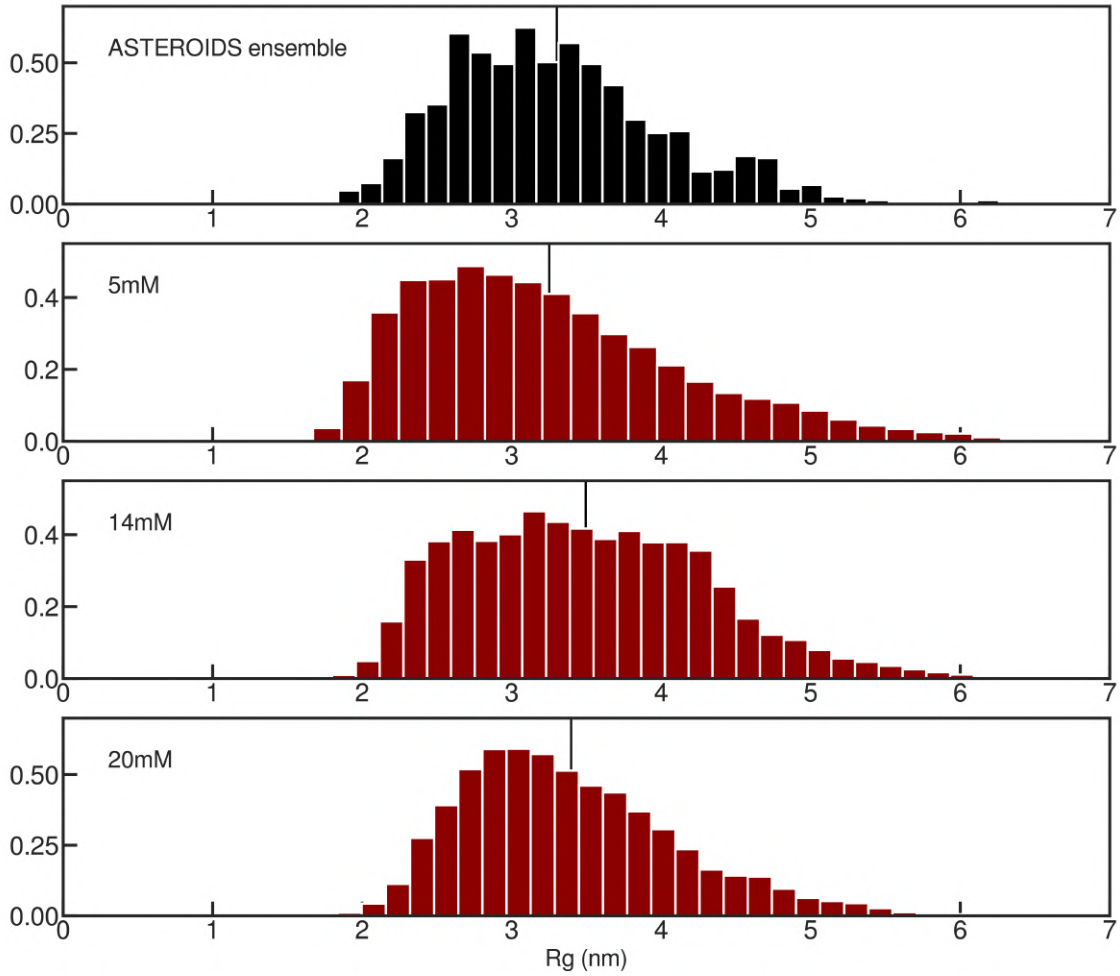


Figure A.9: Radius of gyration distribution in MeV Ntail's ASTEROIDS ensemble (top) and in the concentrated Ntail simulations at 5, 14 and 20 mM respectively from top to bottom (red histograms). The average value is indicated with a vertical black bar

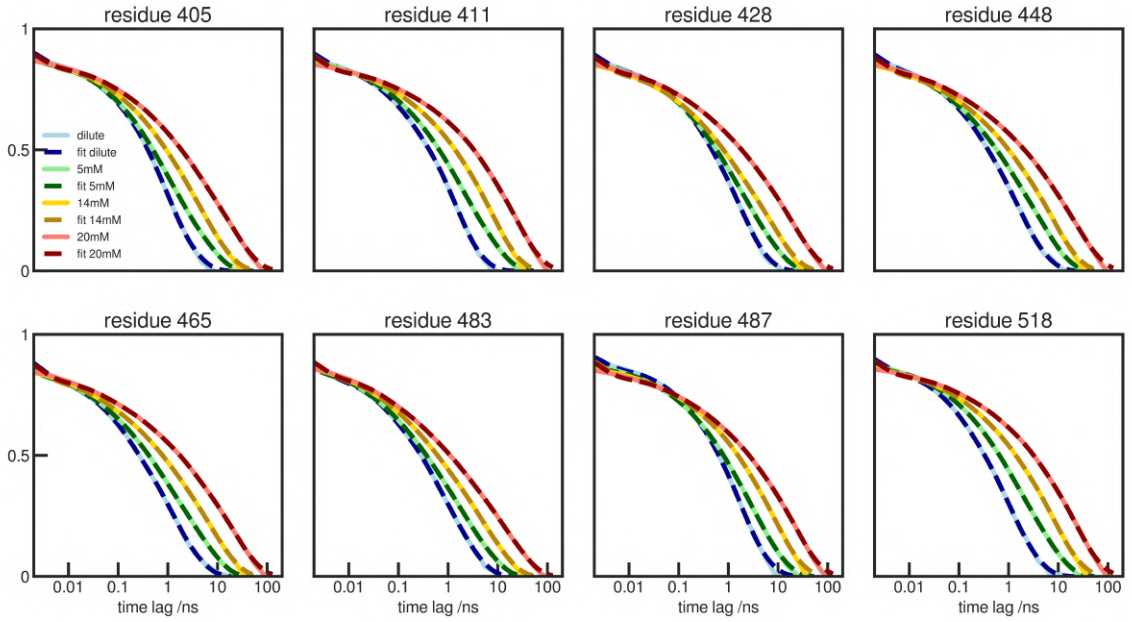


Figure A.10: calculated (continuous lines) and fitted (dashed lines) N-H bond vector rotational time auto-correlation functions in randomly selected residues of MeV Ntail's simulations in the dilute state (blue), at 5 mM (green), at 14 mM (orange) and at 20 mM (red).

A.2 Chapter 4

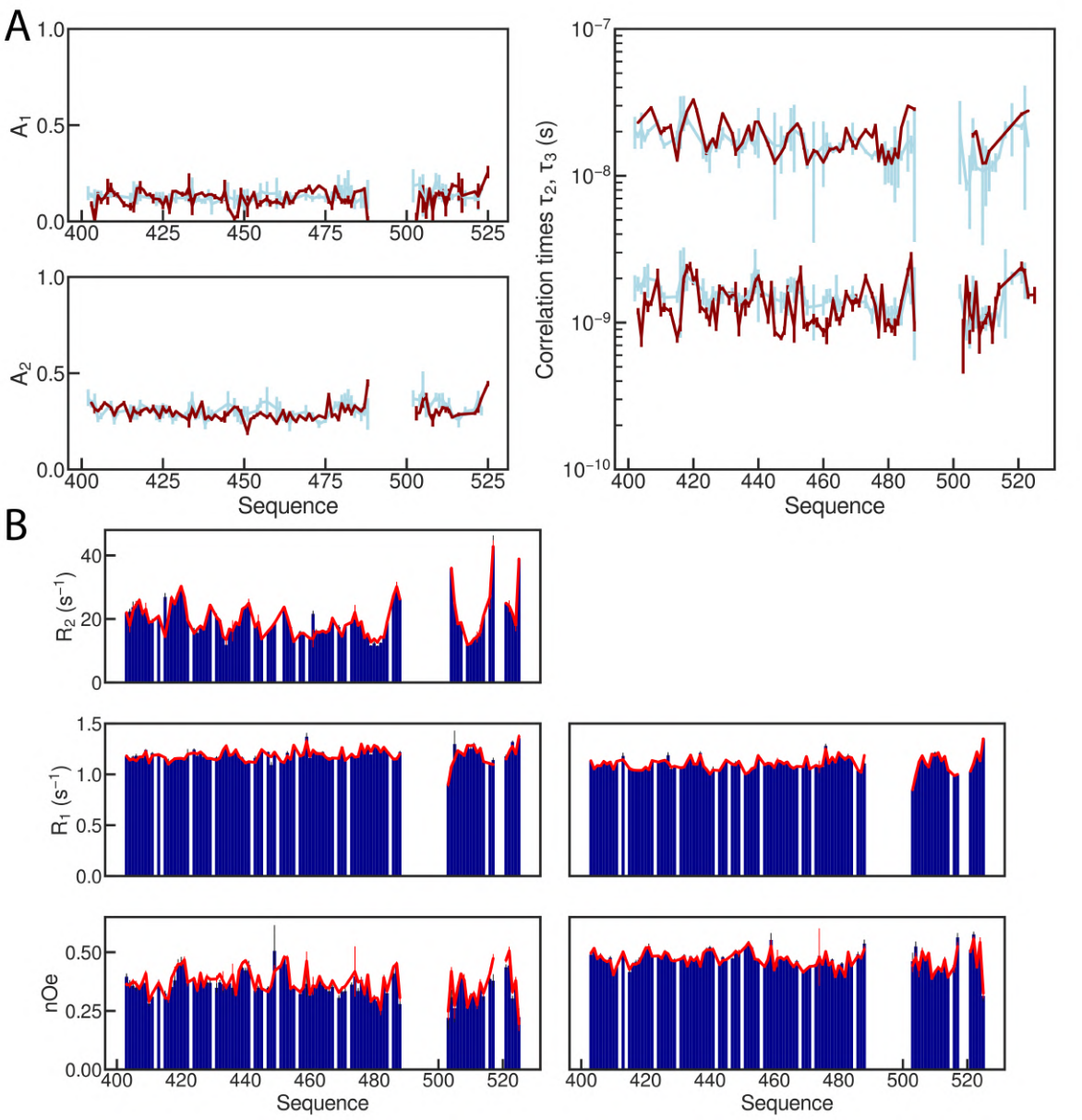


Figure A.11: A: Model-Free parameters of MeV Ntail in the dense phase prepared in Guseva & Schnapka *et al.* (Light blue) and prepared as described here (Dark red). B: Fitted relaxation rates (Red lines) compared with experimental relaxation rates (Blue bars) for the model-free analysis in the condensed phase. The fir was performed by fixing the fast correlation time to 100 ps and by neglecting hypothetical exchange contributions to R_2.

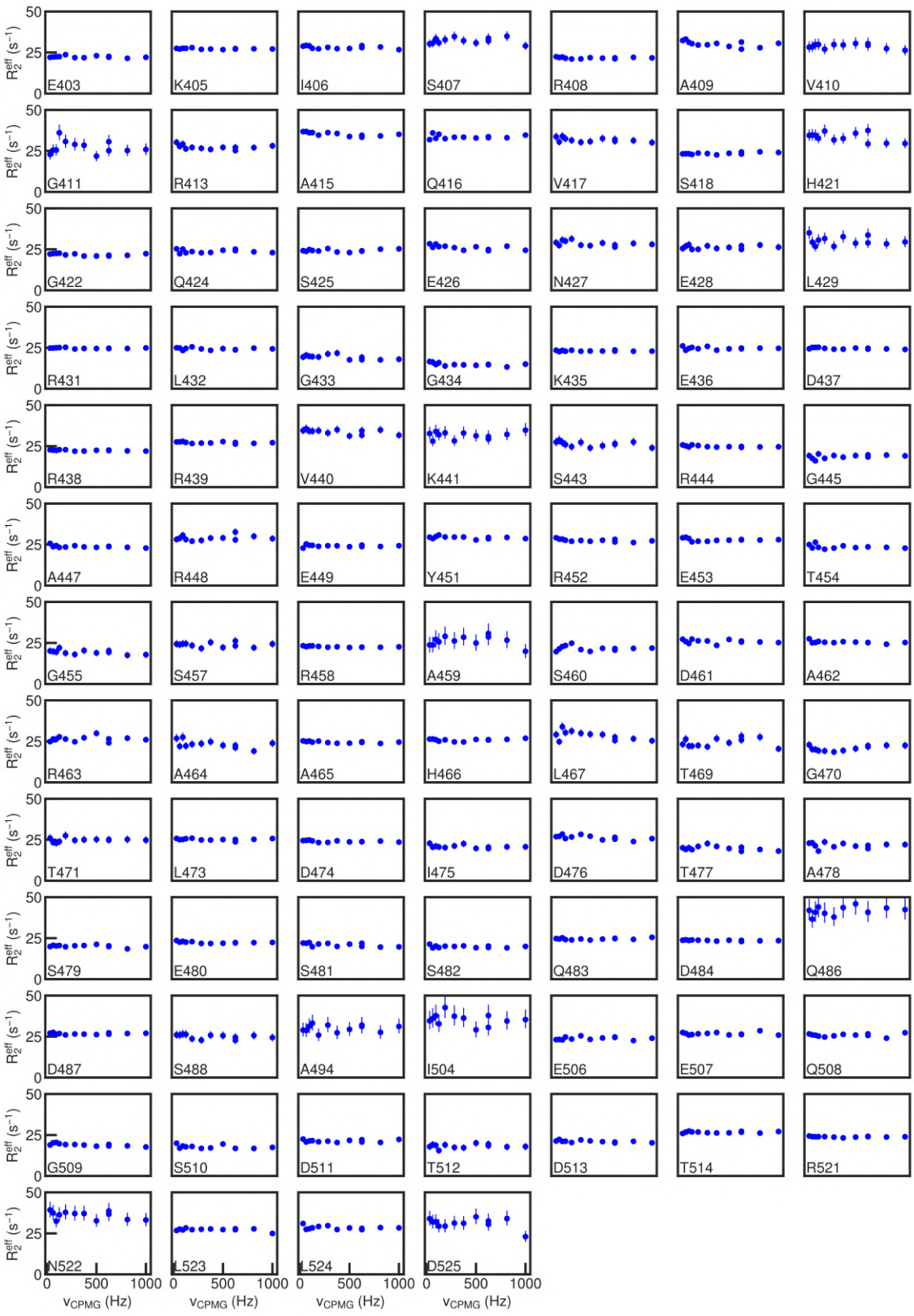


Figure A.12: CPMG relaxation dispersion profiles of each visible residues in MeV Ntail condensate.

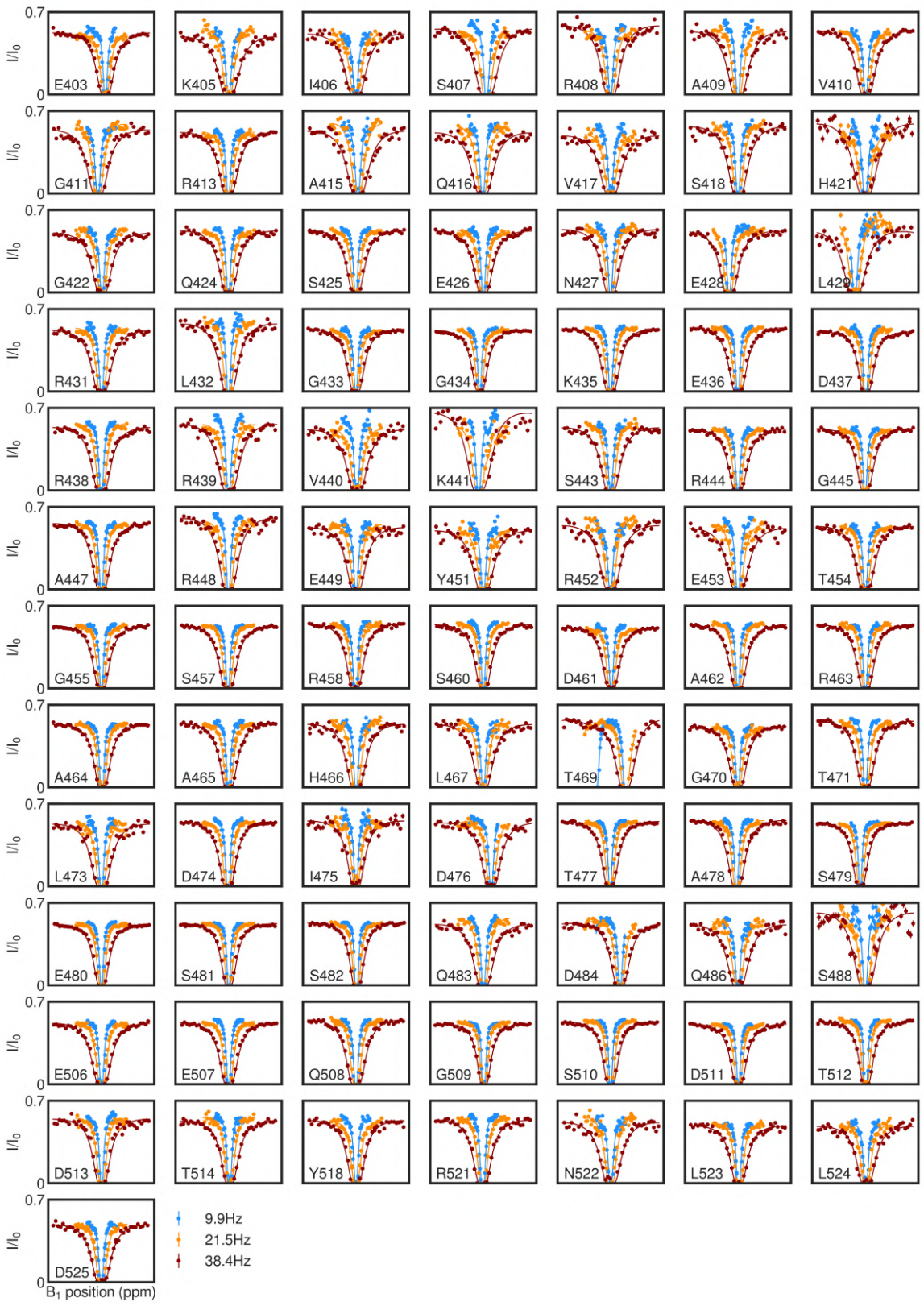


Figure A.13: DANTE-CEST profiles of each visible residues in MeV Ntail condensate, at B1 fields calibrated around 10 Hz (Blue), 20 Hz (Orange) and 40 Hz (Red).

A.3 Chapter 5

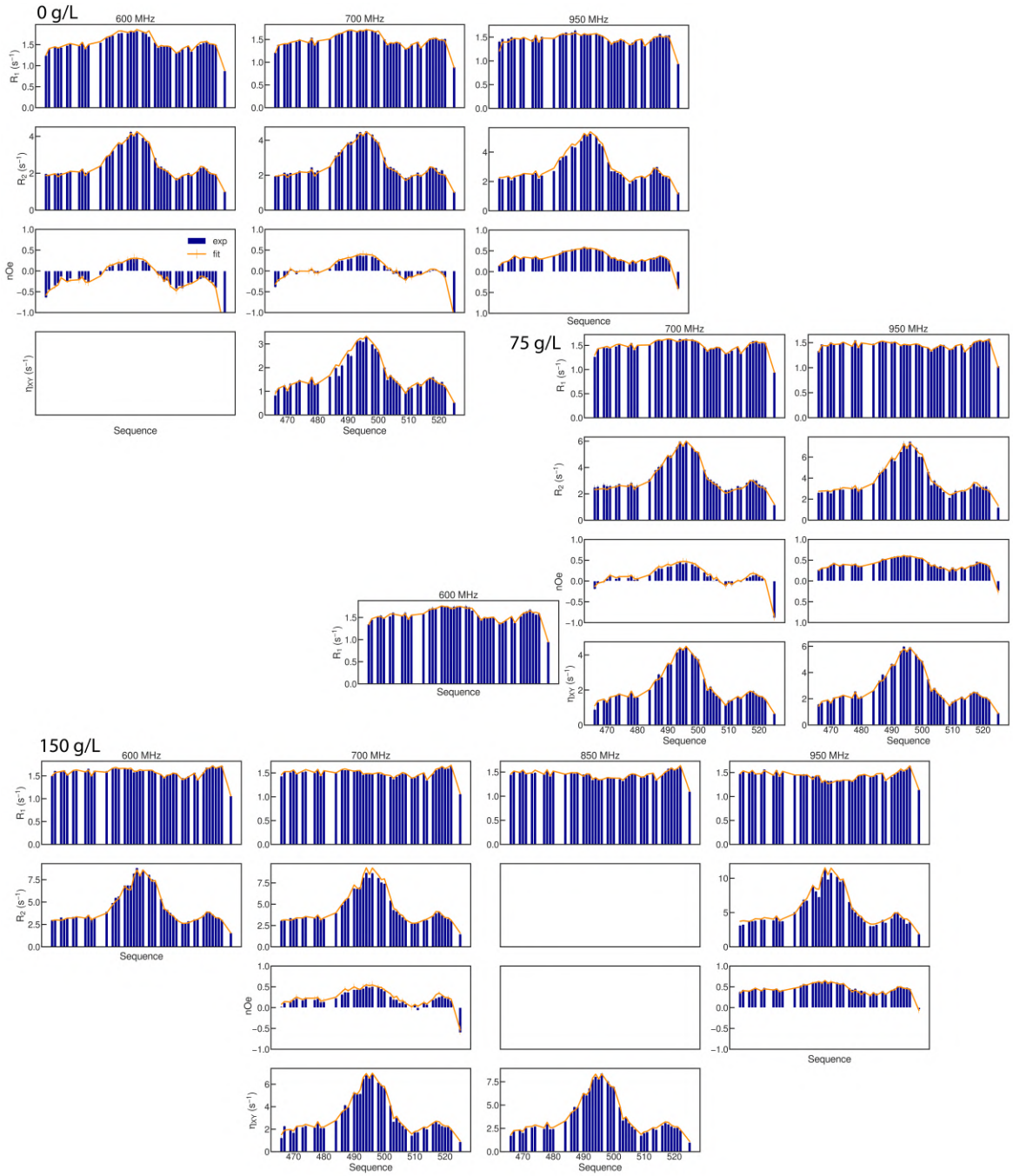


Figure A.14: Experimental (orange lines) and fitted (blue bars) relaxation rates from the viscosity Model-Free analysis at 298.1 K from 0 to 150 g/L PEG.

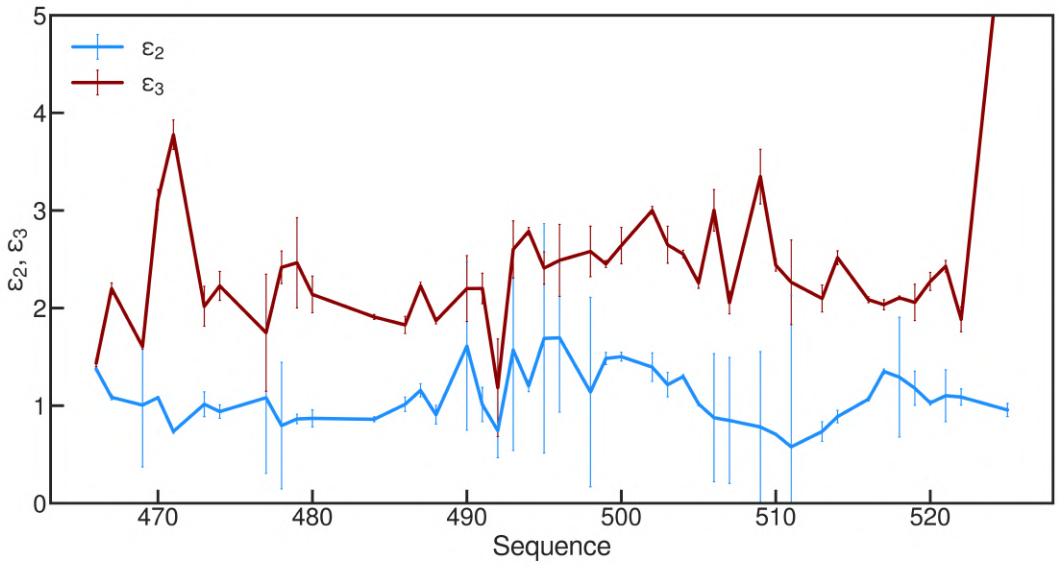


Figure A.15: Fitted friction coefficients from viscosity Model-Free analysis from 0 to 150 g/L PEG.

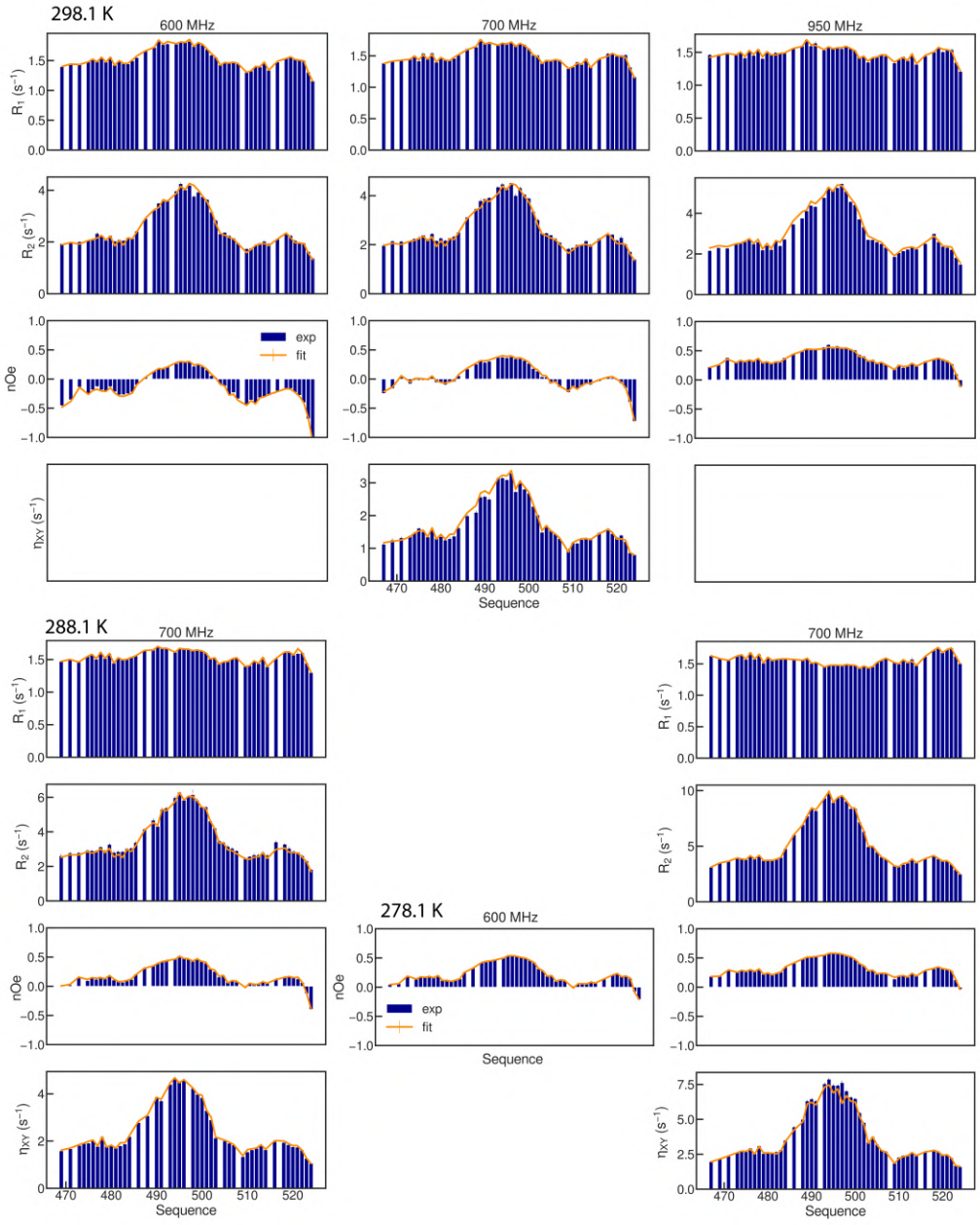


Figure A.16: Experimental (orange lines) and fitted (blue bars) relaxation rates from the Arrhenius Model-Free analysis with 0 g/L PEG.

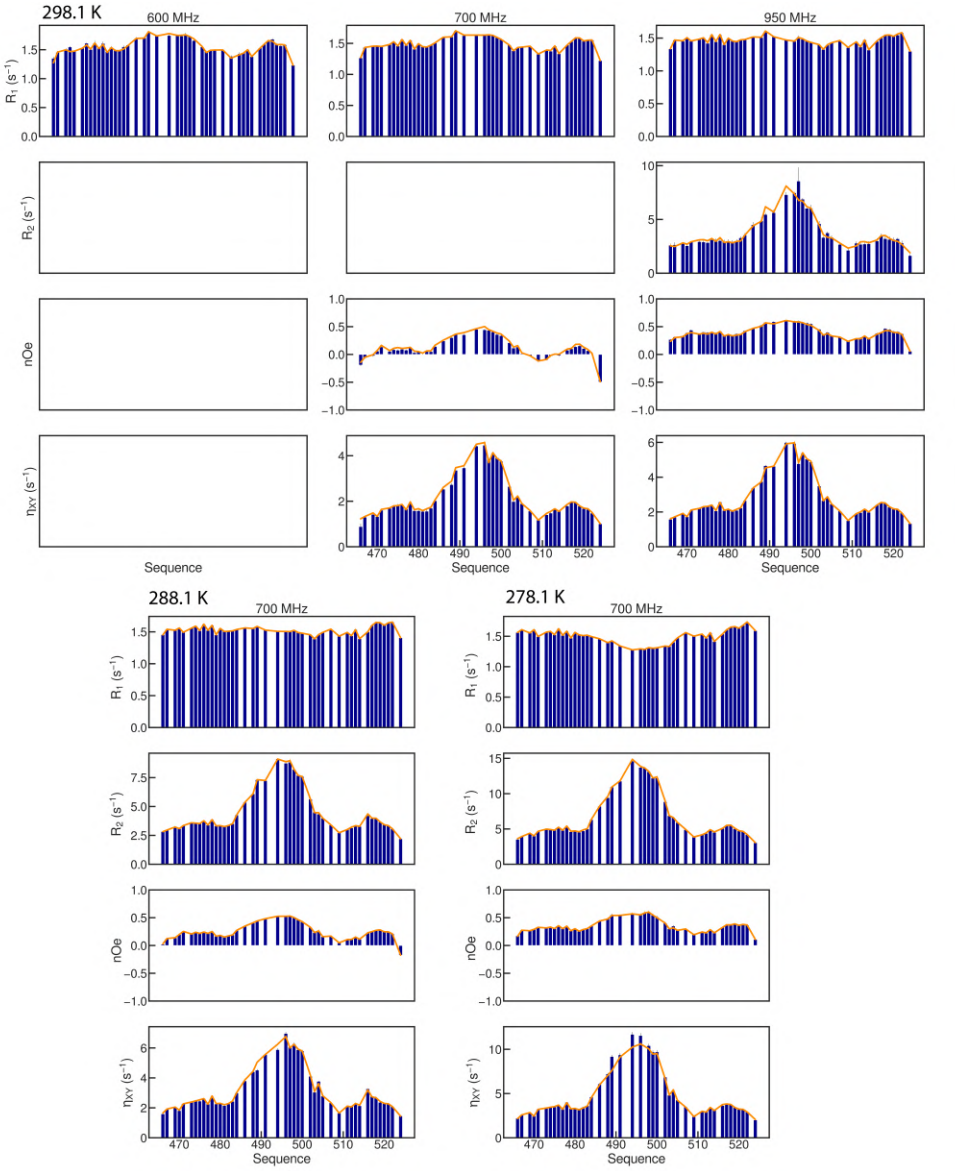


Figure A.17: Experimental (orange lines) and fitted (blue bars) relaxation rates from the Arrhenius Model-Free analysis with 75 g/L PEG.

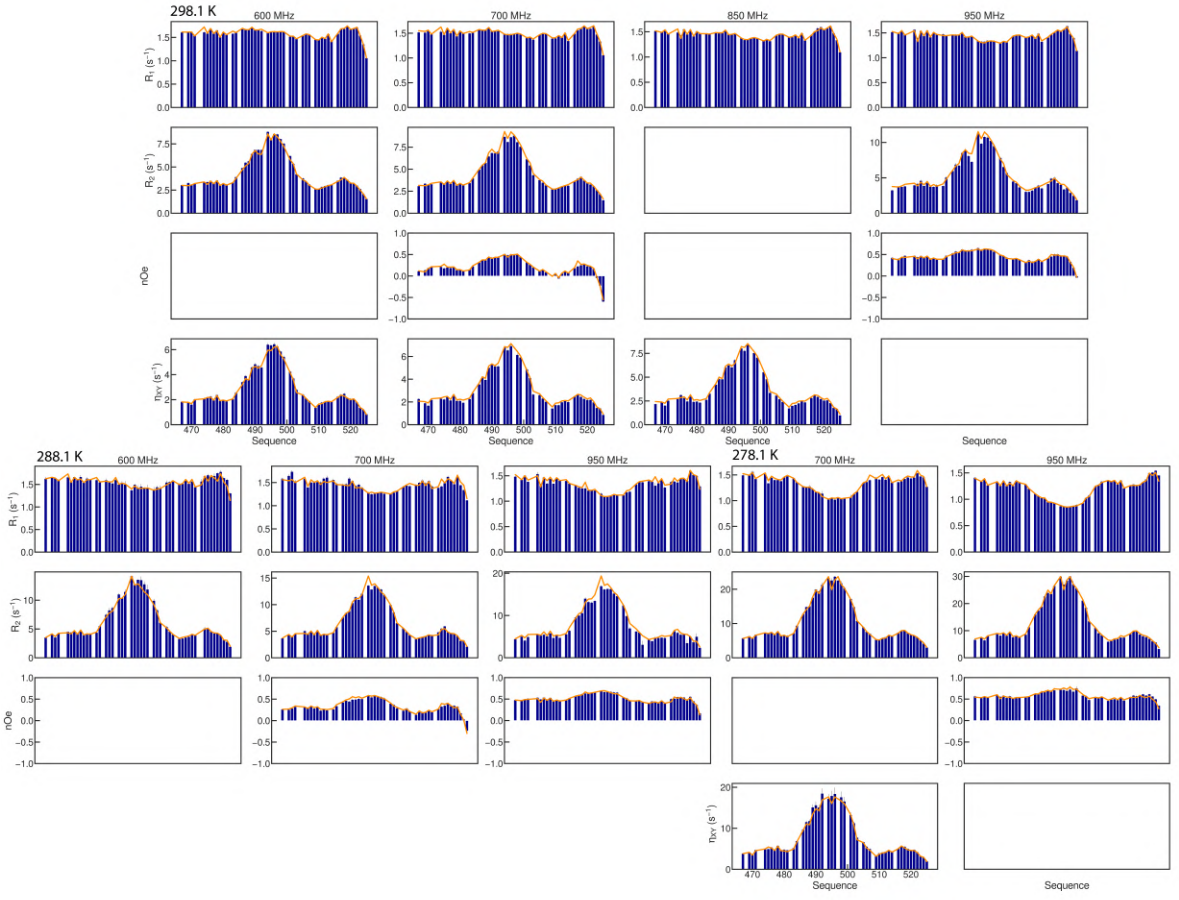


Figure A.18: Experimental (orange lines) and fitted (blue bars) relaxation rates from the Arrhenius Model-Free analysis with 150 g/L PEG.

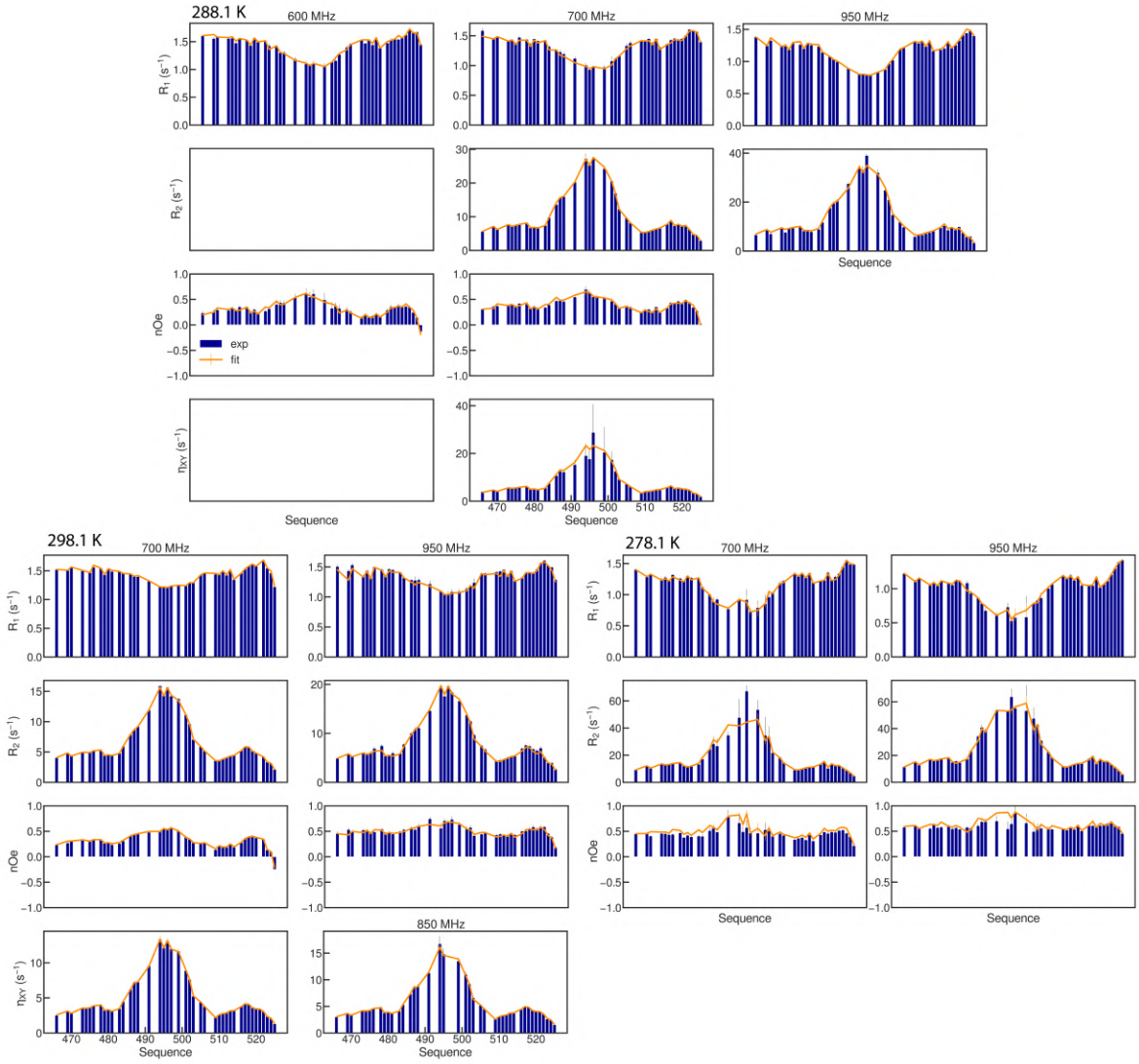


Figure A.19: Experimental (orange lines) and fitted (blue bars) relaxation rates from the Arrhenius Model-Free analysis with 250 g/L PEG.

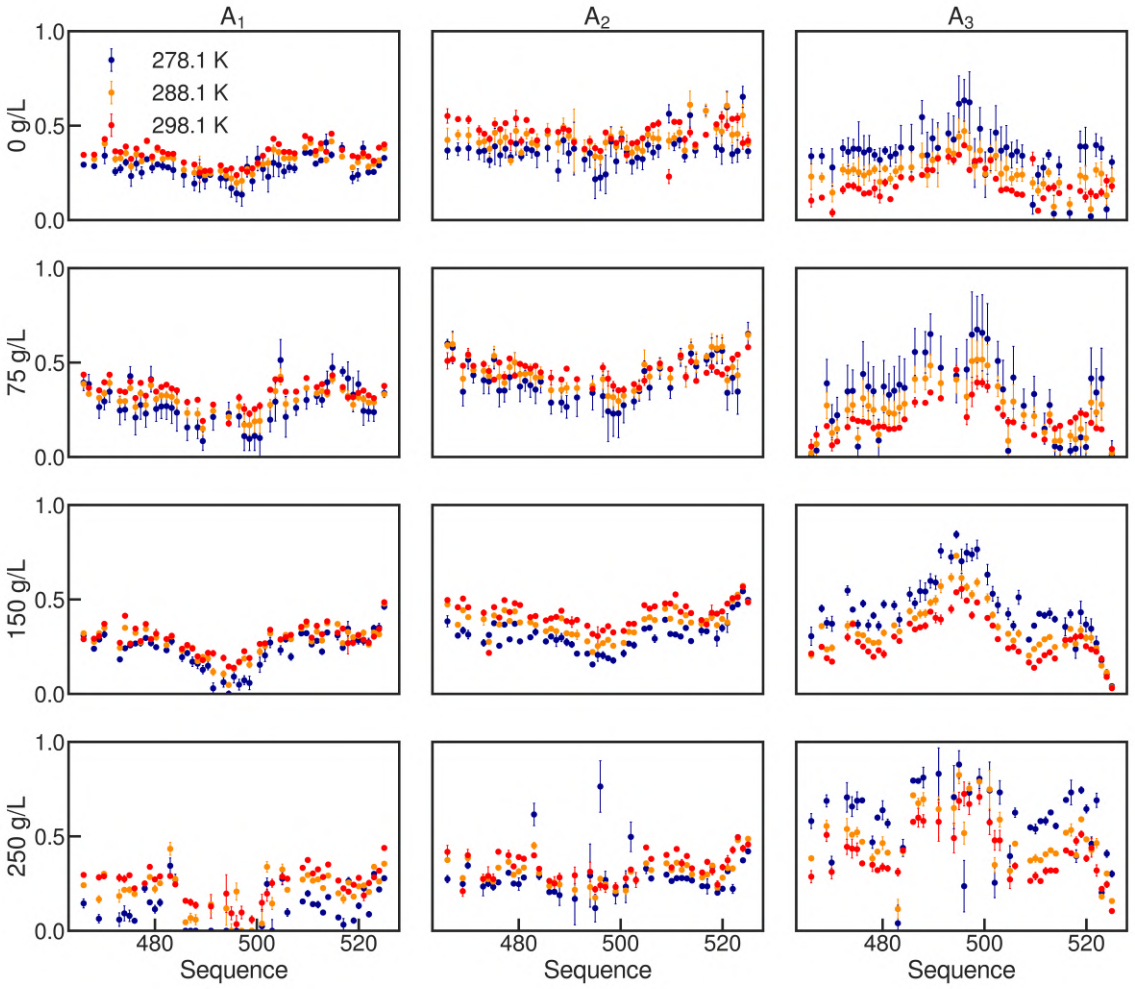


Figure A.20: Temperature dependence of the dynamic modes amplitudes from 0 to 250 g/L of PEG.

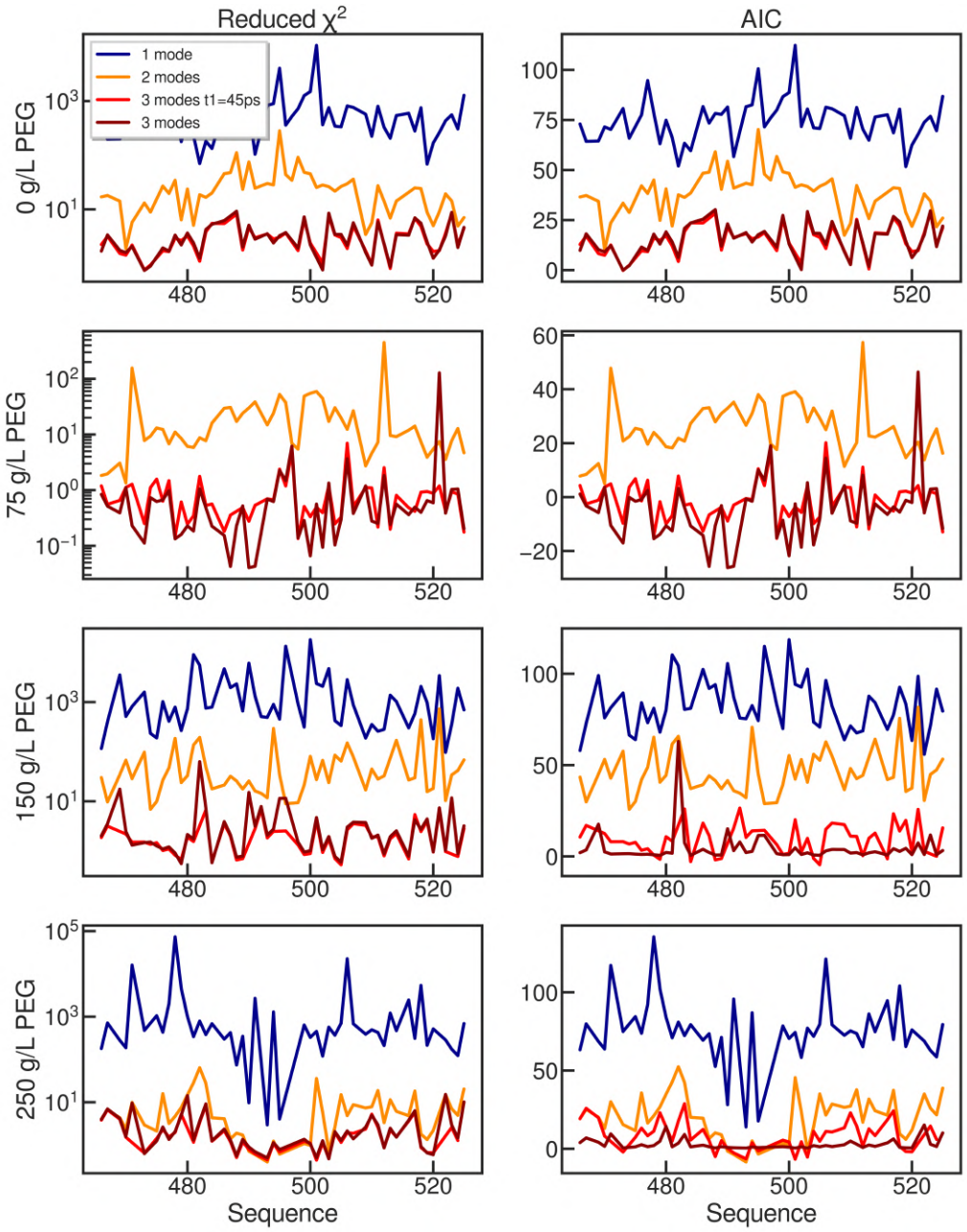


Figure A.21: Reduced chi-square (left) and Akaike Information Criterion (AIC, right) over the sequence associated with Model-Free analysis of MeV Ntail 465-525 from 0 to 250 g/L of PEG (top to down) with 1 (blue), 2 (orange) and 3 (red) dynamic modes.

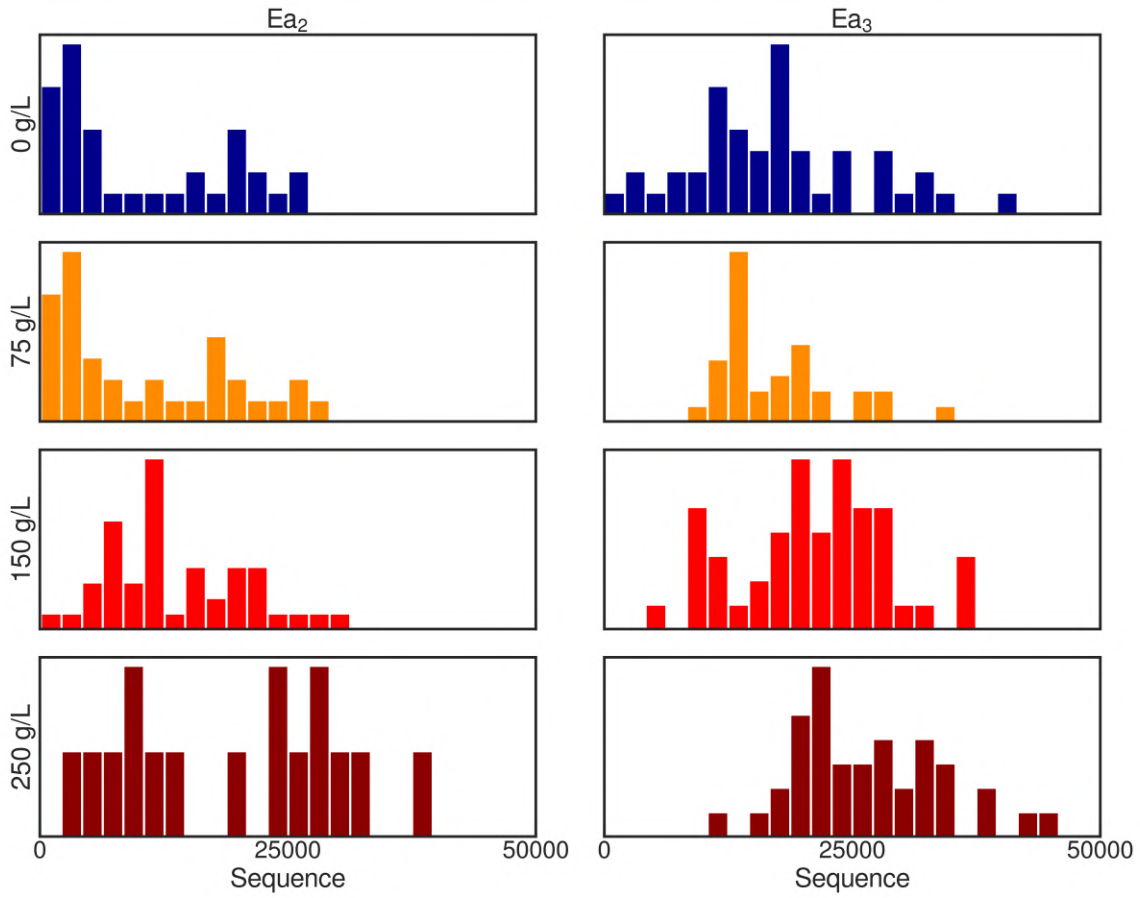


Figure A.22: Distribution of the fitted activated energies from the different Arrhenius Model-Free analyses. This parameter is ill-defined due to the lack of experimental data.

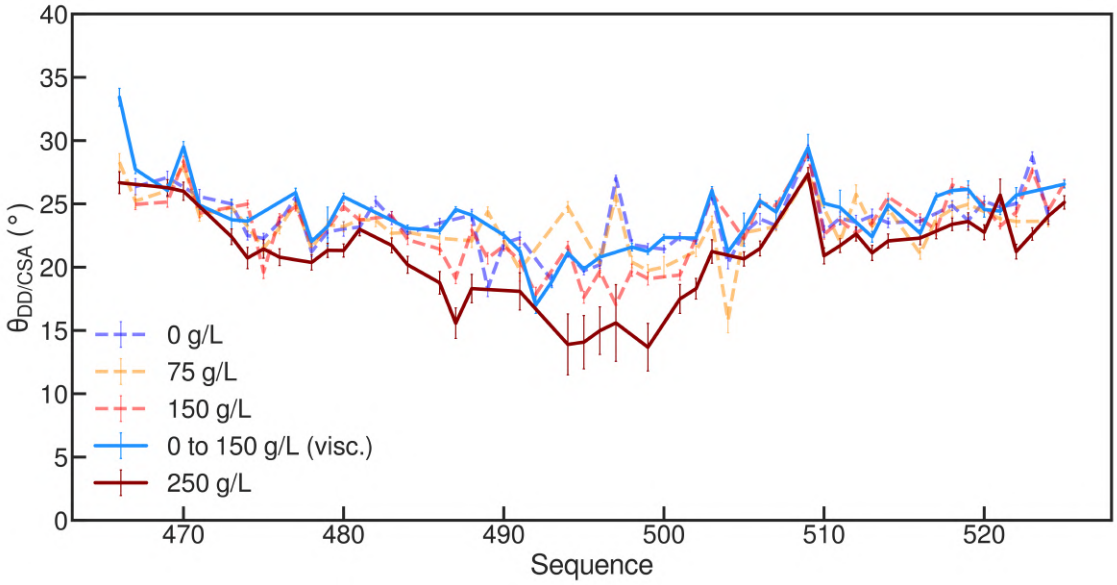


Figure A.23: Fitted DD/CSA angles from the different Model-Free analyses.

A.4 Chapter 6

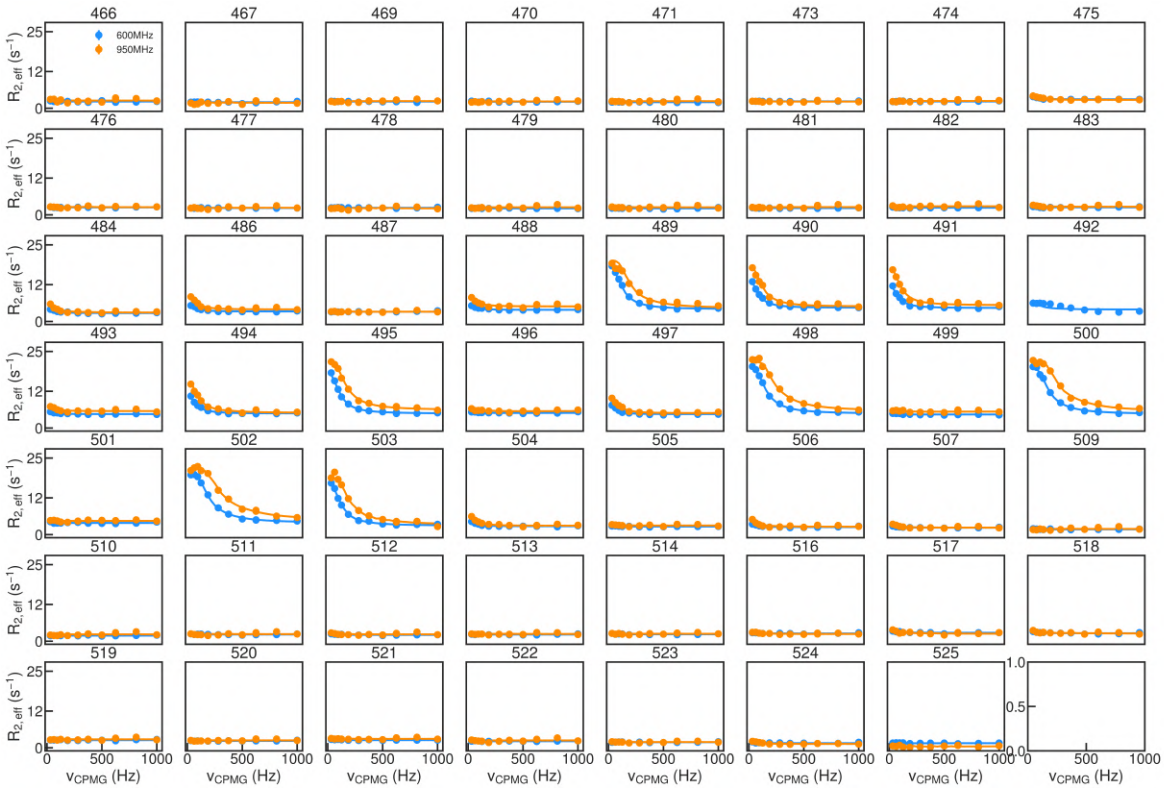


Figure A.24: Experimental (dots) and fitted (lines) CPMG profiles of Ntail 465-525 with 5% of XD and 0 g/L of PEG at 600 (blue) and 950 (orange) MHz proton frequency.

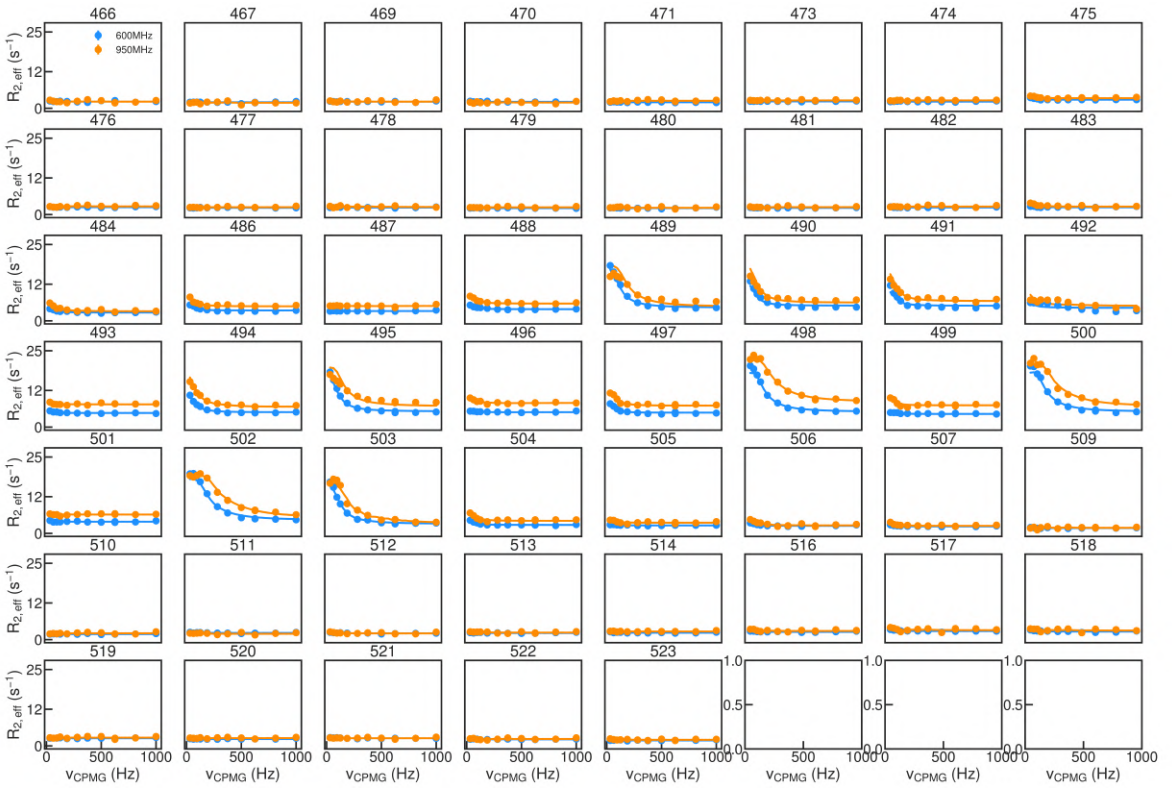


Figure A.25: Experimental (dots) and fitted (lines) CPMG profiles of Ntail 465-525 with 5% of XD and 75 g/L of PEG at 600 (blue) and 950 (orange) MHz proton frequency.

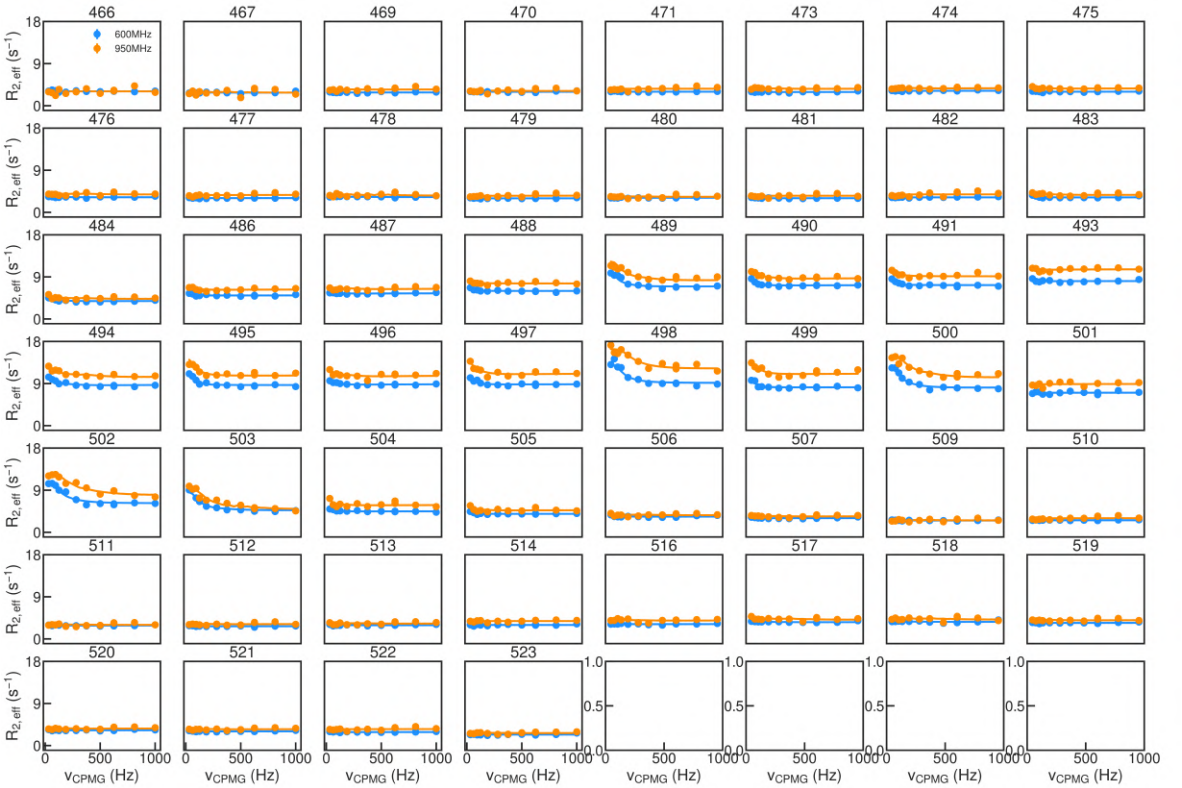


Figure A.26: Experimental (dots) and fitted (lines) CPMG profiles of Ntail 465-525 with 5% of XD and 150 g/L of PEG at 600 (blue) and 950 (orange) MHz proton frequency.

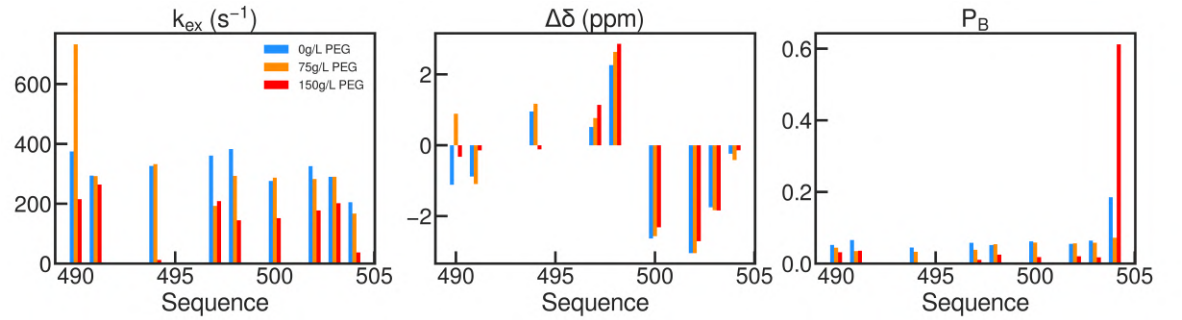


Figure A.27: Residue specific fitting of the helical propensity domain of MeV Ntail 465-525 with 5% XD and 0 (blue), 75 (orange) or 150 (red) g/L of PEG. Left: Exchange rates. Center: Chemical shift differences between free and bound state. Right: Bound state populations.

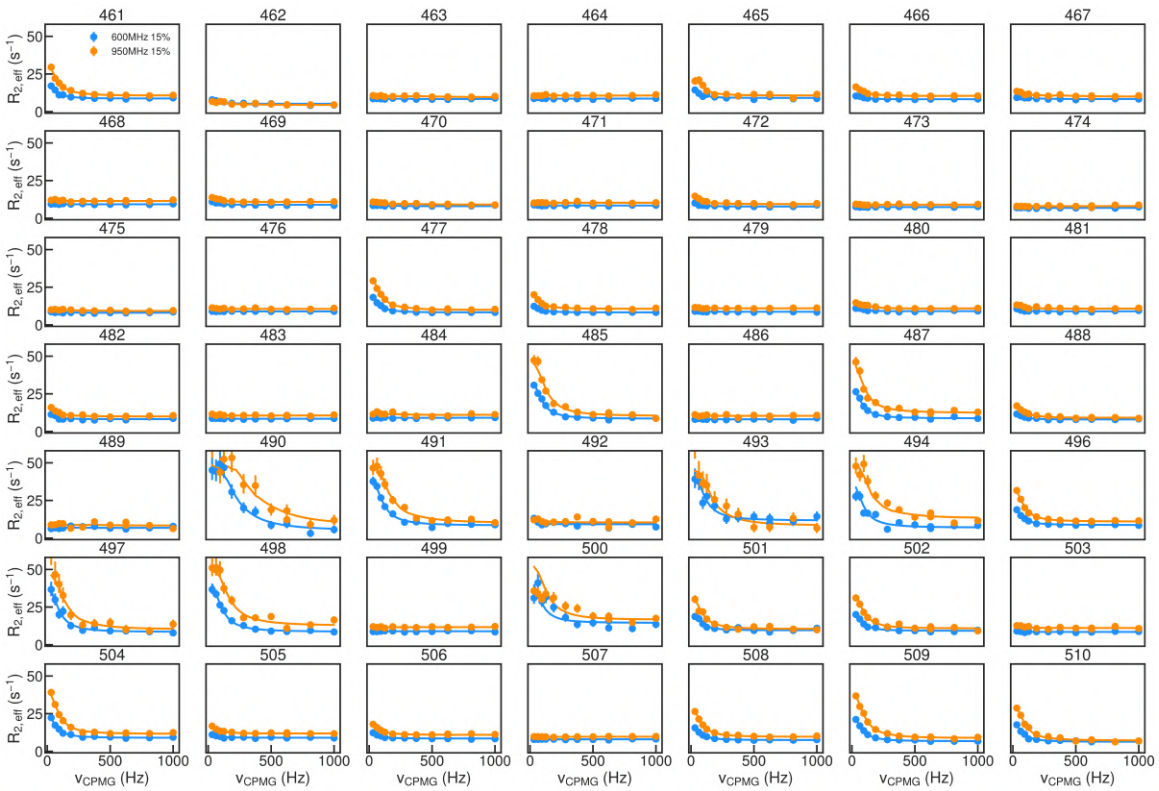


Figure A.28: Experimental (dots) and fitted (lines) CPMG profiles of XD with 15% of Ntail 465-525 and 0 g/L of PEG at 600 (blue) and 950 (orange) MHz proton frequency.

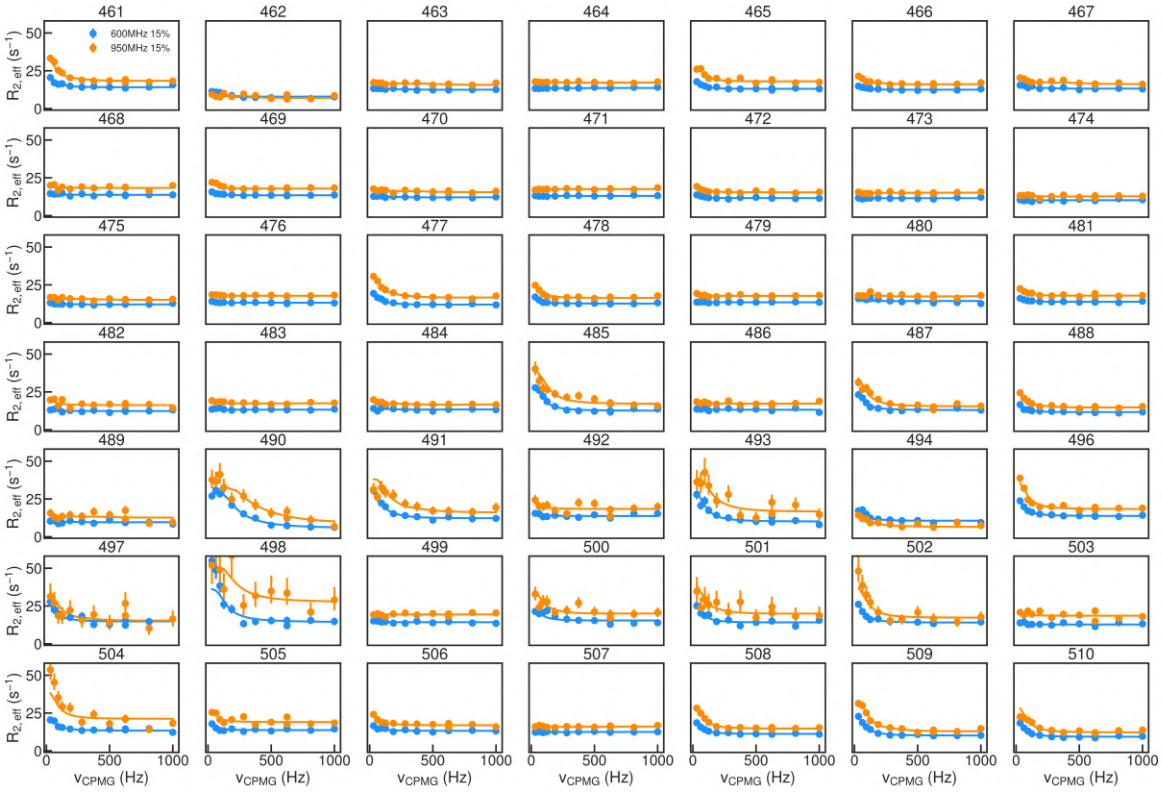


Figure A.29: Experimental (dots) and fitted (lines) CPMG profiles of XD with 15% of Ntail 465-525 and 75 g/L of PEG at 600 (blue) and 950 (orange) MHz proton frequency.

A.5 Chapter 7

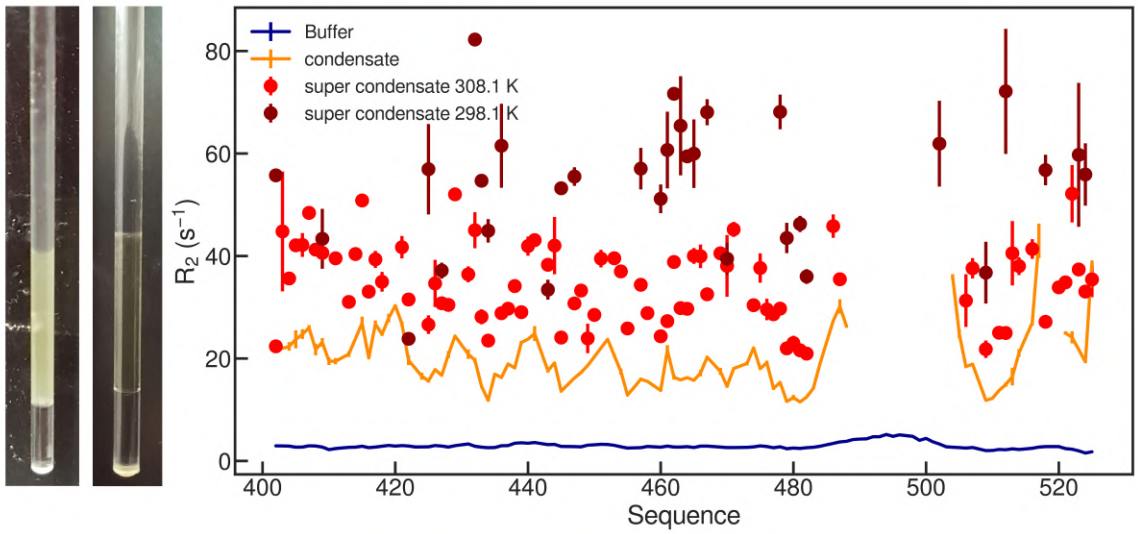


Figure A.30: Left: NMR samples of a "super" biomolecular condensate and a biomolecular condensate respectively. Right: ^{15}N R_2 relaxation rates in MeV Ntail condensate with an initial MeV Ntail concentration of 2 mM at 298.1 K (darkred) and 308.1 K (red), with an initial MeV Ntail concentration of 0.5 mM at 298.1 K (orange), compared with MeV Ntail in the dilute state (blue).

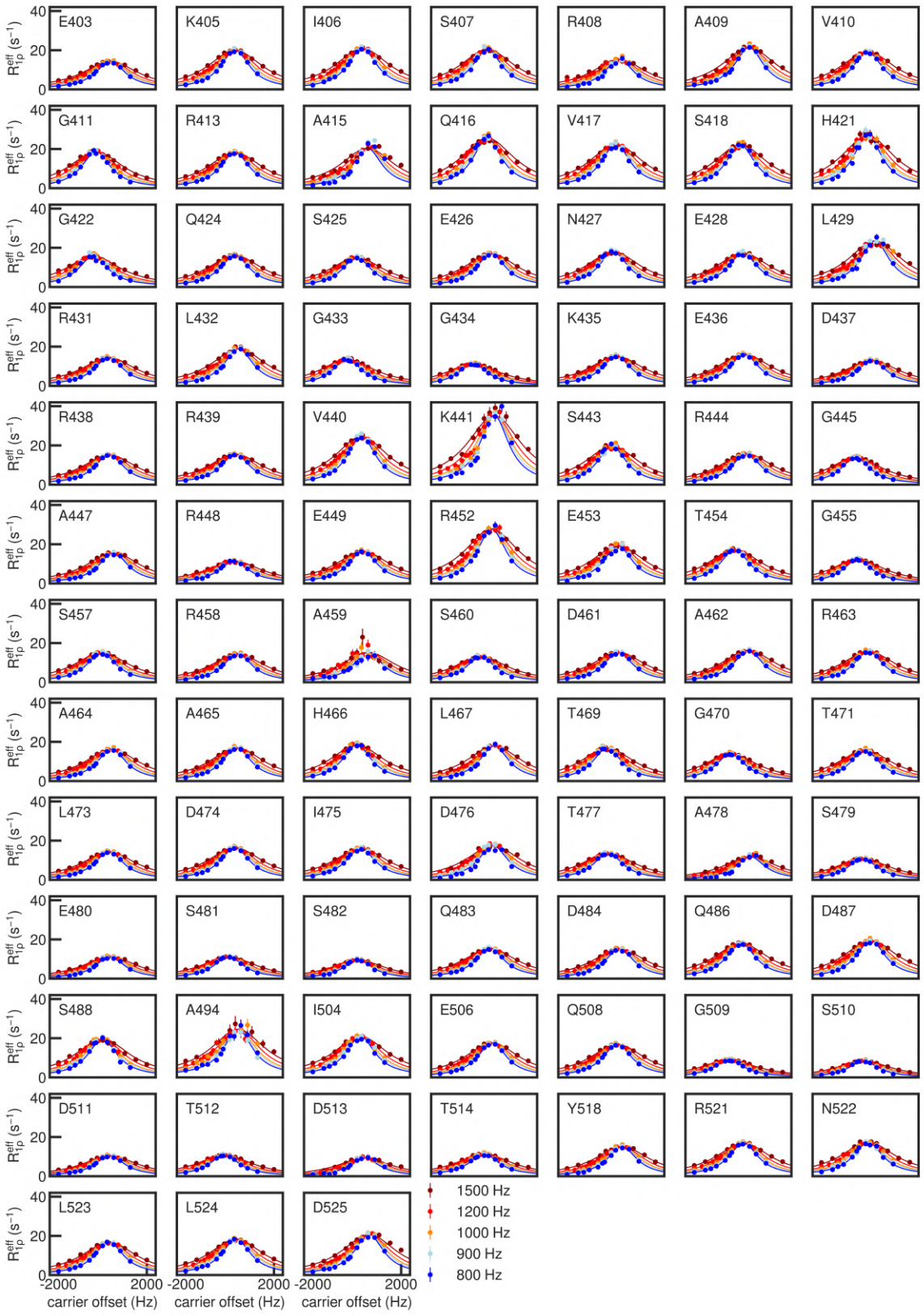


Figure A.31: ^{15}N off-resonance $R_{1\rho}$ relaxation dispersion in MeV Ntail condensate after a one month long room temperature incubation.

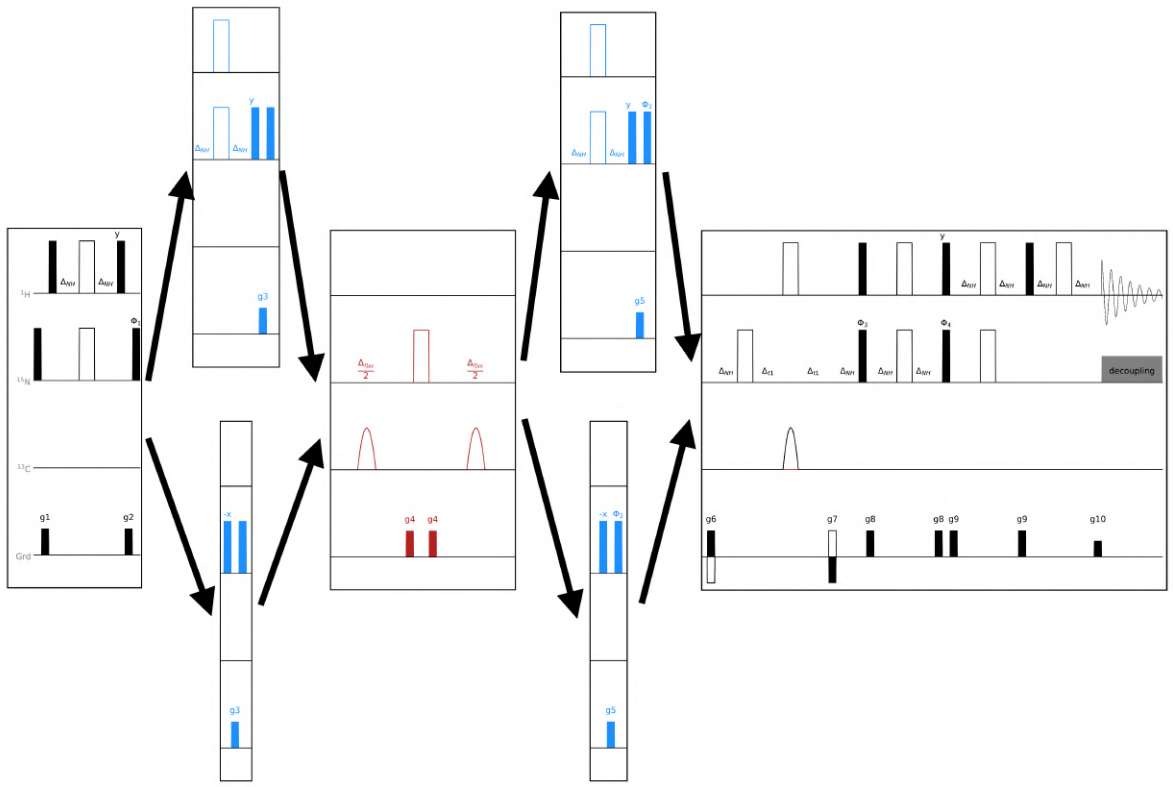


Figure A.32: Pulse sequence for a sensitivity enhanced symmetrical reconversion scheme for measurement of transverse DD/CSA cross correlated relaxation rates in backbone amides. The first block is the excitation and transfer to the nitrogen block. The following blue blocks correspond to conversion or selection of the magnetization. The red block corresponds to the relaxation period, followed by another conversion or selection block to finish with a transfer back to proton and acquisition with nitrogen decoupling. The vertical bars in the fourth line of each block are gradients. Gradients 6,7 and 10 are the first order coherence selection gradients set to 80, -80 and 16.2% respectively while the rest are purging and artifact correction gradients. The other filled and unfilled vertical bars are 90 and 180 degree pulses respectively and the shaped pulses in the carbon channel are adiabatic decoupling pulses. Δ_{NH} is the half NH J-coupling evolution delay, $\Delta_{\eta_{XY}}$ is the relaxation delay and Δ_{t1} is the incremented indirect dimension evolution delay. Φ_1 is [+x, +x, +x, +x, -x, -x, -x, -x], Φ_2 is [+y, -y], Φ_3 is [+x, +x, -x, -x] and Φ_4 is [-y, -y, +y, +y]. Finally, the receiver phase is [+x, -x, -x, +x, -x, +x, +x, -x].

Appendix B

ModFree, a flexible command-line relaxation analysis framework

---

When dealing with an increasingly high amount of relaxation data, it becomes important to progressively analyze the data and assess its quality. To do so, a flexible tool that allows extensive data modeling with different sets of parameters, models and data without a time consuming modification of a complicated algorithm written in a low-level programming language such as C or Fortran becomes significantly useful. Several model-free analysis frameworks already exist such as relax, which offers a broad range of analysis tools for relaxation data. However, the Arrhenius and viscosity extensions introduced recently for IDPs remain unavailable in a simple accessible and user-friendly program to the scientific community. Here, I introduce ModFree, a command-line framework for model-free analysis and reduced spectral density mapping. ModFree is built upon state of the art python libraries and provides a flexible analysis framework allowing least-square minimization of relaxation data on Lorentzian spectral density functions and reduced spectral density mapping using the method of Farrow *et al.*. The error bars are estimated using Monte Carlo analysis. Such framework is a valuable addition to the already existing frameworks since it introduces the possibility to fit the DD/CSA angle, timescale activation energies using multi-temperature relaxation data and friction coefficient in multi-viscosity data. In addition, it is user-friendly, easy to install and written in a flexible programming language.

B.1 Overview of the program

ModFree is a command-line module for analysis of backbone amide nitrogen relaxation data. It can take longitudinal relaxation rates R_1, transverse relaxation rates R_2, nuclear Overhauser enhancements nOe which depends on the dipolar cross-relaxation rate σ_{NH} and R_1, and finally, it can take the longitudinal and transverse cross-correlated relaxation rates η_Z and η_{XY} respectively. ModFree contain several tools to fit relaxation data with a model-free analysis, to plot the resulting fits, to perform reduced spectral density mapping and finally to generate random relaxation data. These different tools are now going to be presented.

B.2 modfree fit

The "fit" function performs a multi-Lorentzian model-free analysis of relaxation data using a Levenberg-Marquardt non-linear Least-square minimization algorithm from Python's LMfit module that minimizes the following function:

$$\chi^2 = \sum_i \left(\frac{R_i^{exp} - R_i^{calc}}{\sigma_i^{exp}} \right)^2$$

Where R_i^{exp} and σ_i^{exp} are experimental relaxation data and the corresponding estimated error and R_i^{calc} is the calculated rate from the model. The residues are summed over all the experimental data introduced in a given model fitting. The expressions used for the different relaxation rates are the following:

$$R_2 = \frac{1}{20}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2 \left(\sigma_\parallel - \sigma_\perp\right)^2 (4J(0) + 3J(\omega_N))$$

$$R_1 = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2 \left(\sigma_\parallel - \sigma_\perp\right)^2 J(\omega_N)$$

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N))$$

$$\eta_{xy} = \frac{1}{15} P_2(\cos\theta)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)\left(\sigma_\parallel - \sigma_\perp\right)\omega_N (4J(0) + 3J(\omega_N))$$

Where μ_0 is the void's permeability, $\hbar$ is the reduced Planck constant, P_2 is the second order Legendre function, γ_H and γ_N are the gyromagnetic ratios of proton and nitrogen respectively, r_{NH} is the average N-H bond distance, $\sigma_\parallel - \sigma_\perp$ is the nitrogen's chemical shift anisotropy and θ is the DD/CSA angle.r_{NH}, $\sigma_\parallel - \sigma_\perp$ and θ's values can be fixed to a chosen value in the input files of the program. and θ can be fitted simultaneously with the different model-free parameters is desired. The general expression for the fitted spectral density function is as follows:

$$J(\omega) = \sum_k \frac{A_k \tau_k}{1 + \tau_k^2 \omega^2}$$

Where A_k and τ_k are the fitted amplitudes and correlation times of a given dynamic mode k. Multiple temperature relaxation data can be fitted with the following Arrhenius relationship:

$$\tau_k = \tau_{k,\infty} exp\left(\frac{E_{a,k}}{RT}\right)$$

Where $E_{a,k}$ is an optimized activation energy parameter, R is the ideal gas constant and T is temperature in Kelvin. Similarly, relaxation data taken with different viscosities can be fitted simultaneously provided that the relative solvent frictions are known with the expression:

$$\tau_k = \tau_{k,0}(1 + \varepsilon_k \rho)$$

Where ε_k is an optimized solvent friction parameter and ρ is the solvent friction of the solution. Finally, data in complex environments can be fitted with the unified expression:

$$\tau_k = \tau_k^{ref}(1 + \varepsilon_k \rho) exp\left(\frac{E_{a,k}}{RT}\right)$$

One must note that an Arrhenius relationship optimizes different amplitudes parameters for each temperatures while the viscosity model assumes a negligible amplitude modulation as a function of solvent friction at a given temperature. Among all the parameters that are optimized, the user can choose to fix some of them to a given value as well as the number of dynamic modes to fit, allowing a useful amount of flexibility.

The "fit" tool takes several input arguments: A parameter file in which is indicated the number of dynamic modes, whether or not to fit θ or to fix a given parameter, the number of Monte-Carlo iterations and the values for r_{NH}, $\sigma_\parallel - \sigma_\perp$ and, if fixed, θ. The second output is a file containing the path of every single relaxation data set including the corresponding magnetic field and, if needed, the corresponding temperatures and solvent frictions. Finally, a last argument allows to choose the model between Standard, Arrhenius, Viscosity and Arrhenius-Viscosity. The module outputs all the optimized parameters in individual files containing three columns: The residues, the values and the errors. It also outputs the used experimental and calculated relaxation rates. In addition, the program outputs the reduced χ^2, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for statistical analyses. Finally, for each residues, the program outputs the Monte-Carlo simulation results including the statistics parameters to allow an analysis of the χ^2 distribution and other correlation analyses.

B.3 modfree plot

The "plot" function plots the results of a model-free analysis performed with this framework. It takes as argument the output directory containing the model-free results and plots the optimized parameters, the statistics parameters, the overlapped experimental and calculated relaxation rates and the correlation plots between experimental and calculated relaxation rates.

B.4 modfree sdm

The "sdm" function performs a reduced spectral density procedure described by Farrow *et al.* as the so-called "Method 3" in their paper. Presenter in chapter 2, the method takes several nOe values to estimate the derivative of the spectral density function at high frequency values and estimates $J(0.87\omega_H)$ using nOe values, then $J(\omega_N)$ using R_1 values and the estimated $J(0.87\omega_H)$ and finally $J(0)$ using R_2 and the previous estimations. The error is estimated using Monte-Carlo simulations again. Similarly to the "fit" function, the input parameters are the parameter file

indicating the number of Monte-Carlo iterations to perform and the directory file indicating the path to the different relaxation data and their corresponding magnetic fields.

B.5 modfree generate

The ”generate” function generates random relaxation data using a multi-Lorenztian spectral density function, along with the directory file and parameter file templates to perform a fit using the ”fit function”. This functions is useful to generate templates for the input files. As inputs, one can specify the number of dynamic modes for the generation, the magnetic fields, and possibly different temperatures and solvent frictions. The generation parameters are randomly set to correspond to a realistic situation but so far, the resulting relaxation rates are not guaranteed to be realistic.

Appendix C

Acronyms

ABSURD	Average Block Selection Using Relaxation Data
ASTEROIDS	Selection Tool for Ensemble Representations Of Intrinsically Disordered States
CEST	Chemical Exchange Saturation Transfer experiment
CPMG	Carr-Purcell-Meighboom-Gill relaxation dispersion experiment
CSA	Chemical Shift Anisotropy
CSI	Chemical Shift Index
DANTE	
DD	Dipole-Dipole interaction
FTIR	Fourier Transform Infrared Resonance
FRET	Förster Resonance Energy Transfer
HSQC	Heteronuclear Single Quantum Correlation experiment
IDP	Intrinsically Disordered Protein
IMPACT	Interpretation of Motions by a Projection onto an Array of Correlation Times
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
LLPS	Liquid-Liquid Phase Separation
MD	Molecular Dynamics
MeV	Measles Virus
MeV N	Measles Virus Nucleoprotein
MeV P	Measles Virus Phosphoprotein
MF	Model-Free
NMR	Nuclear Magnetic Resonance
nOe	nuclear Overhauser enhancement
NOESY	Nuclear Overhauser Effect SpectroscopY
NPT	isobaric-isothermic ensemble
NVE	Microcanonical ensemble
NVT	Canonical ensemble
PEG	PolyEthylene Glycol
POMT	Potential of Mean Torque
R_1	Longitudinal spin relaxation rate
R_2	Transverse spin relaxation rate
RDC	Residual Dipolar Coupling
SAXS	Small Angle X-ray Scattering
SeV	Sendai Virus
sNT	MeV Ntail 465-525 (short Ntail)
SSP	Secondary Structure Propensity
TROSY	Transverse Relaxation-Optimized SpectroscopY
XD	XD domain of the MeV Phosphoprotein
η_{xy}	Transverse DD/CSA cross-correlated spin relaxation rate
η_z	Longitudinal DD/CSA cross-correlated spin relaxation rate

Appendix D

Publication

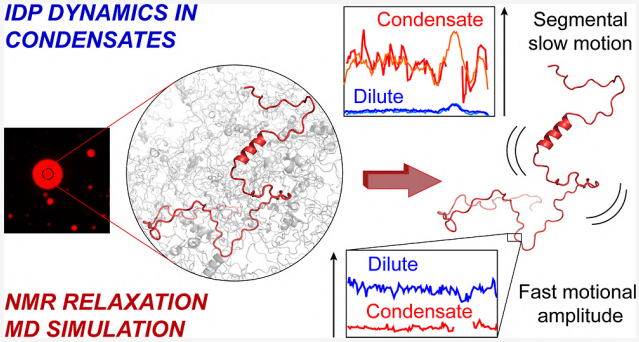
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Liquid–Liquid Phase Separation Modifies the Dynamic Properties of Intrinsically Disordered Proteins

Serafima Guseva,[‖] Vincent Schnapka,[‖] Wiktor Adamski, Damien Maurin, Rob W. H. Ruigrok, Nicola Salvi, and Martin Blackledge*

Cite This: J. Am. Chem. Soc. 2023, 145, 10548–10563

ABSTRACT: Liquid–liquid phase separation of flexible biomolecules has been identified as a ubiquitous phenomenon underlying the formation of membraneless organelles that harbor a multitude of essential cellular processes. We use nuclear magnetic resonance (NMR) spectroscopy to compare the dynamic properties of an intrinsically disordered protein (measles virus N_{TAIL}) in the dilute and dense phases at atomic resolution. By measuring ^{15}N NMR relaxation at different magnetic field strengths, we are able to characterize the dynamics of the protein in dilute and crowded conditions and to compare the amplitude and timescale of the different motional modes to those present in the membraneless organelle. Although the local backbone conformational sampling appears to be largely retained, dynamics occurring on all detectable timescales, including librational, backbone dihedral angle dynamics and segmental, chainlike motions, are considerably slowed down. Their relative amplitudes are also drastically modified, with slower, chain-like motions dominating the dynamic profile. In order to provide additional mechanistic insight, we performed extensive molecular dynamics simulations of the protein under self-crowding conditions at concentrations comparable to those found in the dense liquid phase. Simulation broadly reproduces the impact of formation of the condensed phase on both the free energy landscape and the kinetic interconversion between states. In particular, the experimentally observed reduction in the amplitude of the fastest component of backbone dynamics correlates with higher levels of intermolecular contacts or entanglement observed in simulations, reducing the conformational space available to this mode under strongly self-crowding conditions.



INTRODUCTION

Intracellular compartmentalization is essential for maintaining specific environmental conditions for many cellular processes. Along with the well-described membrane-bounded compartments, over the last 10 years, it has become increasingly clear that membraneless organelles, formed via liquid–liquid phase separation (LLPS), a spontaneous and reversible demixing of component polymers that circumvents the energetic cost of membrane formation, provide ubiquitous microenvironments essential for cellular function. Many biochemical reactions have been shown to occur within such condensates where the concentration of the reagents can be several orders of magnitude higher than the average concentration in the cell, allowing for tight temporal and spatial regulation of a large number of biochemical processes.[1]

Phase separation is a well-understood phenomenon in polymer physics and has been thermodynamically described over 80 years ago by Flory[2] and Huggins,[3] whose mean-field descriptions are still used to describe phenomenological observations of phase separation. The realization of the extent of this phenomenon has transformed our perspective on cellular biochemistry,[4,5] stimulating intense investigation of the molecular and mechanistic details underpinning their physicochemical stability.[6−11] While localization of physiological cofactors in biomolecular condensates has been shown to increase enzymatic efficiency,[12−14] understanding the true functional advantage offered by such phenomena requires a description of their behavior in the dense phase.[6,15] LLPS relies on multivalent interactions between component biomolecules which often show high levels of flexibility, at least in the dilute phase. Examples of such flexibility can be found in single-stranded RNA and intrinsically disordered proteins (IDPs),[16,17] both of which are observed to participate in the formation of physiologically relevant condensates. The biomolecular function of such molecules is determined by

Received: December 22, 2022
Published: May 5, 2023

their intrinsic dynamic behavior, which is highly sensitive to changes in the environment, for example, temperature, ionic strength, and viscosity.[18]

Due to the importance of biocondensation for understanding cellular function, a great deal of effort has been devoted to describing the behavior of IDPs within membraneless organelles. The concentration of proteins in droplets has been estimated on the basis of numerous techniques (e.g., UV absorbance, FCS[19]) and can be up to 3 orders of magnitude higher than the co-existing dilute phase. The importance of intermolecular interaction for the stabilization of the dense phase has been investigated in great detail, for example, concerning aromatic and charged amino acids, in particular arginines, which have been shown to be essential to trigger and maintain condensation.[9,20−24] These interactions do not appear to affect secondary structure propensities along the sequence of the IDP. Indeed, backbone chemical shifts indicate that many IDPs including ddx4,[20] hnRNPA2,[25] FUS,[26] and CAPRIN1[27] appear to maintain random-coil-like features already observed in the dilute phase throughout the protein. Long-range structure in IDPs has also been studied in the dense phase using fluorescence,[28] paramagnetic nuclear, electron magnetic resonance,[29] and intermolecular nuclear Overhauser effects either under free or phase-separated conditions.[20,30−32] In general, multivalent inter-protein interactions that are thought to stabilize membraneless components are considered to be weak, allowing for the rapid association and dissociation rates necessary to maintain the liquid-like nature of the droplet, despite the very high concentration of protein.[9,10]

In addition to understanding the impact on the free energy landscape, it is equally important to investigate the impact of the dense phase on the dynamics of the phase separating proteins.[15] The rates of interconversion between functional states are expected to play an important role in the efficiency of molecular reactions occurring within the droplets, as well as the underlying stability of the phase. The majority of experimental studies of liquid droplets probe the translational diffusion of component proteins within the droplets using fluorescence methods, most commonly FRAP (fluorescence recovery after photobleaching)[33] or diffusion ordered spectroscopy nuclear magnetic resonance (NMR).[20,30] These methods reveal a significant slowing down of translational diffusion, of between two and three orders of magnitude compared to the dilute phase, while nevertheless demonstrating the liquid-like nature of the coacervate. Conformational dynamics involving large-scale reorganization can also be investigated using distance-dependent approaches, such as paramagnetic relaxation enhancement,[34] electron paramagnetic resonance,[35] fluorescence correlation spectroscopy, and single molecule FRET.[36]

Reorientational dynamics of IDPs under phase-separating conditions have been investigated using time-resolved fluorescence spectroscopy techniques, for example, in α-synuclein, where decreased chain flexibility was observed in LLPS compared to dilute conditions,[37] and tau protein, where fluorescence and EPR spectroscopies[28,38] were used to characterize modulation of rotational correlation times upon phase separation. EPR was also used to study IDP dynamics and in arginine-rich peptides.[39] Along with experimental methods, computational approaches have been used to investigate protein condensates.[11,40−44] While coarse-grained methods[45−47] have been successfully exploited to reproduce condensate properties, atomistic simulation[48] can also provide detailed biophysical insight into the equilibrated condensed phase at atomic resolution.

NMR spectroscopy is perhaps the ideal tool for studying the dynamic properties of IDPs on timescales from tens of picoseconds to tens of nanoseconds that are precisely encoded in experimentally accessible spin relaxation rates. Atomic resolution backbone dynamics of ddx4,[20,49] hnRNPA2,[25] elastin,[50] FUS,[26] and CAPRIN1[32,51] have been measured using ^{15}N relaxation, resulting in diverse observations, possibly testifying to the different experimental conditions (temperature, magnetic field strength) employed in the different studies. These ground-breaking studies highlight the remarkable potential of NMR to investigate the dynamics of IDPs in phase separating conditions. An analytical comparison of molecular motions in the condense and dilute phases, which is necessary to provide a complete understanding of the stabilizing forces within biocondensates, is currently lacking.

While the physical origin of individual relaxation rates can be ambiguous, due to the complex dependence on motional frequencies occurring on different timescales, a combination of relaxation rates can provide unique insight into the dynamic behavior of each site in the disordered chain.[52−60] Combining extensive sets of relaxation rates measured at multiple magnetic fields,[58] it is possible to develop a self-consistent physical description of the reorientation properties throughout the IDP, identifying three dynamic modes that contribute to NMR spin relaxation in IDPs.[59] These three distinct components appear to report on librational motions for the fastest component (no temperature or viscosity dependence of the motional timescale), backbone conformational sampling for the intermediate timescale motion (as revealed by viscosity coefficients and activation energies that correlate with expected values from other spectroscopies), and chain-like or segmental motions for the slowest component (as revealed by viscosity coefficients, activation energies, and length-dependence of timescales and segmental persistence lengths).[18,59,61] The response of these distinct motional modes to changes in environmental parameters such as temperature and crowding allows considerable insight into the sequence-dependent behavior of IDPs in complex environments and their coupling to the surrounding solvent.[61] To gain further insight into the physical nature of these dynamic components, approaches based on multi-conformational molecular dynamics (MD) simulation have been developed, providing ensembles of trajectories in agreement with experimental relaxation data that describe the time-dependence of motions in IDPs within their different conformational sub-states.[62,63] This thermodynamic and kinetic description offers new insight into the backbone dynamics of IDPs in solution, for example, highlighting close coupling with solvent behavior.[64]

Here, we apply these approaches to the study of the dynamic behavior of an intrinsically disordered domain of the nucleoprotein of measles virus (N_{TAIL}), which is not only essential for formation of physiological droplets upon mixing with the phosphoprotein in infected cells[12] but also phase separates alone under certain conditions, allowing for the measurement of extensive NMR relaxation data that can be compared with dilute and semi-crowded conditions. We show that while conformational sampling of the backbone is not strongly affected by formation of the dense phase, librational, backbone torsional and segmental dynamics are considerably slowed down, and their contribution to the detected dynamics

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

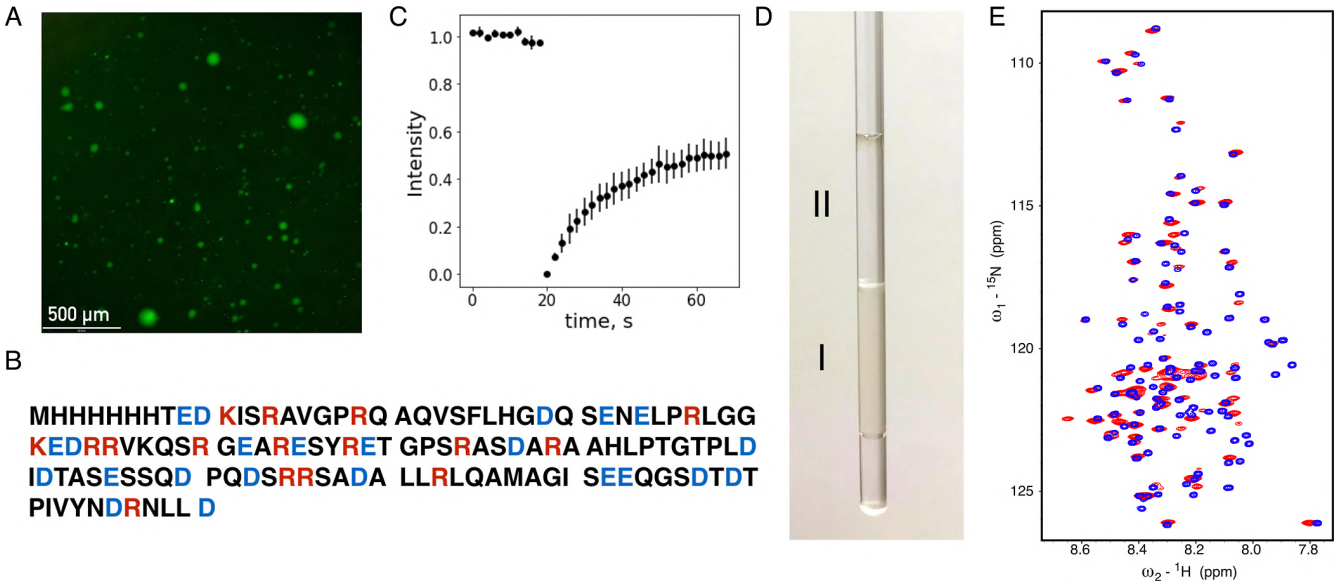


Figure 1. Measles virus N_{TAIL} phase separation upon mixing with PEG. (A) Fluorescence microscopy imaging of a colloid system where N_{TAIL} is labeled with FAM. (B) N_{TAIL} amino acid sequence, including non-cleaved purification tag. (C) FRAP of N_{TAIL} demonstrates the liquid nature of the condensates. (D) Typical NMR sample of N_{TAIL} showing the dense phase (I) separated from the dilute phase (II). Plunger is not yet inserted. (E) $^{15}N-^{1}H$ HSQC spectra of N_{TAIL} in the dilute (blue) and dense (red) phases. The dilute phase spectrum was recorded at 850 MHz, 298 K, and the dense phase at 950 MHz, 298 K.

is significantly redistributed. Extensive MD simulations under self-crowding conditions reproduce the impact of droplet formation on both the free energy landscape and the kinetics of interconversion, correlating with the increase in intermolecular contacts encountered under self-crowding conditions. This combination of experimental NMR and molecular simulation provides unique new insight into the modulation of the dynamic behavior of IDPs in dilute and dense phases.

RESULTS

Identification of a Model System for Studying IDP Dynamics in LLPS. As a model to study protein conformational and dynamic behavior in the dense phase, we identified the disordered domain of measles virus nucleoprotein from 401 to 525 (N_{TAIL}) which phase separates upon mixing with PEG10000 solution and forms a biphasic colloid solution. This phase evolves to a two macro-phase solution: one rich in N_{TAIL} that is yellow and dense and the other with a low N_{TAIL} concentration. The colloid solution was imaged using fluorescently labeled N_{TAIL}, revealing that N_{TAIL} is concentrated inside droplets, while the surrounding solution shows negligible fluorescence (Figure 1A), indicating low N_{TAIL} content. 10% of the amino acids in N_{TAIL} are arginine (Figure 1B), although there are only three aromatic residues. The liquid nature of N_{TAIL} condensates is shown by FRAP, demonstrating exchange between the two phases (Figure 1C).

As phase separation is only observed in the presence of PEG, we first studied its distribution using NMR spectroscopy. As shown from ^{1}H NMR, the presence of PEG in the dense phase was minor, indicating that it is excluded from the dense phase (Figure S1). These results suggest that only N_{TAIL} is required for phase separation, possibly forming a self-scaffold, while PEG acts as a trigger, possibly by replacing water and helping the system to reach a critical water volume fraction.

Phase Diagram of N_{TAIL} in Dilute and Dense States. The phase diagram was established following published approaches.[20] Calculated N_{TAIL} concentrations as a function of salt and temperature are presented in Figure S2. The phase diagram has an upper critical solution temperature; the highest protein concentration was estimated to be 38.9 mM corresponding to 584 mg/mL at 288 K and 118 mM NaCl. Under these conditions, approximately 42% of the volume fraction is solvent (corresponding to 23.3 M). The protein fraction decreases in the dense phase as salt concentration increases, and the volume of the dense phase changes. Using fluorescence microscopy, we observe that increasing NaCl concentration progressively dissolves droplets (Supporting Information Figure S3).

The phase diagram was fitted to the Flory–Huggins model at different NaCl concentrations. The results are plotted together with experimental data in Figure S4. χ shows universally positive values, corresponding to an attractive interaction promoting phase separation. Comparison of entropic and interacting components using this simple comparison (assuming negligible changes in conformation) confirms thermodynamic promotion of demixing under our experimental conditions.

NMR Spectroscopy of the N_{TAIL} in the Dense Phase. In order to examine changes in conformation due to condensation, we have compared the chemical shifts of N_{TAIL} in the dilute and dense phases. Single dense macroscopic phases were produced for NMR measurement comprising 10% $^{15}N-^{13}C$-labeled N_{TAIL} and 90% of unlabeled N_{TAIL} (see Materials and Methods) (Figure 1D). Comparison of ^{15}N HSQC spectra reveals that almost all peaks observed in dilute N_{TAIL} (300 μM) are present in dense phase spectra (Figure 1E) apart from the region between 488 and 499, which corresponds to a helical element that is known to be populated to >80% for a stretch of 6 contiguous amino acids in the dilute phase. The absence of this region may be due to slow tumbling or its involvement in interactions that are important for phase separation. ^{13}C backbone resonances are also very similar in

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

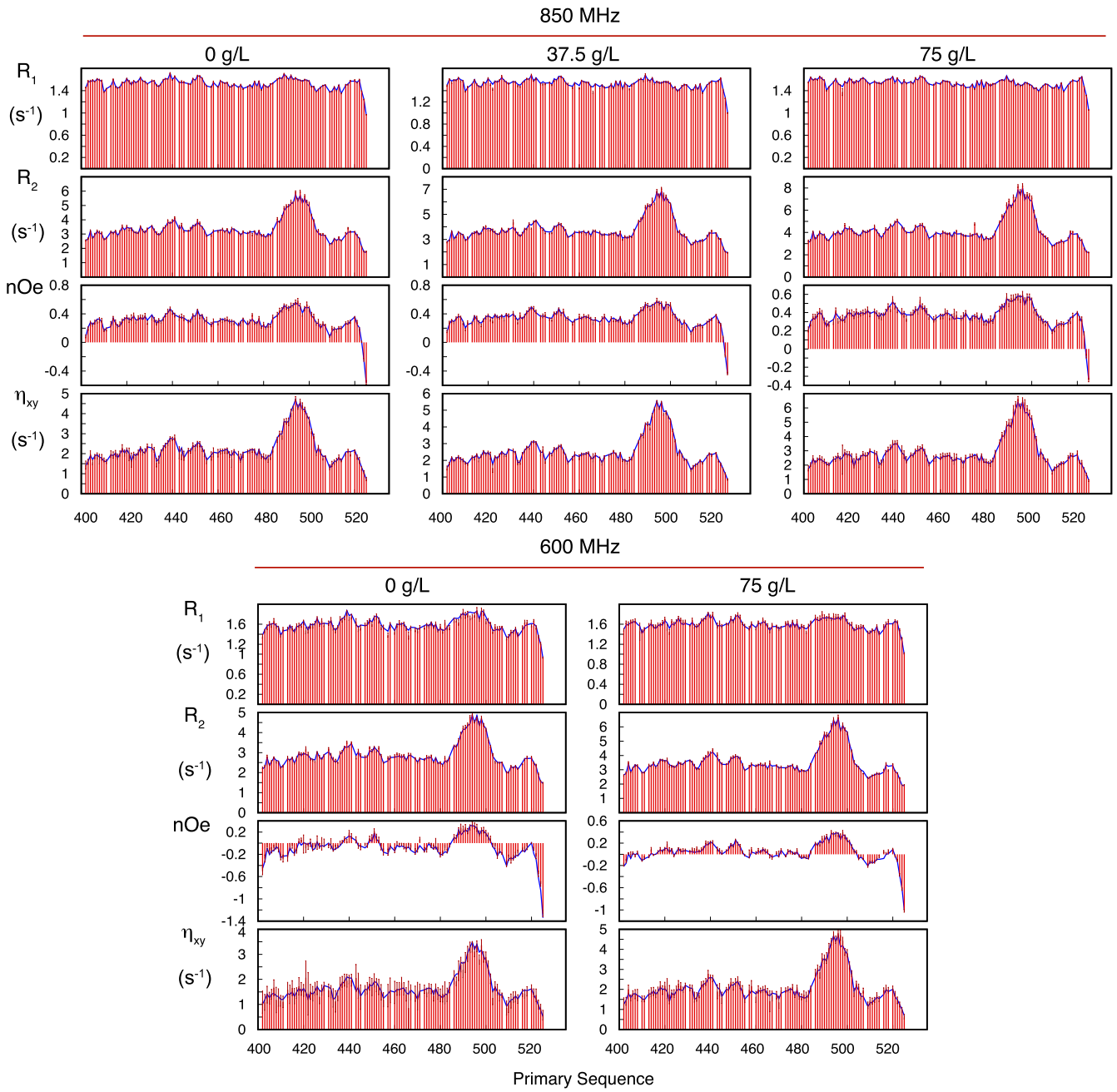


Figure 2. Experimental ^{15}N relaxation rates of N_{TAIL} in the dilute phase, under crowding conditions not-inducing phase separation. Longitudinal (R_1) and transverse (R_2) autocorrelated relaxation, transverse cross-correlated DD/CSA (η_{xy}), and heteronuclear $\{^1H\}-^{15}N$ nuclear Overhauser enhancement (nOe) measured at two magnetic field strengths (600 and 850 MHz) as a function of viscogen concentration (PEG10000). Red bars show the experimental rates, and blue lines show values calculated using a simultaneous model-free fit of all data. Additional R_1 and nOe were measured at 700 MHz in the absence of crowder and used for cross-validation. These values are shown in Figure S7.

the dilute and dense phases (Figures S5 and S6), with only three amino acids in the C-terminal region (^{518}YND^{520}) showing continuous differences, uniquely in the C′ chemical shifts of the dense phase. We do not know the origin of these localized differences, although they may be due to intermolecular interactions in the dense phase. The remainder of the protein exhibits remarkably similar chemical shifts, as has been observed for a number of phase-separated IDPs.[20,25,26]

NMR Relaxation of N_{TAIL} in the Dilute Phase and in Crowded Conditions. In order to describe the conformational dynamics of N_{TAIL}, ^{15}N backbone relaxation rates of the 130 amino acid protein were measured. Transverse (R_2), longitudinal (R_1), cross-relaxation (heteronuclear nOe), and transverse cross-correlated dipole−dipole/chemical shift anisotropy (CSA) cross-relaxation (η_{xy}) were measured at 600, 700, and 850 MHz in the dilute phase and at 600 and 850 MHz at 0, 37.5, and 75 g/L PEG concentrations (Figure 2), at which N_{TAIL} demonstrably does not phase separate. The presence of the helical region is evident, while maxima in the profile of the heteronuclear nOe reveal more ordered regions at ^{439}RVKQ^{442} and ^{450}SYR^{452}. The remainder of the protein exhibits classical relaxation characteristics of an IDP.

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

Model-Free Analysis of N_{TAIL} in the Dilute Phase. The correlation function was modeled as described previously[18,53] using three components represented by distinct exponential time constants whose relative amplitude A_k are governed by $\sum_k A_k = 1$, correlating the timescales of the dynamics of intermediate and slow timescale motions (τ_k) in buffer and under the two crowding conditions with the nano-viscosity of the surrounding solvent as determined from ^{1}H relaxation of water.[61] This analysis, which has recently been applied to the related NT from Sendai virus nucleoprotein and the disordered domain of MAP kinase MKK4, allows us to determine characteristic timescales, and the relative amplitude of their contributions to the angular correlation function, throughout the primary sequence. 20 experimental rates from 600 to 850 MHz were fitted for each site in the protein (reproduction of experimental rates is shown in Figure 2). Relaxation rates at 700 MHz that were not used in the fitting step were compared to values back-calculated from the model. Experimental values are predicted within experimental error, cross-validating the analysis (Figure S7). The distribution of fast (A_1, τ_1), intermediate (A_2, τ_2), and slow (A_3, τ_3) timescale motions (Figures 3 and S8) resembles recent analyses of multi-field relaxation data from other IDPs at 298 K, with fast motions occurring on timescales of approximately 50 ps, intermediate motions on timescales around 1 ns, and the slowest component varying between 5 and 10 ns as a function of viscogen concentration. The viscosity coefficients follow a similar pattern compared to disordered NT and MKK4 domains,[61] with coefficients of intermediate and slow timescale motions separated by approximately a factor of three (Figure S8), a difference that we have interpreted in terms of the dimensions of the reorienting moiety (individual peptide plane and multi-peptide segment, respectively). The amplitudes of the slowest motions are dominated by the helix and to a lesser extent a highly charged segment ^{435}KEDRRVK^{441} (Figure S9). The angle between CSA and dipolar tensors was well defined (26.5 ± 2.2)° over the protein, again reproducing the range estimated from similar analysis of NT and MKK4 disordered domains. These site-specific values were incorporated in the analysis of dense phase relaxation data.

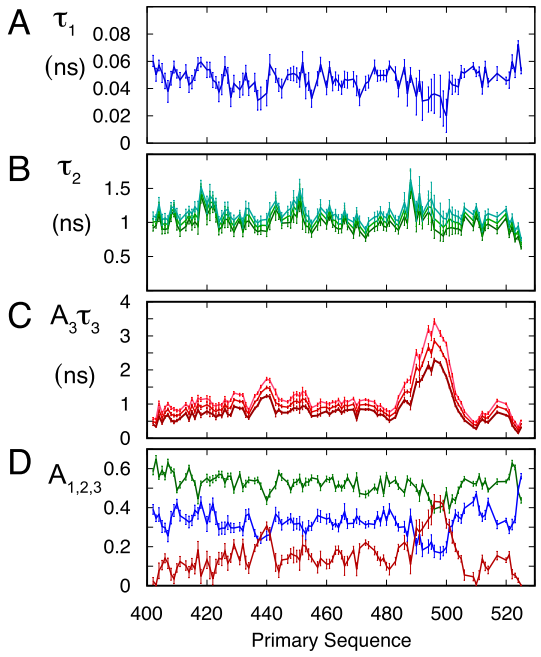


Figure 3. Model-free analysis of dynamic behavior of N_{TAIL} as a function of viscosity. 20 relaxation rates were fitted to eqs 7a and 7b using eq 8 to relate data measured in the presence of different concentrations of viscogen. 8 parameters were optimized for each ^{15}N−^{1}H spin pair (ε_3, ε_2, $\tau_{3,\infty}$, $\tau_{2,\infty}$, τ_1, θ, A_2, and A_3—see Materials and Methods). (A) Residue-specific correlation times of the fast motional mode (τ_1). (B) Residue-specific correlation times of intermediate, backbone motions (τ_2). 0 (dark green), 37.5 (green), and 75 g/L PEG10000 (light green). (C) Residue-specific correlation times of segmental backbone motions (τ_3). 0 (dark red), 37.5 (red), and 75 g/L PEG10000 (light red). (D) Distribution of the residue-specific amplitudes of the different motional modes (A_1—blue, A_2—green, A_3—red).

NMR Relaxation of N_{TAIL} in the Dense Phase. Relaxation data were measured at the same magnetic field strengths in the dense phase (Figure 4). Although the absolute values are very different compared to those measured in the dilute phase samples, the distribution of values along the sequence shows some similar features. For example, increased heteronuclear nOe and transverse relaxation rates appear in the same segments of the protein in both phases (around 439 and 451). Predominantly hydrophobic sites in the terminal regions of the protein show relatively higher transverse relaxation rates in the dense phase compared to that in the dilute phase, for example, ^{406}KI^{407}, ^{416}QV^{417}, and ^{516}IVY^{518}, indicating that, unlike in the dilute phase, fast and intermediate motions are partially damped in the highly flexible termini, thereby allowing slow motions to contribute more significantly to relaxation.

Model-Free Analysis of N_{TAIL} in the Dense Phase. A model-free analysis of the backbone dynamics of N_{TAIL} in the dense phase (Figure 5, data fits in Figure S10) was performed on the basis of these experimental data. A systematic discrepancy between R_2 and η_{xy} was noted when both were included in the target function, with η_{xy} overestimated and R_2 underestimated, suggesting a small contribution from exchange. We therefore removed R_2 from the analysis and compared to the value calculated from the fit of the remaining rates that are not sensitive to such exchange processes. The minor but systematic difference between calculated and experimental R_2 at the two fields supports the existence of a small exchange contribution, with a weak field dependence throughout most of the protein (Figure S10), suggesting a slow or slow-intermediate timescale. Intriguingly, this apparent exchange contribution has maxima in the vicinity of arginine residues. It may be relevant that a similar slow exchange contribution was more quantitatively identified in the dense phase formed by the protein ddx4.[49] Off-resonance rotating frame relaxation was not measured as in the ddx4 study; however, we have measured ^{15}N relaxation dispersion in the dense phase, which would be sensitive to exchange events occurring on timescales in the millisecond to hundreds of millisecond regime, and see no evidence of intermediate exchange (Figure S11). Although the limited number of individual rates included in the model-free analysis of the dense state precludes detailed cross-validation, removal of η_{xy} measured at 600 or 850 MHz from the analysis and fitting to the remaining data sets predicts measured values within 5% accuracy, attesting to the overall coherence of the two data sets (Figure S12).

The model-free analysis reveals significant differences both in the correlation times of the different components and their amplitudes, compared to the dilute phase. The correlation time

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

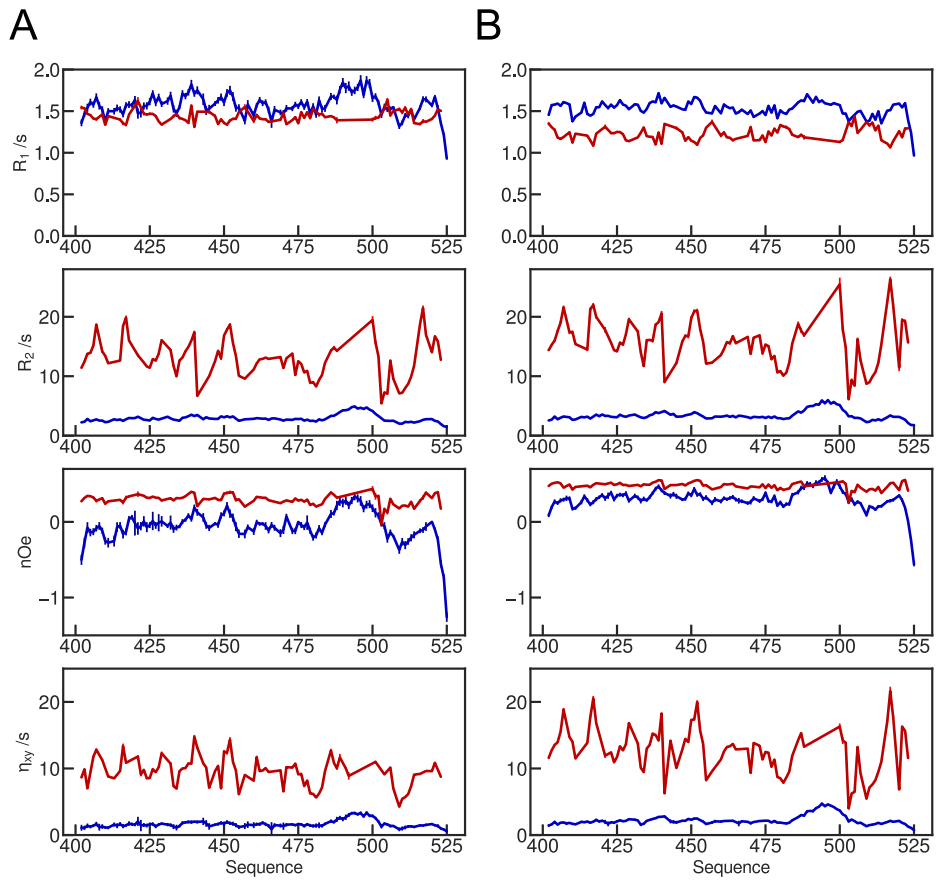


Figure 4. Comparison of experimental relaxation rates in the dilute and dense phase. Longitudinal (R_1) and transverse (R_2) autocorrelated relaxation, transverse cross-correlated DD/CSA (η_{xy}), and heteronuclear $\{^1H\}-^{15}N$ nuclear Overhauser enhancement (nOe) measured at two magnetic field strengths (A-850 and B-600 MHz). Red—in the dense phase, blue—in the dilute phase.

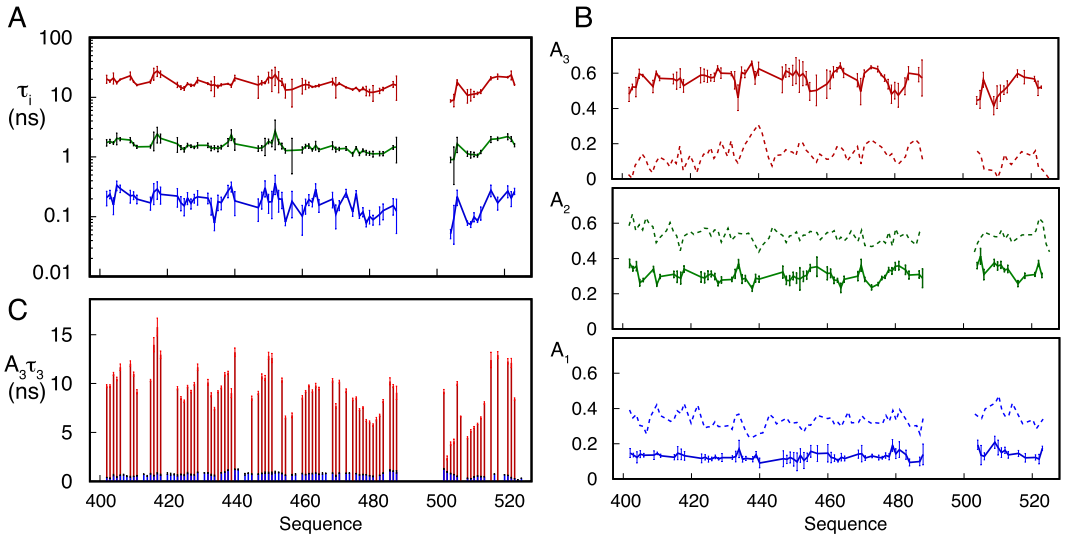


Figure 5. Model-free analysis of dynamic behavior of N_{TAIL} in the dense phase. Relaxation rates were fitted to eq 7a. 5 parameters were optimized for each $^{15}N-^1H$ spin pair (τ_1, τ_2, τ_3, A_2, and A_3). In all cases, error bars represent the uncertainty in the dynamic parameters as estimated from Monte-Carlo noise-based least squares analysis respecting 95% confidence limits. (A) Residue-specific correlation times of the three motional modes. τ_1 (fast)—blue, τ_2 (intermediate)—green, τ_3 (slow)—red. (B) Residue-specific amplitudes of the three motional modes. A_1 (fast)—blue, A_2—(intermediate) green, A_3 (slow)—red. The amplitudes of the three motional modes characterizing the dilute state (Figure 3D) are shown as dashed lines for ease of comparison. (C) Comparison of the product $A_3\tau_3$ in the dilute (blue) and dense (red) phases.

of the fastest component, shown to sample values tightly distributed around 50 ps throughout the protein in the dilute phase, independently of temperature and viscogen concentration, systematically falls in the 100−200 ps range in the dense phase. Intermediate timescale contributions sample values in the range of 1.5 ns, compared to the dilute phase

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

values around 1 ns. The slower contribution is considerably longer, sampling timescales over 20 ns, compared to 5−10 ns in the dilute phase (Figure S8). Perhaps most remarkably, the amplitudes of the different components (A_1, A_2, and A_3) are drastically redistributed, with fast motional amplitudes being considerably more restricted in the dense phase. As a result, the residual contribution to the correlation function, representing all detectable slower motions, is significantly larger. Notably, the product $A_3\tau_3$ is more than an order of magnitude higher for the dense with respect to the dilute phase (Figure 5).

We note that attempts to quantify the nano-viscosity within the dense phases by measuring the R_1 of water were unsuccessful, due to a strong magnetic field dependence of the measured rate, most likely due to additional exchange effects contributing to the measured rate.[65−67]

MD Simulation of Dilute Solutions of N_{TAIL}. In order to further investigate the physical origins of the model-free parameters, we turned to MD simulation.[68] We initially compiled an ensemble of trajectories of free N_{TAIL} in the dilute phase, using the ABSURD genetic optimization algorithm, as previously applied to NT. 30 trajectories of 200 ns were calculated using different starting structures from an ASTEROIDS-derived ensemble, selected to be in agreement with experimental chemical shifts and residual dipolar couplings.[69,70] ABSURD reweighting was performed against R_2 rates measured at 850 MHz proton frequency using 100 ns segments of all calculated trajectories.

Improvement in the agreement with experimental data is observed for each rate measured that was not included in the target function for selection, as has previously been observed for other proteins (Figure S13).[62,64] This ensemble of trajectories is taken to represent the nanosecond dynamic behavior of the ensemble of states that are interchanging on timescales significantly faster than the chemical shift exchange limit (milliseconds) and therefore to provide the best description of the dynamic behavior experienced in the dilute state. Chemical shifts are still reproduced by the selected trajectories in a similar manner to the original ASTEROIDS ensemble (Figure S14).

MD Simulation of Highly Concentrated Solutions of IDPs. In order to simulate the conditions experienced by IDPs in the dense phase, we first needed to create conditions that reproduce the level of crowding likely to be present in this phase. Although important progress has been made in the development of methodologies to describe the behavior of IDPs in liquid droplets using coarse-grained models,[45,47,71,72] which vastly extend accessible timescales, we decided to retain the atomic resolution of so-called all-atom descriptions,[48] to allow for the calculation of NMR relaxation rates, as well as maintain water−protein interactions of force field combinations that we have previously validated against experimental data in the dilute state.[64] In order to retain the conformational sampling that gives rise to the experimentally measured chemical shifts, we used conformers from the ASTEROIDS ensemble and placed them in a box, avoiding steric clashes with neighbors, until the target concentration was attained. In this way, between 125 and 343 copies of N_{TAIL} were placed in boxes representing 5, 14, and 20 mM concentrations (Figure 6) (packing at higher concentrations of protein was challenging, and this range was considered sufficient to detect concentration-dependent traits). Water and charge-neutralizing counter ions were added once the box had been prepared.

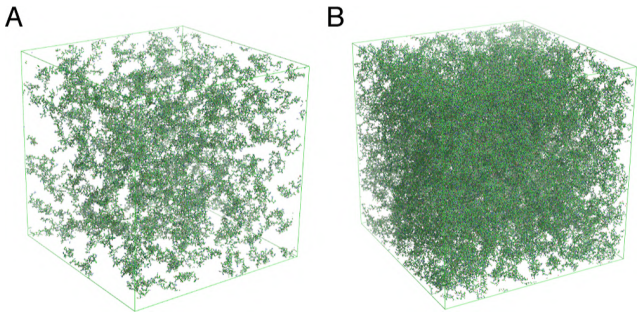


Figure 6. Representation of the starting boxes for self-crowding all-atom MD simulations. (A) 125 molecules of N_{TAIL} randomly selected from the ASTEROIDS ensemble derived from experimental chemical shifts and residual dipolar couplings, representing a concentration of 5 mM. (B) 343 molecules of N_{TAIL} randomly selected from the ASTEROIDS ensemble derived from experimental chemical shifts and residual dipolar couplings, representing a concentration of 20 mM.

Simulations were performed for 200 ns using the same force field as for the dilute phase, except for the 20 mM conditions where 500 ns were used to allow for better sampling of the slowest motions.

^{13}C chemical shifts calculated over all trajectories and all copies of the protein demonstrate that the conformational sampling does not significantly differ from that found in the dilute simulation and experimentally (Figure S14). The average dimensions of the proteins comprising the dense phase ensemble closely resemble those from the ASTEROIDS ensemble (Figure S15), and the average dimensions of the ensemble of proteins remain stable as a function of time in each of the dense phase conditions (Figure S16). In order to compare to experimental conditions, relaxation rates were calculated from the average angular correlation functions calculated for proteins in each box. We note that these rates are statistically highly robust, as they are averaged over an unusually high number of copies of the protein. Examples of autocorrelation functions from the dilute and most concentrated phase simulations are compared in Figure S17. Simulated relaxation rates are compared to each other, and to experimentally determined rates, in Figure 7. The highest concentration simulation exhibits a similar underlying shape with respect to sequence for the predicted $J(0)$-sensitive rates (R_2 and η_{xy}) compared to the experimental rates from the dense phase (neither exhibit the often observed bell-shaped distribution of transverse relaxation often found in the dilute phase[73]), apart from spikes in the experimental data that are present at both magnetic field strengths.

The same simulation is also broadly consistent with the distribution and range of the heteronuclear nOe at both magnetic field strengths. R_1 is in less good agreement, with experimental values, rather coinciding with the simulation of 5 mM N_{TAIL}, suggesting that intermediate timescale motions are less accurately captured by the self-crowded simulations compared to the other two components.

Simulated relaxation rates were used to apply model-free analyses under the three conditions. The slow motional amplitude, A_3, from the most crowded simulation closely reproduces both the sequence-dependence and the absolute experimentally determined values, while A_1 decreases with increasing concentration, mirroring observed difference between model-free parameters corresponding to dilute and dense phase experimental data (Figure 8). We have examined

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

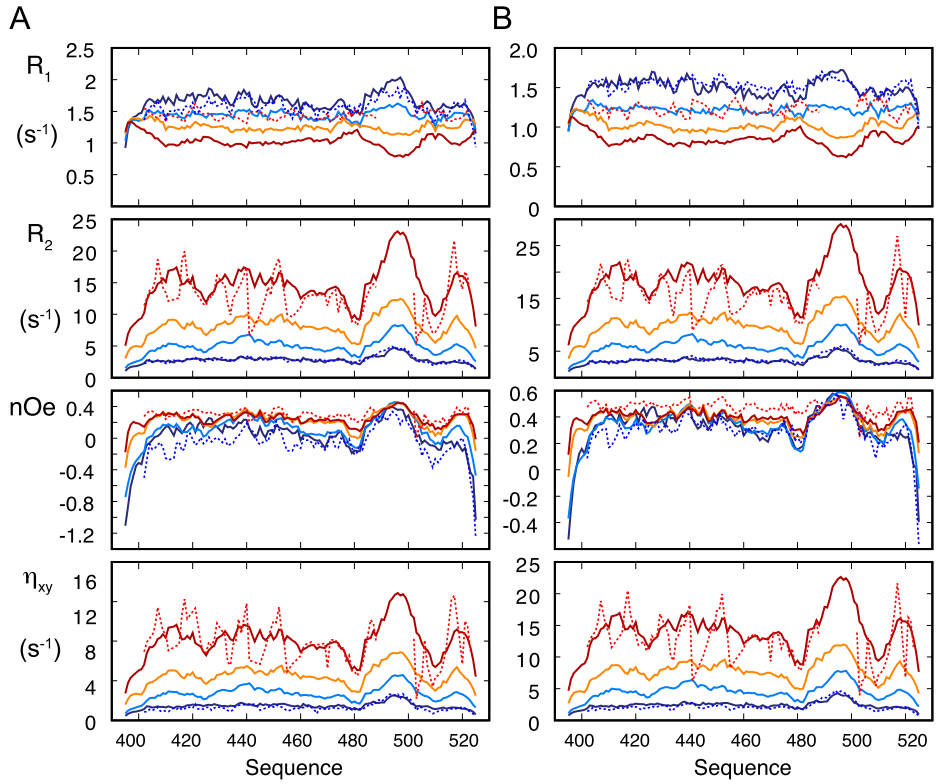


Figure 7. NMR relaxation rates derived from MD simulation as a function of self-crowding. (A) Solid lines show spin relaxation rates at 600 MHz, calculated from MD simulations representing: blue—dilute (1 isolated copy of N_{TAIL}) from ABSURD analysis, light blue—5 mM box (relaxation rates averaged over 125 copies over 230 ns), orange—14 mM box (relaxation rates averaged over 125 copies over 230 ns), and red—20 mM box (relaxation rates averaged over 343 copies over 500 ns). Dotted lines represent experimental rates measured in the dilute (blue) and dense (red) phases. (B) As in (A), for rates simulated and measured at 850 MHz.

the origin of this phenomenon using a similar analysis to an interatomic contact-model, previously demonstrated to predict fast-motional order parameters in folded proteins,[74] whereby spatial restriction is calculated for each available conformation sampled over the trajectory (see Materials and Methods). The distribution of the number of intramolecular contacts across the sequence (Figure 8) shows very similar features to the amplitude of the fast motion (presented here as $1 - A_1$) determined from the autocorrelation functions as a function of concentration and, in particular, exhibits the same tendency of increasing motional restriction as a function of protein concentration in the simulation (correlation plot is shown in Figure S18). This indicates that the experimentally observed reduction in fast motional amplitude results from an increase in intermolecular contacts, reflecting the increased density of heavy atoms, or intermolecular entanglement, in the vicinity of the observed backbone amides.

DISCUSSION

The importance of the phenomenon of protein condensation for a multitude of physiological and pathological intracellular processes is now generally recognized, underlining the importance of developing a physical understanding of the dynamic nature of these highly concentrated, yet fluid functional environments. In this study, we used extensive NMR relaxation measurements to compare the dynamic behavior of N_{TAIL}, a 125aa IDP, in dilute and dense liquid phases. While N_{TAIL} is treated here as a model system, phase separation of measles virus nucleoprotein in the presence of the viral phosphoprotein is suggested to be implicated in viral replication, allowing for genome encapsidation within these condensates.[12] In order to distinguish the effects of inert crowding from condensation, we also investigated the response of N_{TAIL} to increasing viscogen concentration under non-phase-separating conditions, and exploited extensive MD simulations, mimicking concentrations experienced in the dense phase, providing new insight into the physical origin of the observed dynamic parameters.

Although the absolute values of transverse relaxation data measured along the sequence of N_{TAIL} in the dense phase are significantly different from those measured in dilute and weakly crowded conditions, the profile is maintained to some extent. R_2, η_{xy}, and nOe exhibit similarly placed maxima and minima in both dense and dilute phases, with peaks concentrated around stretches of charged residues containing single hydrophobic residues (^{438}RRVK^{441} and ^{449}ESYRE^{453}) and a trough around the less-bulky sequence (^{479}SESS^{482}) (Figure S9). These similarities reproduce similar observations on the protein Caprin1[32] and tend to suggest that the physico-chemical characteristics underlying the dynamic behavior of the chain are not entirely remodeled in the dense state. Nevertheless, additional maxima are also observed in the dense phase around hydrophobic sequences, in particular ^{418}FLH^{420} and ^{517}VYN^{519}, possibly suggesting the existence of additional intermolecular contacts involving aromatic residues that are unique to the droplet form. We also note the striking absence of tapering of relaxation rates toward the termini of the sequence that is very often seen in IDPs in solution[18] (and in N_{TAIL}, as shown above). This indicates that high-amplitude fast motions, termed whip-lash motions in the studies of polymer

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

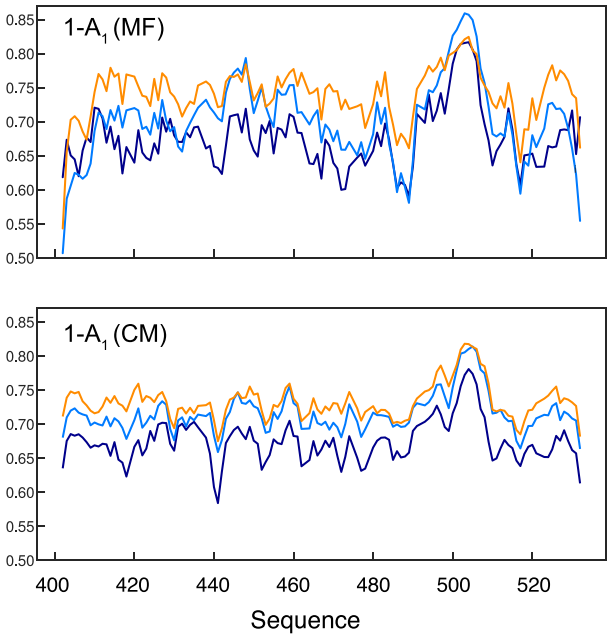


Figure 8. MD-derived interatomic contacts correlate with fast motional amplitudes. Top: Fast motional amplitudes $(1 - A_1)$ derived from the model-free analysis performed on the rates extracted from MD simulations performed as a function of protein concentration (orange—14 mM, light blue—5 mM, dark blue—dilute, the latter is derived from the ABSURD-derived ensemble of trajectories). Bottom: Contact model derived from the same trajectories (calculated from eq 2).

dynamics,[75] are strongly restricted in the dense phase compared to the dilute state and compared to dynamics observed in the presence of viscogen. We also note the presence of an apparent contribution to R_2 from conformational exchange, which appears to be in the slow regime on the NMR time-scale (100 ms to seconds, in view of the static field dependence), and that this contribution shows maxima in the vicinity of charged residues (^{408}R, ^{437}RRVK^{440}, and ^{520}DR^{521}).

In order to interpret the experimental relaxation data in a more quantitative manner, we turned to the model-free approach that models the reorientational autocorrelation function of the relaxation active interaction (in our case characterized by the reorientation of the N−H bond vector) in terms of the sum of exponentials representing independent motional modes. Using this approach, the characteristic correlation times of the different exponential components are determined, as well as their relative contributions to the autocorrelation function. We, and others,[53,57−61] have established and calibrated the validity of the amplitudes and timescales determined from such an approach for the description of IDP dynamics, also comparing to molecular simulation where possible.[55,60,62−64,76] Using extensive cross-validation for three independent IDPs, we have demonstrated the predictive nature of the approach, which is capable of accurately simulating independent relaxation rates, separate to the fitted data. Analysis of temperature-dependence of the resulting timescales and amplitudes of the modes, and as a function of molecular crowding, convincingly identifies three (fast, intermediate, and slow) components as librational, backbone conformational sampling and chain-like or segmental motions, respectively.[59,61] While motions of IDPs on these timescales are potentially much more complex than can be captured by three such generic modes, the proven ability to predict both higher and lower frequency relaxation data,[77] and the coherence of the resulting physical insight, offers convincing arguments to apply this approach to the current comparative study. Indeed, using extensive statistical testing, we have tested the validity of the model, demonstrating the necessity of three dynamic modes to explain experimental data and also showing that additional modes are not statistically required.

Model-free analysis of the multi-field experimental relaxation data measured from N_{TAIL} in the dilute and under weakly crowding conditions mirrors previous analyses of IDPs under similar conditions.[59,61] Fast motional timescales do not evolve as a function of crowding and have a characteristic time constant of approximately 50 ps. Intermediate timescale motions occurring around 1 ns and slower timescale (10 ns) chain-like motions exhibit very similar viscosity coefficients compared to Sendai NT and MKK4,[61] with the timescale of the slowest, segmental component again increasing faster than the intermediate component, a phenomenon previously assigned to the difference in dimensions of single peptide units and multi-peptide segments. The amplitude of the three components does not change as a function of viscogen concentration. Model-free analysis of the dynamics occurring in the dense phase reveals significant differences, including a notable slowing down of timescales associated with all three components of backbone motion, and a strong redistribution of their respective amplitudes. In an inversion of the distribution seen under dilute or weakly crowded conditions, fast, librational motions, occurring on timescales around 200 ps, now have the smallest amplitude, while slower segmental or chain-like motions dominate the correlation function. In order to further investigate the molecular origin of the apparent impact of self-crowding on the nature of dynamics within the dense phase, we ran a series of MD simulations at concentrations increasingly approaching those experienced under experimental phase separated conditions.

The creation of initial conformations for MD simulations of the dense phase is challenging, requiring the replication of known conformational sampling, as exemplified by the observed chemical shifts, while reproducing such high concentrations. Overall, the components of the ensemble should sample the appropriate structural heterogeneity while fitting sterically within the necessary volume. In order to achieve this aim, we used members of a conformational ensemble selected using the ASTEROIDS approach and filled the simulation boxes using random selection of conformers until the required concentration was achieved. The reproduction of experimental chemical shifts by the self-crowded simulations attests to the validity of the local conformational sampling within these self-crowded trajectories.

Although the aim of the simulation is not to reproduce phase separation, these calculations can provide important insight into the impact of self-crowding on IDP dynamics. In particular, 230 or 500 ns simulations provide enough sampling for converged correlation functions describing motions with timescales up to 25 or 30 ns, and the large number of molecules in each simulation (between 125 and 343 independent trajectories) delivers unprecedented statistical robustness for each spin pair.

Comparison of simulated relaxation rates from dilute to most concentrated conditions shows similar trends to differ-

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

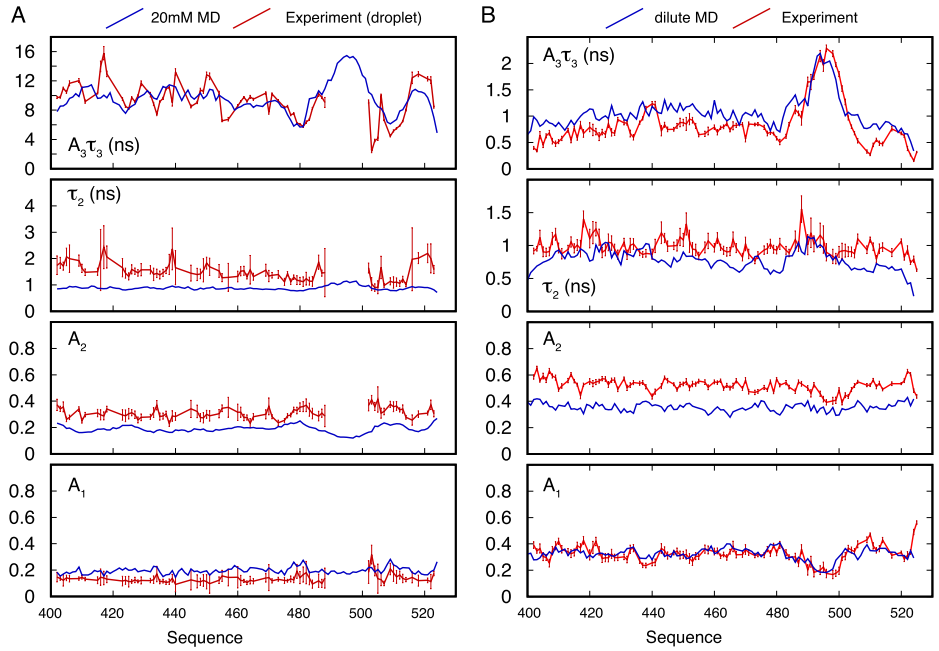


Figure 9. MD simulation of self-crowding reproduces global dynamic features of N_{TAIL} in dilute and dense phases. (A) Residue-specific comparison of dynamic parameters of N_{TAIL} in the dense phase from simulation (blue, parameters calculated from 343 copies of the protein comprising the 20 mM simulation) and experimental data from the dense phase (red). The product $A_3\tau_3$ defines the extent of the slowest component, τ_2 the intermediate timescale motion, and A_2 and A_1 the amplitude of the intermediate and fast timescale motions, respectively. (B) Residue-specific comparison of dynamic parameters of N_{TAIL} in the dilute phase from simulation (blue, parameters calculated from the ABSURD-derived ensemble of trajectories describing the dilute phase) and experiment (red). The product $A_3\tau_3$ defines the extent of the slowest component, τ_2 the intermediate timescale motion, and A_2 and A_1 the amplitude of the intermediate and fast timescale motions respectively.

ences observed between experimental data measured in dilute and dense phases, as summarized in Figure 9. Indeed model-free parameters derived from the simulations reproduce the overall distribution of dynamic parameters derived from experimental observations and in many cases the sequence-dependent variation of motional amplitudes and timescales. For example, glycine, serine, and alanine-containing regions flanking the α-helix (^{487}TASESS^{492}) and (^{509}GSDT^{512}) clearly exhibit higher flexibility (lower A_3) in both simulation and experiment.

Interestingly, A_1 decreases throughout the sequence as the concentration of the IDP increases, mirroring experimental differences between dilute and dense phases. Calculation of local contacts for each amide group suggests that the restricted order measured as a function of self-crowding, at least in part, results from the extensive intermolecular contacts with neighboring molecules. As noted above, this is not observed in conditions of inert crowding, either here or in our recent study of two other IDPs,[61] suggesting that self-crowding of IDPs, as experienced in liquid droplets, involves enhanced intermolecular entanglement and even self-solubilization, involving very close intermolecular contacts. Simulations of the most crowded conditions broadly reproduce the range of experimental R_2 and nOe but provide a poorer reproduction of experimental R_1, which are systematically higher in the dense phase, suggesting that the impact of self-crowding on nanosecond motions is not as accurately represented in this series of simulations as the faster and slower components. It is possible that the true intermolecular interactions that stabilize the droplet involve complex correlated nanosecond and supra-nanosecond motions. Indeed, at this level of protein concentration and entanglement, it seems highly likely that both intermediate and slow motional components comprise a significant component of longer-range collective dynamics.

CONCLUSIONS

We have compared the dynamic properties of an IDP that undergoes LLPS under dilute, crowded and phase-separated conditions using a combination of experimental and simulation approaches developed to understand the physical origin of NMR relaxation of IDPs. Despite barely impacting local backbone dihedral sampling of the free energy landscape of the protein, formation of the dense liquid phase nevertheless has a major impact on its dynamic properties. We observe a systematic slowing down of all three previously identified dynamic contributions, reporting on librational fluctuation, backbone dihedral angle dynamics and segmental or chain-like, and possibly collective dynamics. In addition to this overall lengthening of characteristic timescales, the relative amplitudes of the distinct components are also strongly affected by phase separation, with fast fluctuations being significantly restricted compared to both dilute and viscogen-crowded environments, suggesting that inter-molecular entanglement in the highly concentrated environment is responsible for restriction of rapid fluctuations. Overall, the combination of NMR relaxation and molecular simulation in increasingly self-crowded environments reproduces the overall features of the experimental dynamic parameters and provides new insight into the transformation of the dynamics of IDPs in biomolecular condensates.

MATERIALS AND METHODS

Protein Preparation. The N_{TAIL} domain (residues 401−525) of the nucleoprotein of Measles virus was cloned into pET22b with an

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

N-terminal 6His-tag and transformed into *Escherichia coli* BL21 (Novagen) for expression as described.[12] Briefly, the protein was expressed in *E. coli* Rosetta TM(lDE3)/pRARE (Novagen). Cell cultures were incubated at 37 °C until reaching an optical density of 0.6 and then at 20 °C overnight after induction with 1 mM isopropyl-*b*-D-thiogalactopyranoside. After cells were centrifuged at 5000 rpm for 20 min, pellets were resuspended in lysis buffer (150 mM NaCl, 20 mM Tris-HCl, pH 8.0) sonicated and centrifuged again at 18,000 rpm for 50 min. Proteins were subjected to nickel affinity chromatography and eluted with the lysis buffer containing 500 mM imidazole. The resulting solution was concentrated and subjected to size exclusion chromatography (SEC, Superdex75) in NMR buffer (150 mM NaCl, 50 mM sodium phosphate pH 6.0, and 2 mM dithiothreitol).

Establishment of the Phase Diagram of N_{TAIL} Condensation. Protein was mixed with PEG10000 solution at various NaCl concentrations, so that final N_{TAIL} was 300 μM and PEG 200 mg/mL. PEG and protein solutions were gently mixed by pipetting until homogeneous turbid solution was obtained. The dense phase was sedimented down by incubating tubes overnight in a PCR machine at the given temperature and briefly sedimented at 11,000 rpm. Volumes of both phases were measured, and then protein concentration was measured in the presence of guanidinium chloride. The dense phase was diluted 40 times before measurement and the dilute phase twice. This procedure was performed at four temperatures (282, 288, 298, and 308 K) and four NaCl concentrations (118, 181, 244, and 377 mM NaCl).

The phase diagram was fitted to the Flory–Huggins model at different NaCl concentrations using the expression for the energy of mixing

$$\Delta F_{mix} = kT\left(\frac{\phi}{N_{N_{TAIL}}}\ln\phi + \frac{1-\phi}{N_{solvent}}\ln(1-\phi) + \chi\phi(1-\phi)\right) \quad (1)$$

where ϕ is the volume fraction of the protein in the dense phase, N the polymer length, assuming that $N_{N_{TAIL}} = 130$, the number of amino acids and $N_{solvent} = 1$, and χ the interaction parameter.

Protein concentration was measured using nanodrop; three measurements for each sample were recorded and averaged. Extinction coefficients $\varepsilon = 2980\ M^{-1}\ cm^{-1}$ were used to determine concentrations from absorbance at 280 nm. Protein density of 1.35 g/cm^3[78] was used to calculate the volume fraction of the protein in both phases.

The impact of crowding on the dynamic behavior of N_{TAIL} was measured using PEG concentrations that lie in the dilute region of the phase diagram (0, 37.5, and 75 g/L PEG).

Fluorescence Microscopy. Fluorescence imaging and FRAP of fluorescein-labeled N_{TAIL} were performed with an Olympus IX81 spinning-disk confocal microscope equipped with a Yokogawa CSUX1 spinning disk head and diode-pumped solid-state lasers 488 nm (50 mW Coherent Sapphire). Emission light is filtered through a 520(28) nm filter; acquisition is performed with an iXon Ultra EMCCD (Electron Multiplying Charge-Coupled Device) (Andor) 512 × 512 pixel camera. FRAP was restricted to a region of interest, defined graphically, with a galvanometric FRAP/PA device (Andor) using the same lasers used for imaging with AOTF (Acousto-Optic Tunable Filter) 15% and with a dwell time of 50 ms for 10 repeats before bleaching and with a 2 s interval for 40 repeats after. The laser power and the dwell time during bleaching were optimized to bleach 50% of the initial fluorescence intensity. Images were analyzed using FIJI and python plugin ImageFRAP (https://imagej.net/Analyze_FRAP_movies_with_a_Jython_script); the recovery was fitted with a single exponent.

Biomolecular Condensate Preparation for NMR. Mixtures of ^{13}C,^{15}N-labeled and unlabeled N_{TAIL} (total concentration 2 mM, diverse admixtures) and PEG10000 (concentration 400 mg/mL) were mixed 1:1, inducing phase separation. The suspension was then sedimented at 1000 rpm and 298 K for 3 h. During this time, small droplets of dense phase present in the tubes coalesce into a single large droplet at the bottom of the tubes. Approximately 80 μL of dense N_{TAIL} phase was obtained from approximately 4 mL of 2 mM dilute phase solution and transferred into a 3 mm Shigemi tube using a glass capillary and manual centrifuge. Use of the Shigemi tube helps achieve a better static filled homogeneity ("shim") during the NMR measurements and results in narrower lineshapes in the ^1H dimension.

NMR Spectroscopy. NMR experiments were performed on Bruker spectrometers operating at ^1H frequencies of 600, 700, and 850 MHz. Spectra were processed using NMRPipe[79] and analyzed using NMRFAM-Sparky.[80] Relaxation rates were measured in NMR buffer, in dilute (no crowder), crowded (37.5 and 75 g/L PEG) samples, and in the dense phase, always at 298.1 K. ^{15}N R_1, $\{^1H\}-^{15}N$ heteronuclear NOE and $R_{1\rho}$ (with a spin lock of 1.5 kHz) were measured as described.[81] $^{15}N-^1H$ CSA/DD transverse cross-correlated (η_{xy}) cross-relaxation rates were measured as described.[82] Relaxation delays for R_1 and $R_{1\rho}$ in the dense phase were [0, 0.6, 0.08, 1.6, 0.4, 0.32, 0.1, 0.2, 0.6]s and [1, 15, 35, 120, 75, 22, 6, 15, 190]ms, respectively (including repetition of one delay). $R_{1,water}$ was measured using saturation recovery with 20 delays from 1 ms to 12 s. Data were analyzed using NMRPipe and NMRglue.[83] ^{15}N relaxation dispersion was measured in the dense phase at 950 MHz as previously described,[84] using 14 points at CPMG frequencies between 31 and 1000 Hz with a constant-time relaxation delay of 32 ms.[85]

Preparation of Boxes of Concentrated Proteins. Each box was generated using dimensions calculated from the number of proteins and the target protein concentration. Coordinates were then defined for each box as a 3-dimensional grid, and a molecule of N_{TAIL} was randomly chosen and iteratively inserted at the given location with a randomly chosen conformation from an ASTEROIDS ensemble.[69,70] The process was repeated until no clash was found, using reduced (0.15) van der Waals radii, between the new protein and the already added proteins. Using this method, three boxes of 5 mM, 14 mM, and 20 mM were generated with a total of 125, 125, and 343 proteins, respectively. Water molecules and Na^+ and Cl^- ions were then added in the box to neutralize the charges, and the system energy was minimized using standard procedures from GROMACS. Chemical shift calculation over the resulting boxes verified that the starting configurations were in agreement with experimentally measured backbone chemical shifts.

MD Simulation of Single Molecule of N_{TAIL} in the Dilute Phase. The rectangular cuboid simulation boxes were composed by a single N_{TAIL} construct with approximately 100,000 water molecules with Na^+ and Cl^- ions corresponding to a salt concentration of 150 mM. 30 independent boxes were generated with a different protein conformation randomly selected among a previously derived ASTEROIDS ensemble.[69,70] GROMACS version 2019.4[86] was used to calculate the trajectories. Each box was subjected to an energy minimization process using steepest descent with 10,000 steps followed by 500 ps and 2 ns equilibration steps in the *NVE* and *NPT* ensembles, respectively. The trajectories were calculated in the *NVT* ensemble for 200 ns with a time step of 2 fs. The coordinates were saved every 2 ps. CHARMM36m (C36m) was combined with TIP4P/2005 water as we have previously shown this combination to yield good results for IDPs, including the homologous N_{TAIL} protein from Sendai virus.[64] Holonomic constraints were applied on the hydrogen-heavy atom bond terms to remove fast modes of oscillation, and a cutoff of 1.2 nm was used for the Lennard-Jones and electrostatic interactions[87] (this has been shown to exert negligible impact on calculated relaxation rates[88]). Particle-mesh Ewald summation with a grid spacing of 0.16 nm was used to calculate long-range electrostatic interactions.[89] Temperature coupling is performed using velocity rescaling[90] with a time constant of 0.5 ps and a temperature target of 298.15 K. During the *NPT* equilibration, pressure was controlled with a target of 1.0 bar using a Parrinello–Rahman pressure coupling performed with a time constant of 2 ps.[91]

MD Simulation of Boxes of Concentrated Proteins. Prior to production, each box was subjected to an energy minimization procedure using steepest descent with 5000 steps followed by 20 ns

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548–10563

NPT equilibration. Each trajectory was then calculated in the *NPT* ensemble for 200−230 ns with a time step of 1 or 2 fs. The trajectory calculated at the highest concentration (20 mM) was extended to 500 ns to help convergence of angular correlation functions of the slowest motions. The coordinates were saved every 2 ps. Holonomic constraints are applied on the hydrogen-heavy atom bond terms to remove fast modes of oscillation, and a cutoff of 0.9 nm was used for the Lennard-Jones and electrostatic interactions. Particle-mesh Ewald summation with a grid spacing of 0.16 nm was used to calculate long-range electrostatic interactions. In order to maintain stability in the highly crowded solutions, temperature coupling was performed using velocity rescaling[90] with a time constant of 100 ps and a temperature target of 298K. During the *NPT* equilibration, pressure was controlled with a target of 1.0 bar using a Berendsen pressure coupling performed with a time constant of 20 ps.

Calculation of Spin Relaxation Rates. Spin relaxation rates were calculated as described previously.[62] Briefly, for a set of trajectories, rotational autocorrelation functions for the amide bond vectors of each trajectory were calculated up to a maximum lag time corresponding to half the length of the trajectory. A Tikhonov regularization procedure was used to map the average of all the trajectories for each residue onto a grid of 8192 timescales. The associated spectral density function was then computed from the obtained parameters. The resulting spectral density functions were then used to compute all the relaxation rates at the wanted magnetic fields using the conventional formulas (see below). For single protein trajectories, each trajectory was segmented into three 100 ns segments prior to analysis, and the Average Block Selection Using Relaxation Data (ABSURD)[62] procedure was applied as described previously targeting R_2 rates measured at 850 MHz. We note that an ABSURD analysis would only be appropriate for the concentration-dependent study if multiple boxes had been simulated at each concentration, which is outside the scope of currently available calculation. Relaxation rates at each concentration are therefore calculated for each member of the ensemble of conformers in each box and compared as a function of concentration.

Radii of gyration were calculated using the standard GROMACS *gyrate* command. Chemical shift prediction was performed using SPARTA+.[92] Spatial restriction was calculated for each N^H amide i, sampled over the trajectory using the following expression

$$(1 - A_1)_i = \frac{1}{2}\tanh\left(C\sum_j^{\text{atoms}}\left\langle \exp\left(-\frac{kr_{ij}}{m_j}\right)\right\rangle + B\right) + \frac{1}{2} \tag{2}$$

where k is a variable parameter controlling the range of the distance-dependent interaction, r_{ij} and m_j represent the distance to the heavy atoms (j) and their mass, and the sum is taken over all j atoms. Equation 2 was calculated for each snapshot of the trajectories from which the autocorrelation function was calculated and averaged. B and C are adjustable parameters optimized to best agree with the entire dataset.

Distances were calculated using the python library Pytraj. The parameter defined by eq 2 for a given residue was averaged over time and over all the trajectories. Heavy atoms that are at a distance of less than 2.5 Å from the edges of the box were not taken into account for this calculation.

Modeling of ^{15}N Relaxation Rates. ^{15}N relaxation rates are given by the known functions

$$R_1 = \frac{1}{10}\left(\frac{\mu_0\hbar\gamma_H\gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2(\sigma_\parallel - \sigma_\perp)^2 J(\omega_N) \tag{3}$$

$$R_2 = \frac{1}{20}\left(\frac{\mu_0\hbar\gamma_H\gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2(\sigma_\parallel - \sigma_\perp)^2 (4J(0) + 3J(\omega_N)) \tag{4}$$

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0\hbar\gamma_H\gamma_N}{4\pi r_{NH}^3}\right)^2 (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)) \tag{5}$$

$$\eta_{xy} = \frac{1}{15}P_2(\cos\theta)\left(\frac{\mu_0\hbar\gamma_H\gamma_N}{4\pi r_{NH}^3}\right)(\sigma_\parallel - \sigma_\perp)\omega_N(4J(0) + 3J(\omega_N)) \tag{6}$$

$J(\omega)$ is the angular spectral density function at frequency ω, $\hbar$ Planck's constant, μ_0 the permittivity of free space, and θ the angle between the principal axis of the CSA tensor ($\sigma_\parallel - \sigma_\perp = -172$ ppm) and the NH bond vector. r_{NH} is the N−H internuclear distance (assumed to be 1.015 Å) and γ_N and γ_H are gyromagnetic ratios of ^{15}N and ^{1}H, respectively.

The correlation and spectral density functions are modeled as described previously[18] using three components represented by distinct exponential time constants whose relative amplitudes A_k are governed by $\sum_k A_k = 1$.

$$C(t) = \sum_k A_k e^{-t/\tau_k} \tag{7a}$$

and

$$J(\omega) = \sum_k A_k\tau_k/(1 + \omega^2\tau_k^2) \tag{7b}$$

These data, from the dilute samples, were interpreted collectively using a recently proposed approach[61] correlating the timescales of the dynamics of intermediate and slow timescale motions (τ_k) in buffer and under the two crowding conditions with the nano-viscosity of the surrounding solvent

$$\tau_k(C, T) = \tau'_{k,\infty}(\varepsilon_k\rho(C) + 1) \tag{8}$$

$\rho(C) = (\eta_C - \eta_0)/\eta_0 = (R_{1,C} - R_{1,0})/R_{1,0}$ is the solvent friction and η_0 and $R_{1,0}$ are the viscosity and longitudinal relaxation rate of water, $R_{1,C}$ and η_C are the longitudinal relaxation and viscosity at a given crowder concentration. $\tau'_{k,\infty}$ is a pre-factor and ε_k residue and mode-specific friction coefficients that describe how a peptide chain with a given primary sequence responds to increasing viscogen concentration.

An optimization algorithm was written to fit data to the expressions in eqs 3−6 using the definition of the spectral density function associated with eqs 7a and 7b, minimizing the following function

$$\chi_i^2 = \sum_{n=1}^{5}\sum_{m=1}^{N}\left\{\frac{(R_{n,\text{exp}}^m - R_{n,\text{calc}}^m)}{\sigma_{n,\text{exp}}^m}\right\}^2 \tag{9}$$

for each residue i, where n refers to the different rates and m the different conditions (e.g., crowding). In the case of the combined fit of N_{TAIL} in dilute and low concentrations of PEG, 20 experimental rates were fit to an 8-parameter (ε_3, ε_2, $\tau_{3,\infty}$, $\tau_{2,\infty}$, τ_1, θ, A_2, and A_3) fit for each residue, where ε_2 and ε_3 are the viscosity coefficients of the intermediate and slow timescale motional components, respectively, $\tau_{2,\infty}$ and $\tau_{3,\infty}$ are their correlation times at infinite dilution. ε_1 is set to 0 as in previous applications of this procedure.[61] These site-specific derived values of θ were used in the 5-parameter (A_1, A_2, τ_1, τ_2, τ_3) fit of dense phase data.

Errors were estimated using noise-based Monte-Carlo approaches, and all presented data respect 95% confidence limits. Data-fitting programs are available on request.

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c13647.

^{1}H spectra of dilute and dense phases, phase diagram of N_{TAIL} fluorescence images showing FRAP, secondary chemical shifts and correlation spectra in the dilute and dense phases, back-calculated relaxation rates from the dilute phase, model-free analysis of N_{TAIL} as a function of viscosity, Figure 4 with highlighted regions of interest, data fitting, relaxation dispersion and data cross-validation from the dense phase, ABSURD analysis of the dilute phase relaxation data, chemical shifts from the dilute and concentrated simulations, dimensions of the proteins in the dense phase, sample autocorrelation functions, and correlation of A_1 from contact and model-free analyses (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Martin Blackledge − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France;* orcid.org/0000-0003-0935-721X; Email: martin.blackledge@ibs.fr

Authors

Serafima Guseva − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France;* Present Address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, United States

Vincent Schnapka − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France*

Wiktor Adamski − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France;* Present Address: University of Lille, Risk factors and molecular determinants of aging-related disease UMR 1167, Institut Pasteur de Lille, CNRS, Equipe Biologie Structurale Integrative EMR 9002, Campus CNRS de la Haute Borne, F-59658 Villeneuve d'Ascq, France.

Damien Maurin − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France*

Rob W. H. Ruigrok − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France*

Nicola Salvi − *Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS, 38000 Grenoble, France;* Present Address: Bio Structure and Biophysics (BSB), Sanofi R&D, 94400 Vitry-sur Seine, France.

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c13647

Author Contributions

‖S.G. and V.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the European Research Council Advanced Grant DynamicAssemblies under the European Union's Horizon 2020 research and innovation program (grant agreement number 835161) to M.B. This work was supported by the ANR (NanoDisPro ANR-18-CE29-0003, LiquidFact ANR-19-CE15-0024, and TempSens ANR-19-CE20-0021-02), the Labex GRAL (ANR-10-LABX-49-01). We acknowledge the platforms of the Grenoble Instruct European Research Infrastructure Consortium (Integrated Structural Biology Grenoble; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology. Platform access was supported by French Infrastructure for Integrated Structural Biology (ANR-10-INBS-05-02) and the Grenoble Alliance for Integrated Structural and Cell Biology, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). The Institut de Biologie Structurale acknowledges integration into the Interdisciplinary Research Institute of Grenoble. This work was performed using CCRT HPC resource (TOPAZE supercomputer). MD simulations were carried out using the HPC resources of GENCI-TGCC funded via the project 2021SA10spe00014 (Prèparation à l'Exascale"). We acknowledge the platform supported by GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

■ REFERENCES

(1) Hyman, A. A.; Simons, K. Beyond Oil and Water—Phase Transitions in Cells. *Science* **2012**, *337*, 1047−1049.

(2) Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **1942**, *10*, 51−61.

(3) Huggins, M. L. Some Properties of Solutions of Long-Chain Compounds. *J. Phys. Chem.* **1942**, *46*, 151−158.

(4) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176*, 419−434.

(5) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A Framework for Understanding the Functions of Biomolecular Condensates across Scales. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 215−235.

(6) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer Physics of Intracellular Phase Transitions. *Nat. Phys.* **2015**, *11*, 899−904.

(7) Shin, Y.; Brangwynne, C. P. Liquid Phase Condensation in Cell Physiology and Disease. *Science* **2017**, *357*, No. eaaf4382.

(8) Berry, J.; Brangwynne, C. P.; Haataja, M. Physical Principles of Intracellular Organization via Active and Passive Phase Transitions. *Rep. Prog. Phys.* **2018**, *81*, 046601.

(9) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688−699.e16.

(10) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **2020**, *49*, 107−133.

(11) Fawzi, N. L.; Parekh, S. H.; Mittal, J. Biophysical Studies of Phase Separation Integrating Experimental and Computational Methods. *Curr. Opin. Struct. Biol.* **2021**, *70*, 78−86.

(12) Guseva, S.; Milles, S.; Jensen, M. R.; Salvi, N.; Kleman, J.-P.; Maurin, D.; Ruigrok, R. W. H.; Blackledge, M. Measles Virus Nucleo- and Phosphoproteins Form Liquid-like Phase-Separated Compartments That Promote Nucleocapsid Assembly. *Sci. Adv.* **2020**, *6*, No. eaaz7095.

(13) Peeples, W.; Rosen, M. K. Mechanistic Dissection of Increased Enzymatic Rate in a Phase-Separated Compartment. *Nat. Chem. Biol.* **2021**, *17*, 693−702.

(14) Sang, D.; Shu, T.; Pantoja, C. F.; Ibáñez de Opakua, A.; Zweckstetter, M.; Holt, L. J. Condensed-Phase Signaling Can Expand

Kinase Specificity and Respond to Macromolecular Crowding. *Mol. Cell* **2022**, *82*, 3693−3711.e10.
(15) Abyzov, A.; Blackledge, M.; Zweckstetter, M. Conformational Dynamics of Intrinsically Disordered Proteins Regulate Biomolecular Condensate Chemistry. *Chem. Rev.* **2022**, *122*, 6719−6748.
(16) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483*, 336−340.
(17) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.; Longgood, J.; Pei, J.; Grishin, N. V.; Frantz, D. E.; Schneider, J. W.; Chen, S.; Li, L.; Sawaya, M. R.; Eisenberg, D.; Tycko, R.; McKnight, S. L. Cell-Free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **2012**, *149*, 753−767.
(18) Camacho-Zarco, A. R.; Schnapka, V.; Guseva, S.; Abyzov, A.; Adamski, W.; Milles, S.; Jensen, M. R.; Zidek, L.; Salvi, N.; Blackledge, M. NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem. Rev.* **2022**, *122*, 9331−9356.
(19) Wei, M.-T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.-H.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. Phase Behaviour of Disordered Proteins Underlying Low Density and High Permeability of Liquid Organelles. *Nat. Chem.* **2017**, *9*, 1118−1125.
(20) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E. Structural and Hydrodynamic Properties of an Intrinsically Disordered Region of a Germ Cell-Specific Protein on Phase Separation. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E8194−E8203.
(21) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi Contacts Are an Overlooked Protein Feature Relevant to Phase Separation. *eLife* **2018**, *7*, No. e31486.
(22) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Ströhl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Kaminski Schierle, G. S.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **2018**, *173*, 720−734.e15.
(23) Schuster, B. S.; Dignon, G. L.; Tang, W. S.; Kelley, F. M.; Ranganath, A. K.; Jahnke, C. N.; Simpkins, A. G.; Regy, R. M.; Hammer, D. A.; Good, M. C.; Mittal, J. Identifying Sequence Perturbations to an Intrinsically Disordered Protein That Determine Its Phase-Separation Behavior. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 11421−11431.
(24) Greig, J. A.; Nguyen, T. A.; Lee, M.; Holehouse, A. S.; Posey, A. E.; Pappu, R. V.; Jedd, G. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol. Cell* **2020**, *77*, 1237−1250.e4.
(25) Ryan, V. H.; Dignon, G. L.; Zerze, G. H.; Chabata, C. V.; Silva, R.; Conicella, A. E.; Amaya, J.; Burke, K. A.; Mittal, J.; Fawzi, N. L. Mechanistic View of HnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell* **2018**, *69*, 465−479.e7.
(26) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules That Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **2015**, *60*, 231−241.
(27) Kim, T. H.; Tsang, B.; Vernon, R. M.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. Phospho-Dependent Phase Separation of FMRP and CAPRIN1 Recapitulates Regulation of Translation and Deadenylation. *Science* **2019**, *365*, 825−829.
(28) Majumdar, A.; Dogra, P.; Maity, S.; Mukhopadhyay, S. Liquid−Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* **2019**, *10*, 3929−3936.
(29) Emmanouilidis, L.; Esteban-Hofer, L.; Jeschke, G.; Allain, F. H.-T. Structural Biology of RNA-Binding Proteins in the Context of Phase Separation: What NMR and EPR Can Bring? *Curr. Opin. Struct. Biol.* **2021**, *70*, 132−138.
(30) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular Interactions Underlying Liquid−liquid Phase Separation of the FUS Low-Complexity Domain. *Nat. Struct. Mol. Biol.* **2019**, *26*, 637−648.
(31) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and Patterning of Aromatic Residues Determine the Phase Behavior of Prion-like Domains. *Science* **2020**, *367*, 694−699.
(32) Kim, T. H.; Payliss, B. J.; Nosella, M. L.; Lee, I. T. W.; Toyama, Y.; Forman-Kay, J. D.; Kay, L. E. Interaction Hot Spots for Phase Separation Revealed by NMR Studies of a CAPRIN1 Condensed Phase. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2104897118.
(33) Taylor, N. O.; Wei, M.-T.; Stone, H. A.; Brangwynne, C. P. Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys. J.* **2019**, *117*, 1285−1300.
(34) Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid−Liquid Phase Separation of the Microtubule-Binding Repeats of the Alzheimer-Related Protein Tau. *Nat. Commun.* **2017**, *8*, 275.
(35) Emmanouilidis, L.; Esteban-Hofer, L.; Damberger, F. F.; de Vries, T.; Nguyen, C. K. X.; Ibáñez, L. F.; Mergenthal, S.; Klotzsch, E.; Yulikov, M.; Jeschke, G.; Allain, F. H.-T. NMR and EPR Reveal a Compaction of the RNA-Binding Protein FUS upon Droplet Formation. *Nat. Chem. Biol.* **2021**, *17*, 608−614.
(36) Schuler, B.; Hofmann, H. Single-Molecule Spectroscopy of Protein Folding Dynamics–Expanding Scope and Timescales. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36−47.
(37) Ray, S.; Singh, N.; Kumar, R.; Patel, K.; Pandey, S.; Datta, D.; Mahato, J.; Panigrahi, R.; Navalkar, A.; Mehra, S.; Gadhe, L.; Chatterjee, D.; Sawner, A. S.; Maiti, S.; Bhatia, S.; Gerez, J. A.; Chowdhury, A.; Kumar, A.; Padinhateeri, R.; Riek, R.; Krishnamoorthy, G.; Maji, S. K. α-Synuclein aggregation nucleates through liquid−liquid phase separation. *Nat. Chem.* **2020**, *12*, 705−716.
(38) Wen, J.; Hong, L.; Krainer, G.; Yao, Q.-Q.; Knowles, T. P. J.; Wu, S.; Perrett, S. Conformational Expansion of Tau in Condensates Promotes Irreversible Aggregation. *J. Am. Chem. Soc.* **2021**, *143*, 13056−13064.
(39) Hong, Y.; Najafi, S.; Casey, T.; Shea, J.-E.; Han, S.-I.; Hwang, D. S. Hydrophobicity of Arginine Leads to Reentrant Liquid-Liquid Phase Separation Behaviors of Arginine-Rich Proteins. *Nat. Commun.* **2022**, *13*, 7326.
(40) Dignon, G. L.; Zheng, W.; Mittal, J. Simulation Methods for Liquid-Liquid Phase Separation of Disordered Proteins. *Curr. Opin. Chem. Eng.* **2019**, *23*, 92−98.
(41) Ruff, K. M.; Pappu, R. V.; Holehouse, A. S. Conformational Preferences and Phase Behavior of Intrinsically Disordered Low Complexity Sequences: Insights from Multiscale Simulations. *Curr. Opin. Struct. Biol.* **2019**, *56*, 1−10.
(42) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.* **2020**, *71*, 53−75.
(43) Borcherds, W.; Bremer, A.; Borgia, M. B.; Mittag, T. How Do Intrinsically Disordered Protein Regions Encode a Driving Force for Liquid-Liquid Phase Separation? *Curr. Opin. Struct. Biol.* **2021**, *67*, 41−50.
(44) Shea, J.-E.; Best, R. B.; Mittal, J. Physics-Based Computational and Theoretical Approaches to Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2021**, *67*, 219−225.
(45) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence Determinants of Protein Phase Behavior from a Coarse-Grained Model. *PLoS Comput. Biol.* **2018**, *14*, No. e1005941.

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

(46) Choi, J.-M.; Dar, F.; Pappu, R. V. LASSI: A Lattice Model for Simulating Phase Transitions of Multivalent Proteins. *PLoS Comput. Biol.* **2019**, *15*, No. e1007028.
(47) Benayad, Z.; von Bülow, S.; Stelzl, L. S.; Hummer, G. Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. *J. Chem. Theory Comput.* **2021**, *17*, 525−537.
(48) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. *J. Phys. Chem. B* **2020**, *124*, 11671−11679.
(49) Yuwen, T.; Brady, J. P.; Kay, L. E. Probing Conformational Exchange in Weakly Interacting, Slowly Exchanging Protein Systems via Off-Resonance R1ρ Experiments: Application to Studies of Protein Phase Separation. *J. Am. Chem. Soc.* **2018**, *140*, 2115−2126.
(50) Reichheld, S. E.; Muiznieks, L. D.; Keeley, F. W.; Sharpe, S. Direct Observation of Structure and Dynamics during Phase Separation of an Elastomeric Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E4408−E4415.
(51) Wong, L. E.; Kim, T. H.; Muhandiram, D. R.; Forman-Kay, J. D.; Kay, L. E. NMR Experiments for Studies of Dilute and Condensed Protein Phases: Application to the Phase-Separating Protein CAPRIN1. *J. Am. Chem. Soc.* **2020**, *142*, 2471−2489.
(52) Lipari, G.; Szabo, A. Model-Free Approach To The Interpretation Of Nuclear Magnetic-Resonance Relaxation In Macromolecules .1. Theory And Range Of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(53) Brutscher, B.; Brüschweiler, R.; Ernst, R. R. Backbone Dynamics and Structural Characterization of the Partially Folded A State of Ubiquitin by 1H, 13C, and 15N Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **1997**, *36*, 13043−13053.
(54) Yang, D. W.; Mok, Y. K.; FormanKay, J. D.; Farrow, N. A.; Kay, L. E. Contributions to Protein Entropy and Heat Capacity from Bond Vector Motions Measured by NMR Spin Relaxation. *J. Mol. Biol.* **1997**, *272*, 790−804.
(55) Prompers, J. J.; Bruschweiler, R. General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.
(56) Kaderavek, P.; Zapletal, V.; Rabatinova, A.; Krasny, L.; Sklenar, V.; Zidek, L. Spectral Density Mapping Protocols for Analysis of Molecular Motions in Disordered Proteins. *J. Biomol. NMR* **2014**, *58*, 193−207.
(57) Khan, S. N.; Charlier, C.; Augustyniak, R.; Salvi, N.; Dejean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F. Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys. J.* **2015**, *109*, 988−999.
(58) Gill, M. L.; Byrd, R. A.; Palmer III, A. G.; Palmer, I. I. I. Dynamics of GCN4 Facilitate DNA Interaction: A Model-Free Analysis of an Intrinsically Disordered Region. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5839−5849.
(59) Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M. Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J. Am. Chem. Soc.* **2016**, *138*, 6240−6251.
(60) Kämpf, K.; Izmailov, S. A.; Rabdano, S. O.; Groves, A. T.; Podkorytov, I. S.; Skrynnikov, N. R. What Drives 15N Spin Relaxation in Disordered Proteins? Combined NMR/MD Study of the H4 Histone Tail. *Biophys. J.* **2018**, *115*, 2348−2367.
(61) Adamski, W.; Salvi, N.; Maurin, D.; Magnat, J.; Milles, S.; Jensen, M. R.; Abyzov, A.; Moreau, C. J.; Blackledge, M. A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *J. Am. Chem. Soc.* **2019**, *141*, 17817−17829.
(62) Salvi, N.; Abyzov, A.; Blackledge, M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **2016**, *7*, 2483−2489.
(63) Salvi, N.; Abyzov, A.; Blackledge, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angew. Chem., Int. Ed.* **2017**, *56*, 14020.
(64) Salvi, N.; Abyzov, A.; Blackledge, M. Solvent-Dependent Segmental Dynamics in Intrinsically Disordered Proteins. *Sci. Adv.* **2019**, *5*, No. eaax2348.
(65) Persson, E.; Halle, B. Cell Water Dynamics on Multiple Time Scales. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6266−6271.
(66) Kimmich, R.; Fatkullin, N. Self-Diffusion Studies by Intra- and Inter-Molecular Spin-Lattice Relaxometry Using Field-Cycling: Liquids, Plastic Crystals, Porous Media, and Polymer Segments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *101*, 18−50.
(67) Korb, J.-P. Multiscale Nuclear Magnetic Relaxation Dispersion of Complex Liquids in Bulk and Confinement. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *104*, 12−55.
(68) Stenström, O.; Champion, C.; Lehner, M.; Bouvignies, G.; Riniker, S.; Ferrage, F. How Does It Really Move? Recent Progress in the Investigation of Protein Nanosecond Dynamics by NMR and Simulation. *Curr. Opin. Struct. Biol.* **2022**, *77*, 102459.
(69) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W. H.; Blackledge, M. Intrinsic Disorder in Measles Virus Nucleocapsids. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9839−9844.
(70) Jensen, M. R.; Blackledge, M. Testing the Validity of Ensemble Descriptions of Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1557−E1558.
(71) Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V. Intrinsically Disordered Linkers Determine the Interplay between Phase Separation and Gelation in Multivalent Proteins. *eLife* **2017**, *6*, No. e30294.
(72) Boeynaems, S.; Holehouse, A. S.; Weinhardt, V.; Kovacs, D.; Van Lindt, J.; Larabell, C.; Van Den Bosch, L.; Das, R.; Tompa, P. S.; Pappu, R. V.; Gitler, A. D. Spontaneous Driving Forces Give Rise to Protein-RNA Condensates with Coexisting Phases and Complex Material Properties. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 7889−7898.
(73) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. Long-Range Interactions within a Nonnative Protein. *Science* **2002**, *295*, 1719−1722.
(74) Zhang, F.; Brüschweiler, R. Contact Model for the Prediction of NMR N-H Order Parameters in Globular Proteins. *J. Am. Chem. Soc.* **2002**, *124*, 12654−12655.
(75) Kimmich, R.; Fatkullin, N. Polymer Chain Dynamics and NMR. In *NMR 3D Analysi Photopolymerization*; Fatkullin, N., Ikehara, T., Jinnai, H., Kawata, S., Kimmich, R., Nishi, T., Nishikawa, Y., Sun, H.-B., Eds.; Apvances in Polymer Science; Springer: Berlin, Heidelberg, 2004; pp 1−113.
(76) Prompers, J. J.; Brüschweiler, R. Reorientational Eigenmode Dynamics: A Combined MD/NMR Relaxation Analysis Method for Flexible Parts in Globular Proteins. *J. Am. Chem. Soc.* **2001**, *123*, 7305−7313.
(77) Salvi, N.; Zapletal, V.; Jaseňáková, Z.; Zachrdla, M.; Padrta, P.; Narasimhan, S.; Marquardsen, T.; Tyburn, J.-M.; Žídek, L.; Blackledge, M.; Ferrage, F.; Kadeřávek, P. Convergent Views on Disordered Protein Dynamics from NMR and Computational Approaches. *Biophys. J.* **2022**, *121*, 3785−3794.
(78) Fischer, H.; Polikarpov, I.; Craievich, A. F. Average Protein Density Is a Molecular-Weight-Dependent Function. *Protein Sci.* **2004**, *13*, 2825−2828.
(79) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe - A Multidimensional Spectral Processing System Based On Unix Pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.
(80) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31*, 1325−1327.
(81) Lakomek, N.-A.; Ying, J.; Bax, A. Measurement of ^{15}N Relaxation Rates in Perdeuterated Proteins by TROSY-Based Methods. *J. Biomol. NMR* **2012**, *53*, 209−221.

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

(82) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. Symmetrical Reconversion: Measuring Cross-Correlation Rates with Enhanced Accuracy. *J. Magn. Reson.* **2003**, *161*, 258−264.
(83) Helmus, J. J.; Jaroniec, C. P. N. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR* **2013**, *55*, 355−367.
(84) Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D. F.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M. Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *J. Am. Chem. Soc.* **2015**, *137*, 1220−1229.
(85) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. An Improved 15N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins. *J. Phys. Chem. B* **2008**, *112*, 5898−5904.
(86) 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.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(87) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(88) Ali, A. A. A. I.; Hoffmann, F.; Schäfer, L. V.; Mulder, F. A. A. Probing Methyl Group Dynamics in Proteins by NMR Cross-Correlated Dipolar Relaxation and Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2022**, *18*, 7722−7732.
(89) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(90) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(91) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(92) Shen, Y.; Bax, A. SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48*, 13−22.

https://doi.org/10.1021/jacs.2c13647
J. Am. Chem. Soc. 2023, 145, 10548−10563

Supporting Information

Liquid-liquid phase separation modifies the dynamic properties of intrinsically disordered proteins

Serafima Guseva[†], Vincent Schnapka[†], Wiktor Adamski, Damien Maurin, R.W.H. Ruigrok, Nicola Salvi, Martin Blackledge*

Institut de Biologie Structurale,

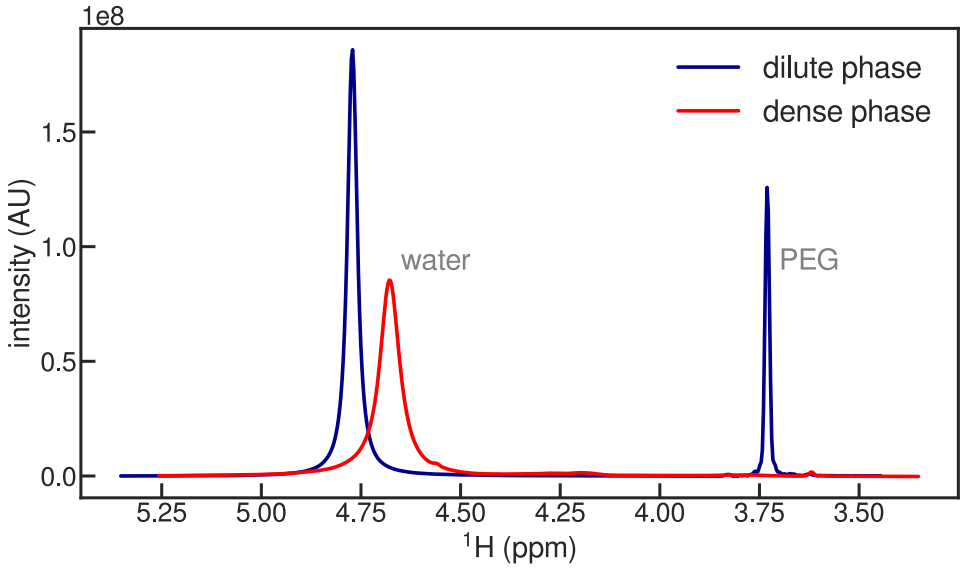
Université Grenoble Alpes-CEA-CNRS

71, Avenue des Martyrs

Grenoble

France

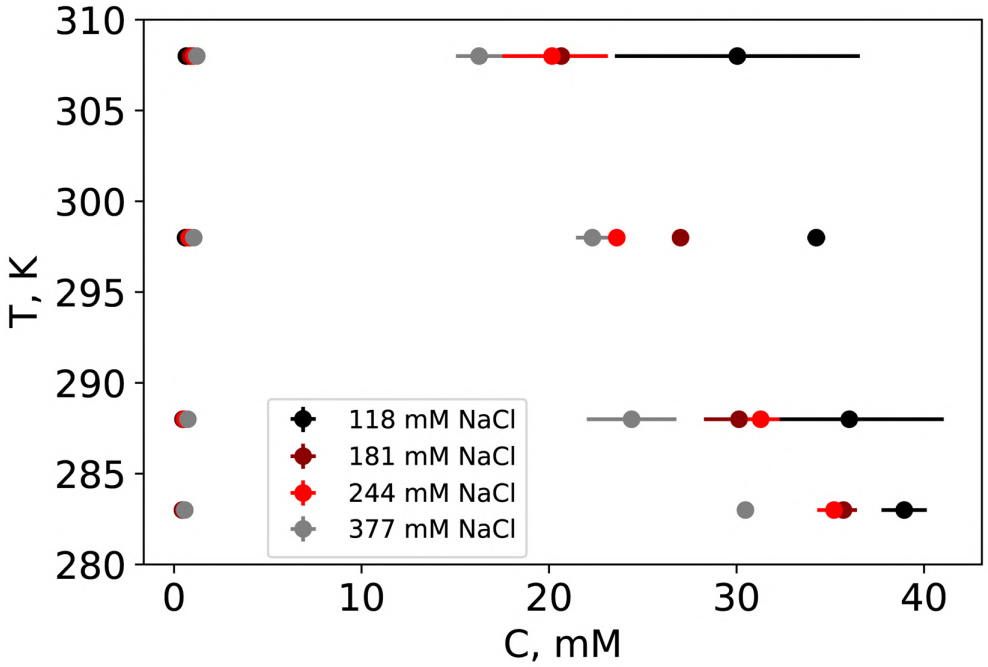
Figure S1



PEG is predominantly localised in the dilute phase of phase-separated N_{TAIL}-PEG mixtures

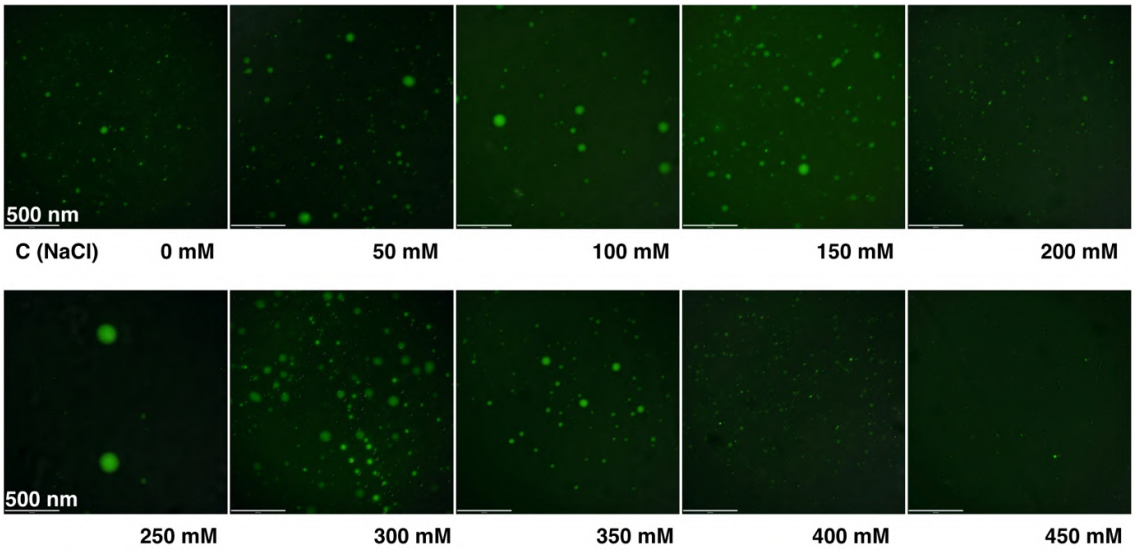
^{1}H NMR spectra of (a) dense phase, and (b) the coexisting dilute, indicating that PEG is predominantly localised to the dilute phase.

Figure S2



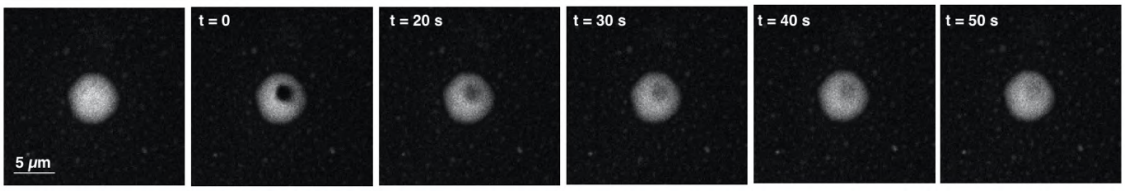
N_{TAIL} phase diagram (molar concentration C) as a dependency on temperature. Different colours correspond to different salt concentrations (black -118, dark red - 181, red - 244 and grey - 377 mM NaCl).

Figure S3A



Fluorescent microscopy images of N_{TAIL} droplets taken at different NaCl concentrations.

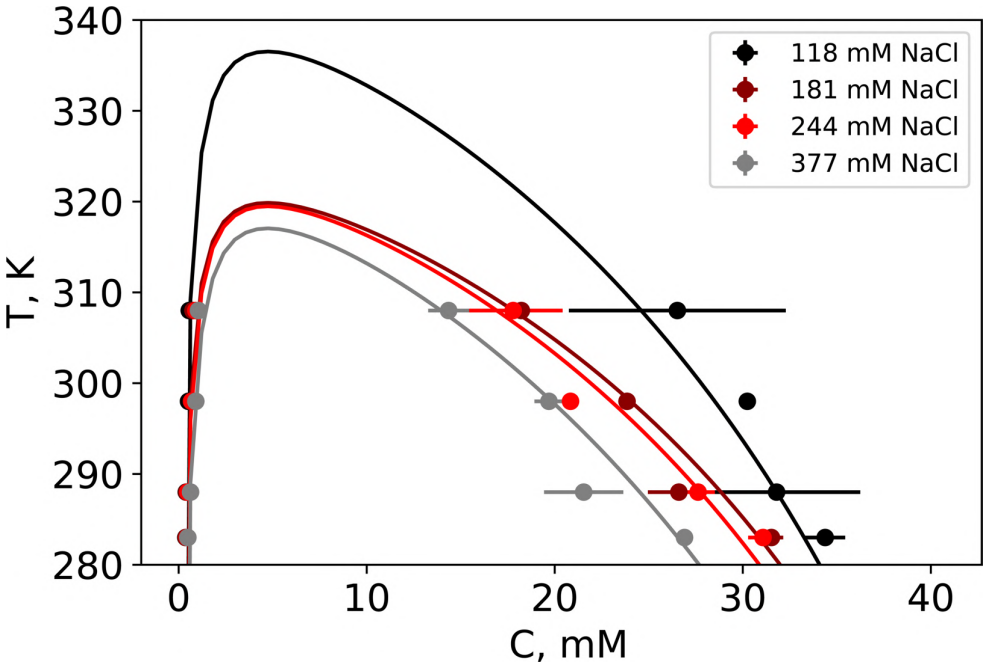
Figure S3B



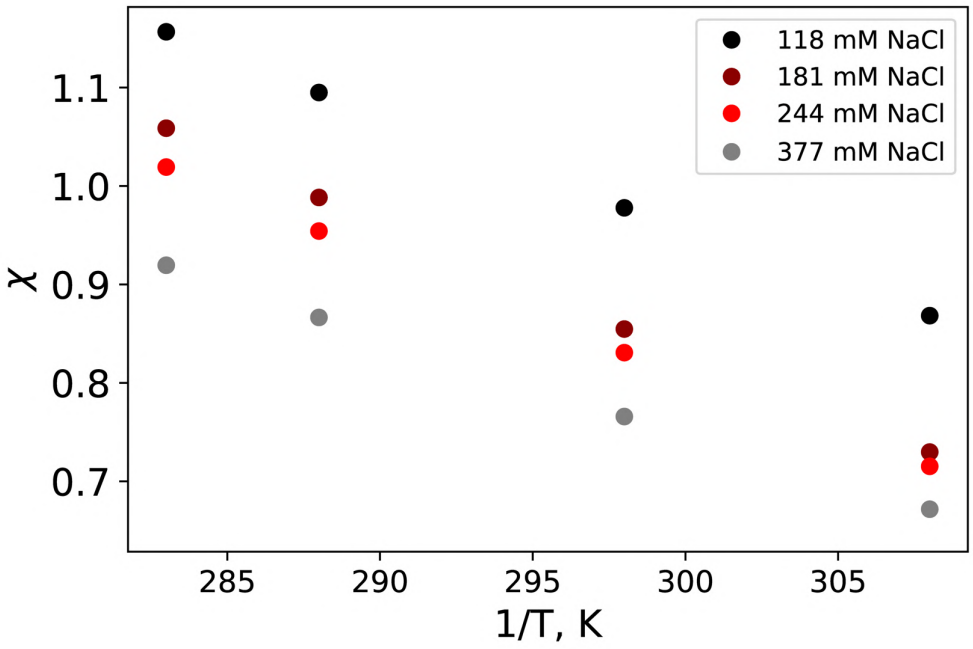
Example of Fluorescence recovery after photobleaching (FRAP) carried out on a single N_{TAIL} droplet (see Methods).

Figure S4

S4A



S4B

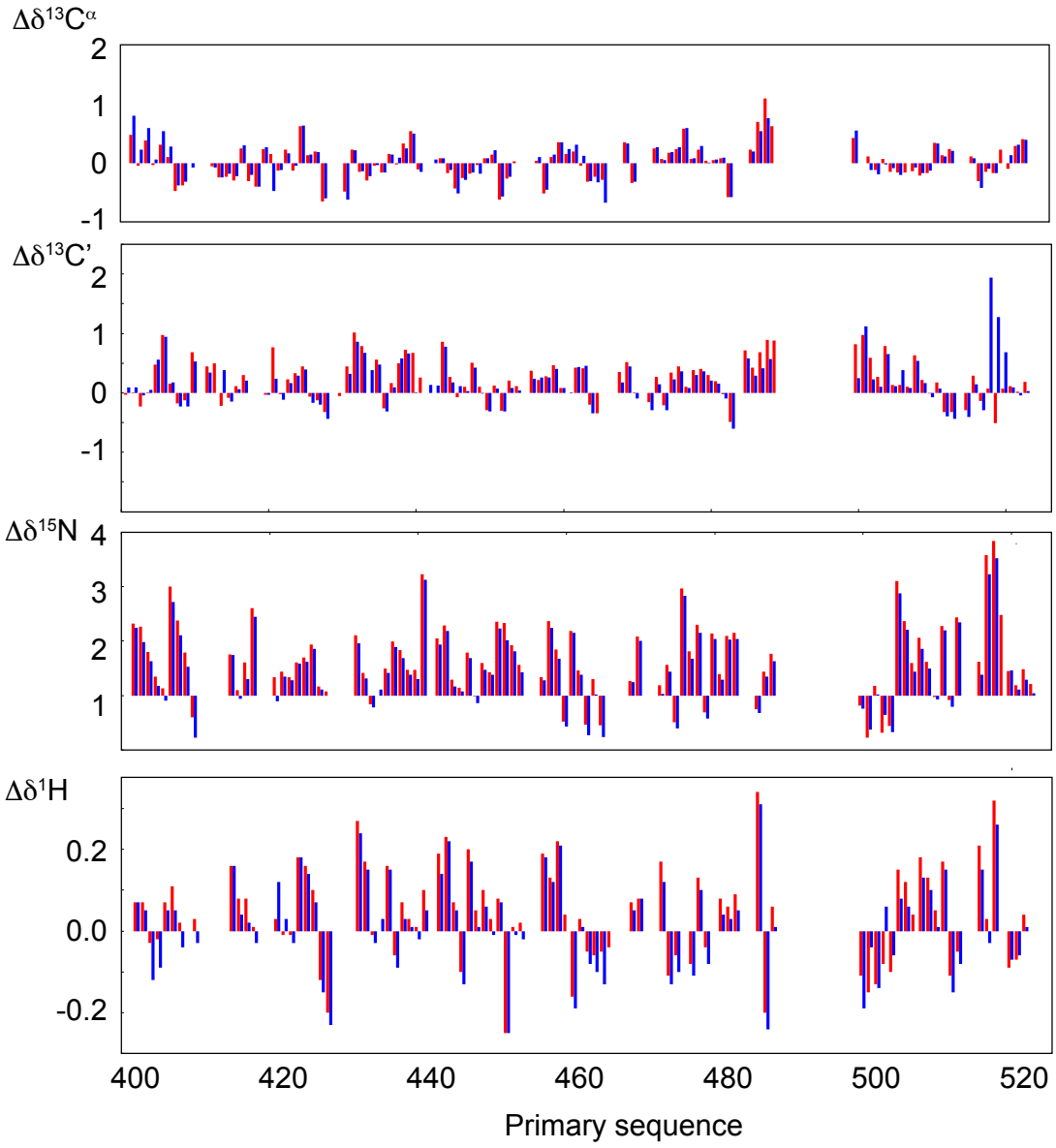


Flory-Huggins model fit of the experimentally determined phase diagram

S4A - Phase diagram calculated from fitting to equation 1.

S4B - Calculated χ parameter for all temperature and salt conditions.

Figure S5

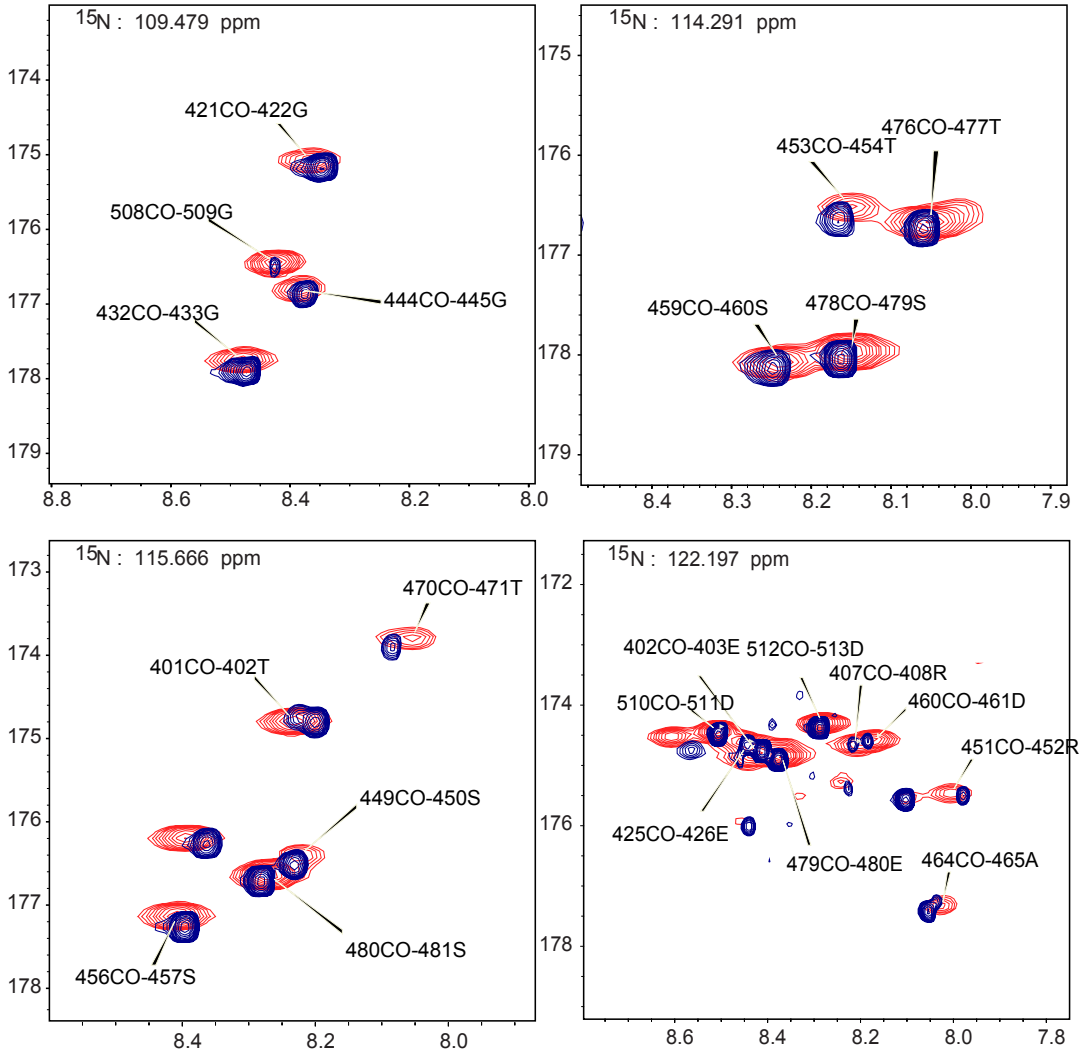


Secondary chemical shifts of N_{TAIL} suggest conserved backbone conformational sampling throughout phase space.

Top to bottom: ^{13}Ca, ^{13}C', ^{15}N and ^{1}H secondary chemical shifts.

Red - N_{TAIL} in the dilute state, blue - N_{TAIL} in the dense phase.

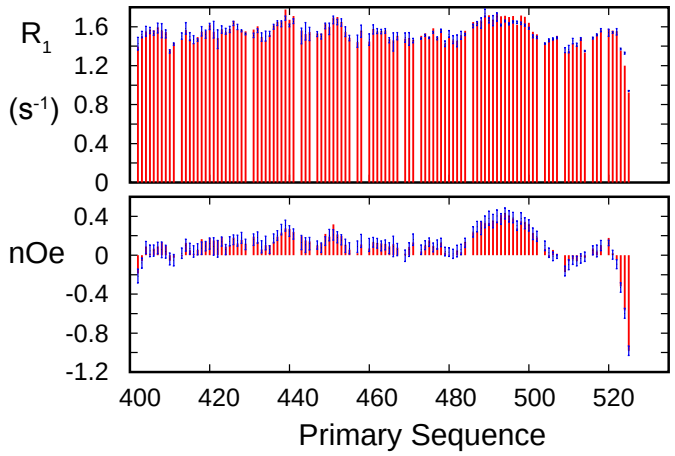
Figure S6



Backbone chemical shifts of N_{TAIL} suggest conserved backbone conformational sampling throughout phase space.

Comparison of ^{15}N planes from triple resonance HNCO experiments on dilute (blue) and dense (red) phase samples showing similarity of ^{N}H-C' correlations under the two conditions.

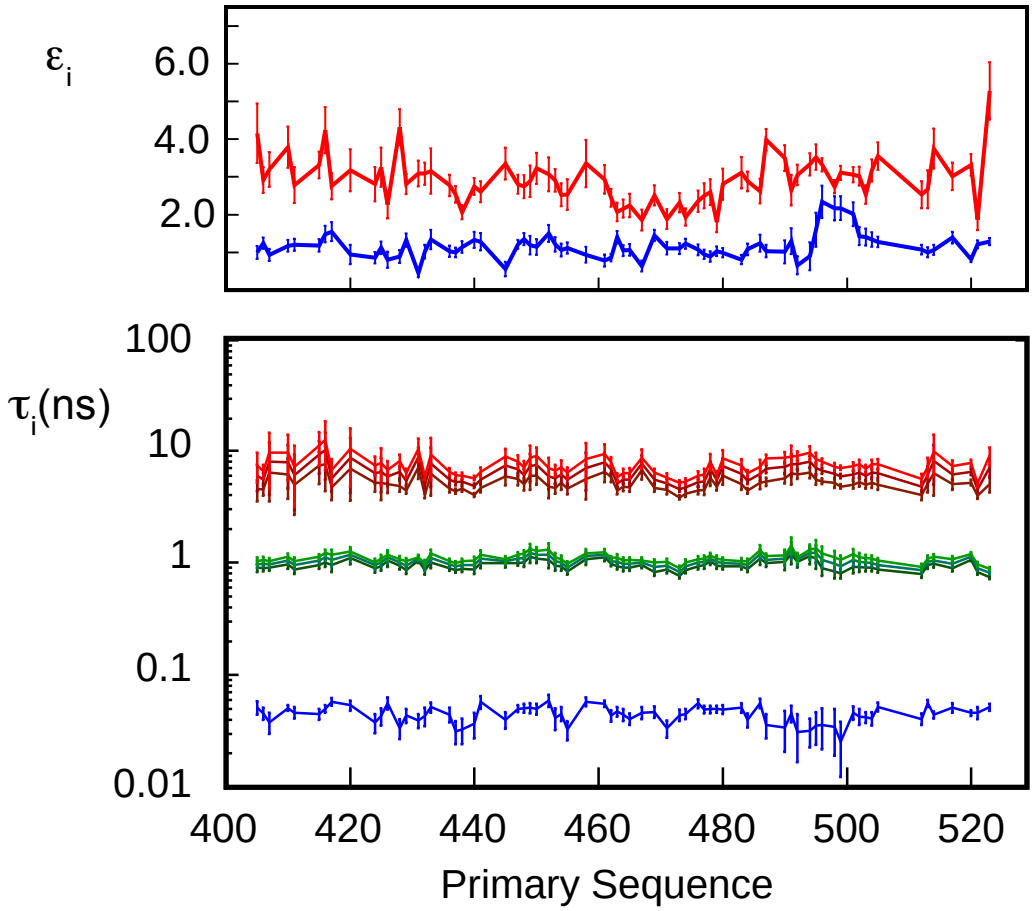
Figure S7



Cross-validation of dynamic model of N_{TAIL} in the dilute phase

Experimental R_1 and heteronuclear ^{1}H-^{15}N nOe measured at 700MHz on the 0g/L (PEG concentration) N_{TAIL} sample in the dilute phase (blue), compared to values back-calculated from the dynamic model-free analysis (red bars) of data measured at 600 and 850 MHz at 0, 37.5 and 75g/L PEG10000). Rmsd values are within the mean experimental error for both R_1 (0.033 compared to 0.0319) and heteronuclear ^{1}H-^{15}N nOe (0.035 compared to 0.042).

Figure S8

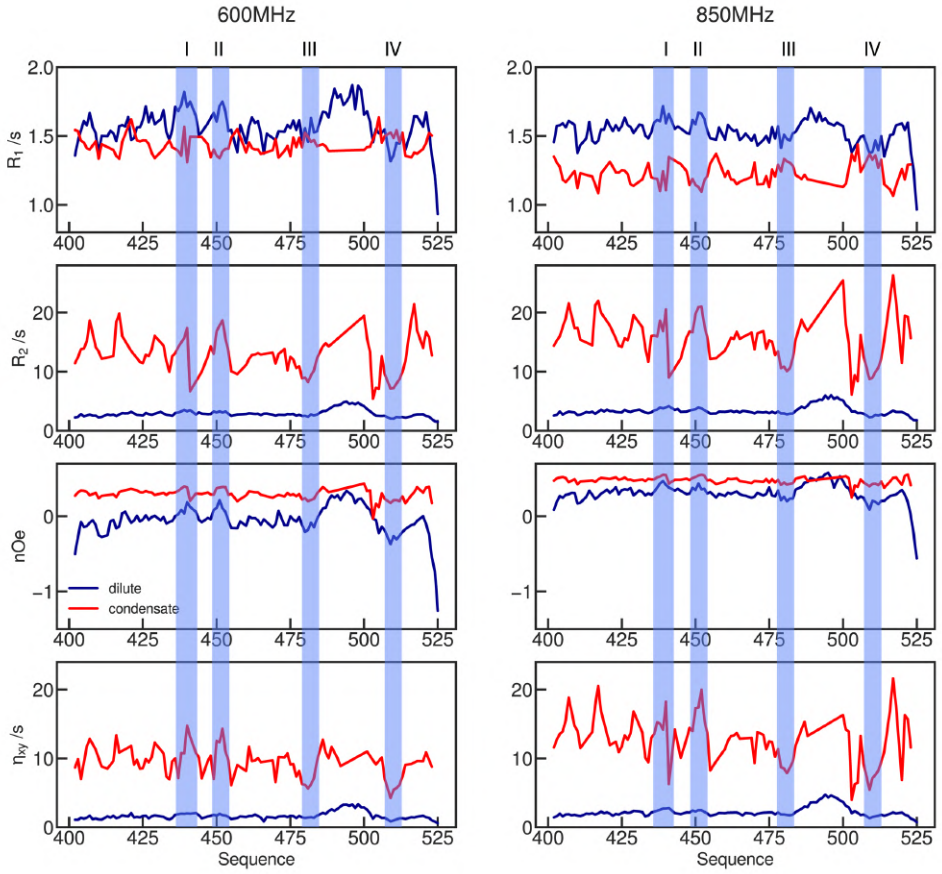


Model-free analysis of dynamic behaviour of N_{TAIL} as a function of viscosity

Top – Residue-specific viscosity coefficients for the intermediate (blue) and slow (red) dynamics modes.

Bottom – Residue-specific correlation times of segmental backbone motions (τ_3)- 0 g/L (dark red), 37.5 g/L (red) and 75 g/L PEG10000 (light red), intermediate, backbone motions (τ_2) - 0 g/L (dark green), 37.5 g/L (green) and 75 g/L PEG10000 (light green) and fast motions (τ_1) (blue)

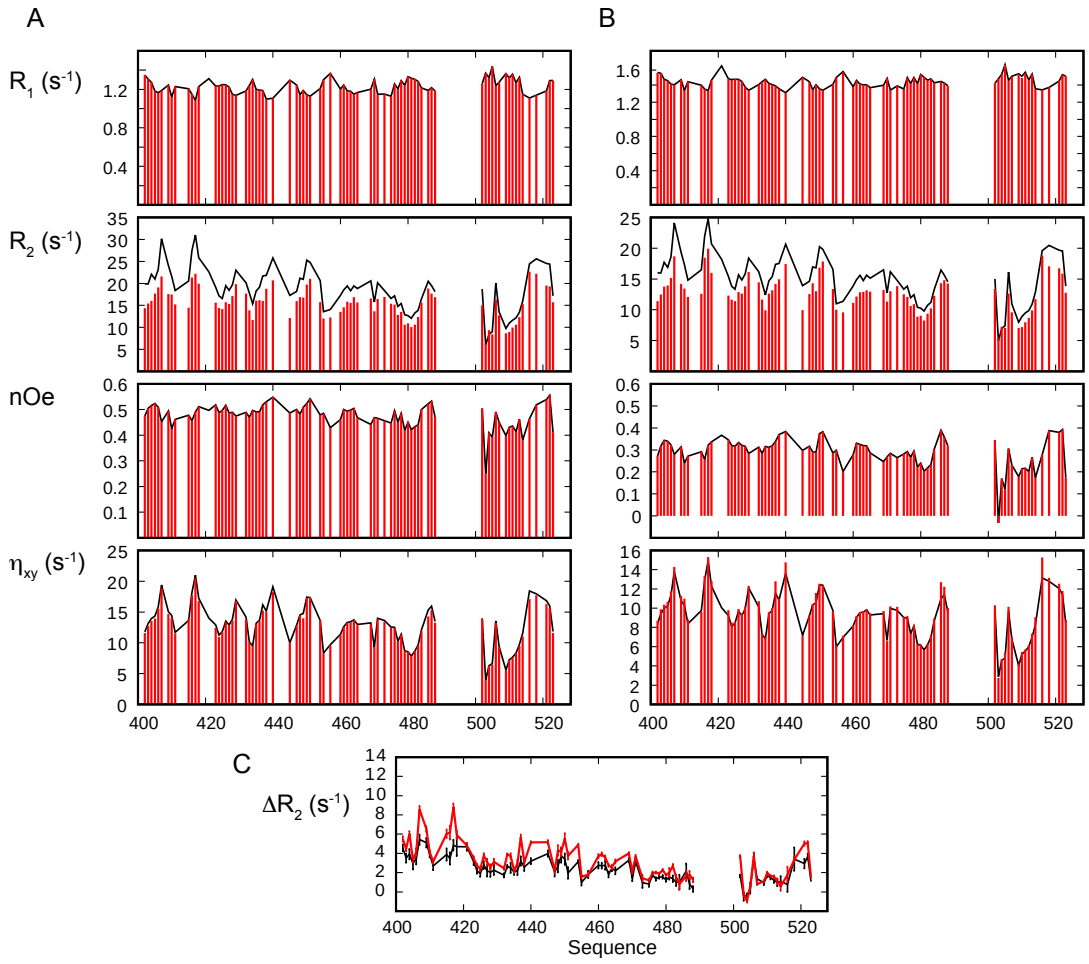
Figure S9



Identification of regions of N_{TAIL} exhibiting specific dynamic behaviour

Figure 4 of the manuscript is reproduced highlighting regions ^{438}RRVK^{441} (I) and ^{449}ESYRE^{453} (II), both exhibiting elevated transverse relaxation rates in both dilute and dense phases, and ^{487}TASESS^{492} (III) and ^{509}GSDT^{512} (IV) that clearly exhibit higher flexibility

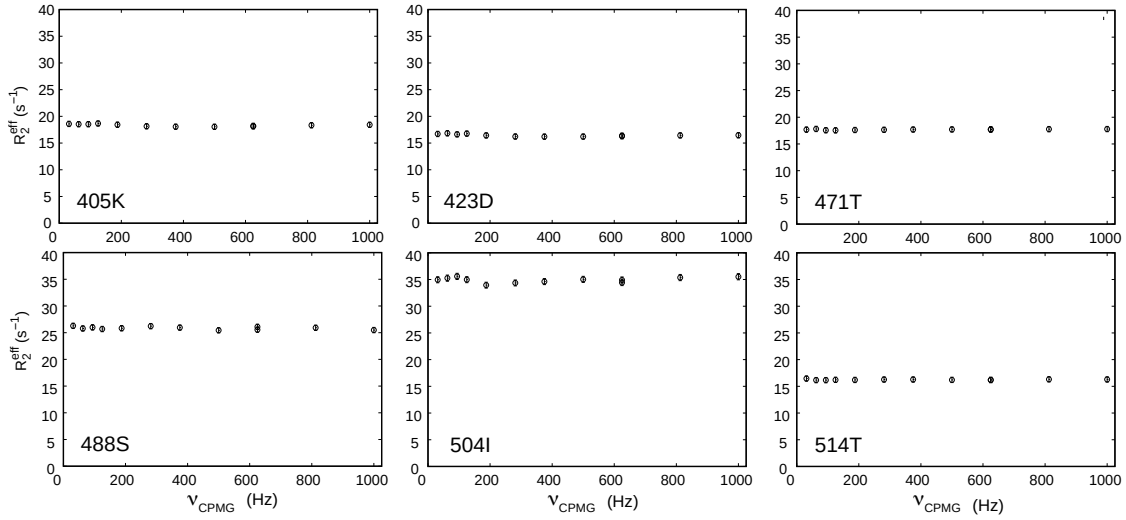
Figure S10



Reproduction of experimental data by model-free analysis of data from the dense phase

88 experimental relaxation rates (black lines) were fitted to equation 7. 5 parameters were optimized ($\tau_1, \tau_2, \tau_3, A_2$ and A_3) by fitting to experimental (black) R_1, η_{xy} and heteronuclear $\{^1H\}$-^{15}N nOe at two magnetic field strengths (A-850 and B-600 MHz). Calculated values are shown as red bars. C - R_2 values were back-calculated from this model-free analysis, and compared to experimental values ($\Delta R_2 = R_{2,calc} - R_{2,exp}$) (red 850MHz, black – 600 MHz).

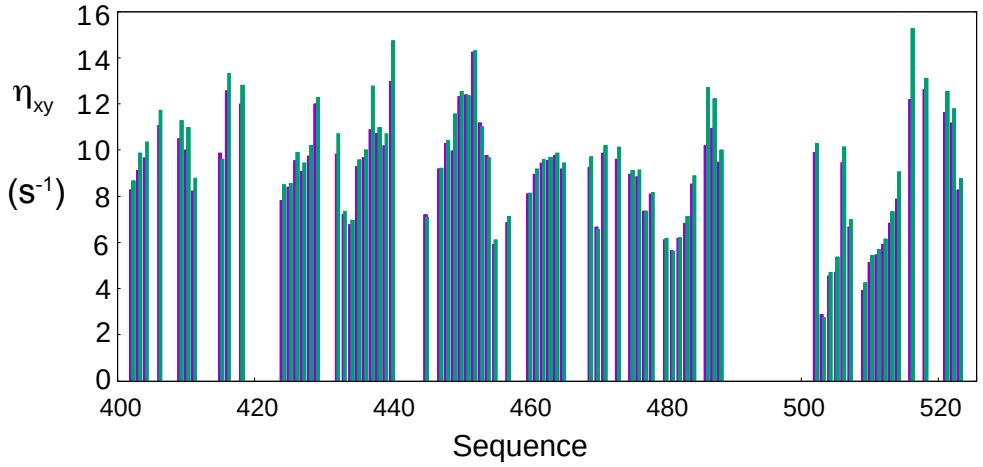
Figure S11



Relaxation dispersion CPMG experiments carried on the N_{TAIL} in the dense phase

Relaxation dispersion CPMG was carried out at 950 MHz in the dense phase, revealing no evidence for significant chemical shift exchange, as illustrated from 6 randomly selected amino acids along the primary sequence.

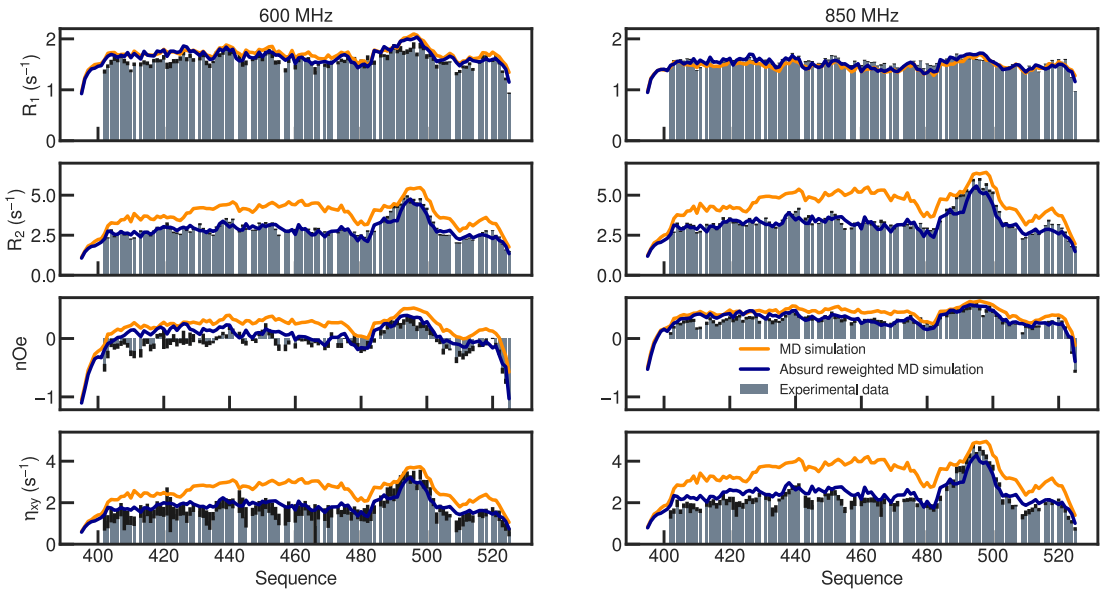
Figure S12



Reproduction of experimental data by model-free analysis of data from the dense phase

Experimental η_{xy} rates (blue bars) that were removed from the fit equation 7 and predicted from the fit of R_1, and heteronuclear $\{^1H\}$-^{15}N nOe at 600 MHz and R_1, η_{xy} and heteronuclear $\{^1H\}$-^{15}N nOe at 850MHz. Calculated values are shown as green bars.

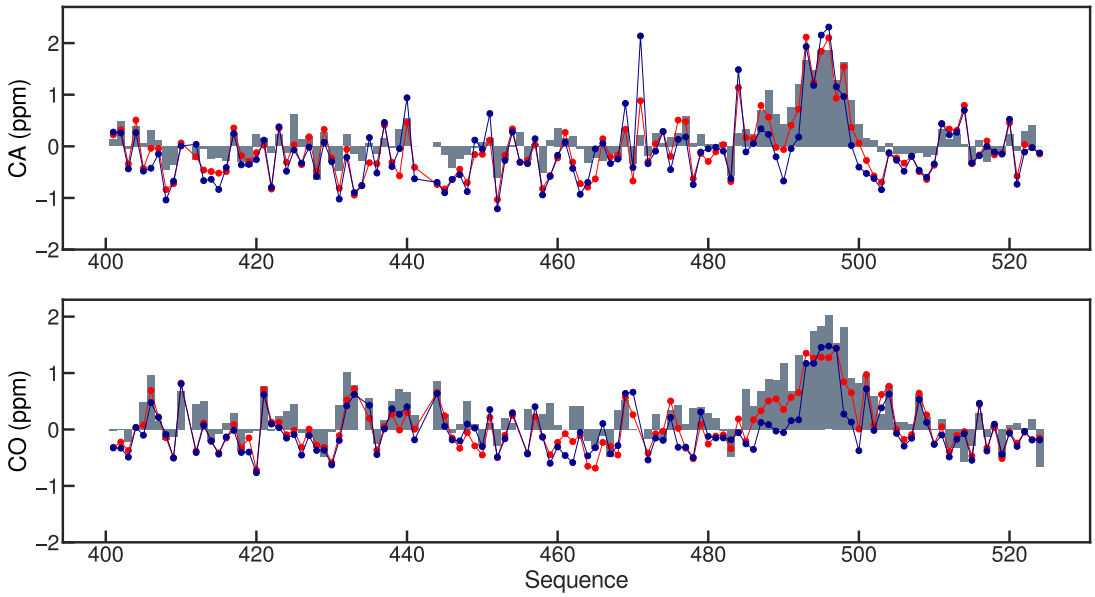
Figure S13



Reproduction of experimental relaxation rates using ABSURD ensemble trajectory analysis

MD simulation of were performed with CHARMM36m (C36m) in combination with the TIP4P/2005 (t4p2005) water model. 30 trajectories of 200ns were calculated and analyzed. Predicted rates (orange) failed to accurately reproduce experimental rates (bars), especially those sensitive to J(0). The ABSURD genetic algorithm targetting R_2 at 850MHz selects the combination of trajectories that best reproduce this rate. Improvements in the agreement with all other experimental data is also observed (blue lines), indicating a better representation of the dynamic ensemble.

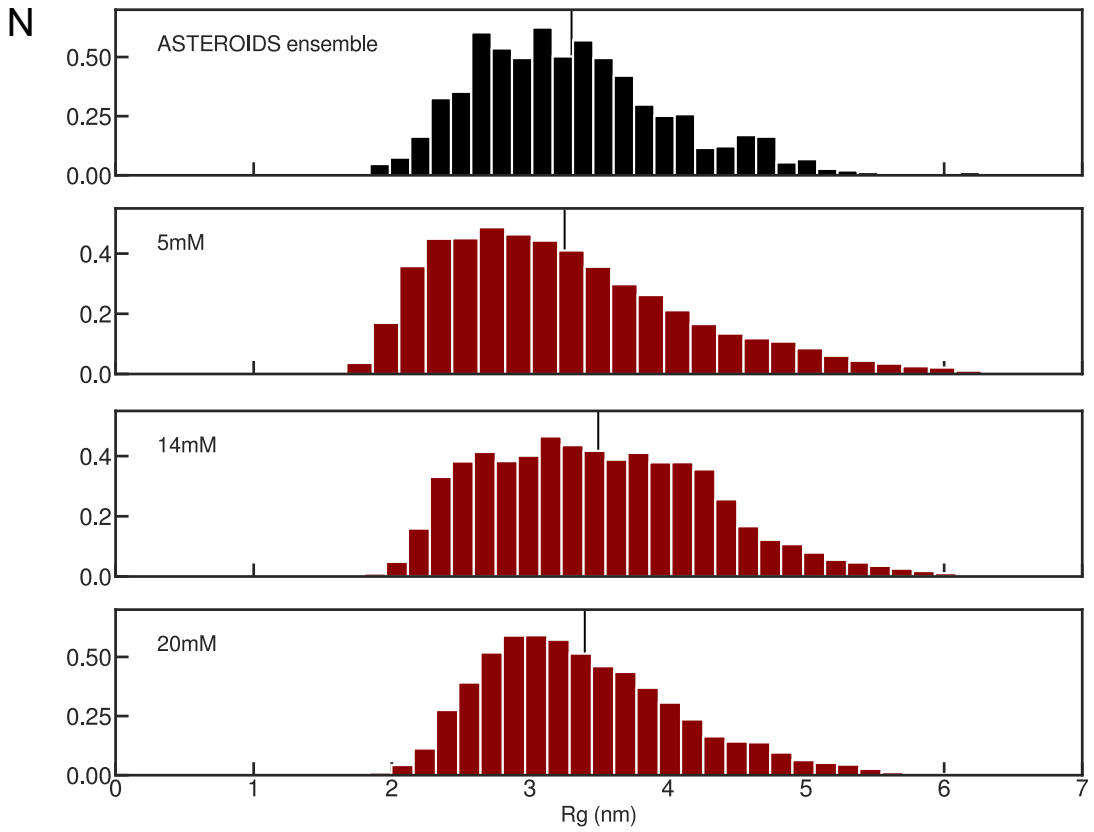
Figure S14



Reproduction of experimental chemical shifts using ABSURD ensemble trajectory analysis and MD simulations of highly concentrated N_{TAIL}

Chemical shifts prediction derived from MD simulations of dilute (ABSURD ensemble of trajectories - blue), 20mM N_{TAIL} concentration (red) compared with experimental data from the dilute phase (grey bars).

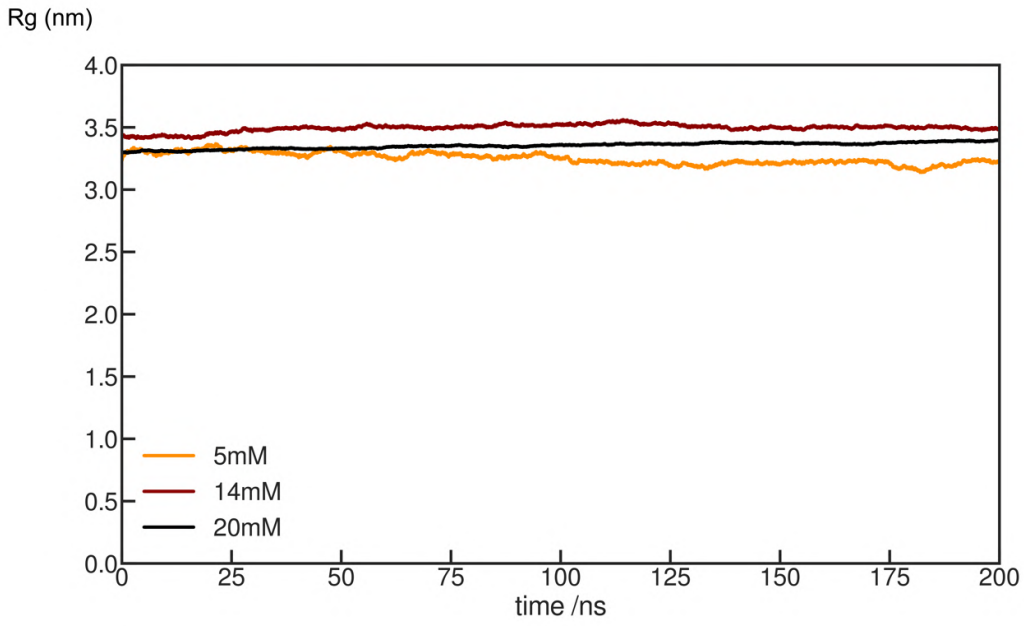
Figure S15



Distribution of radii of gyration of the ASTEROIDS ensemble and MD simulations of highly concentrated N_{TAIL}

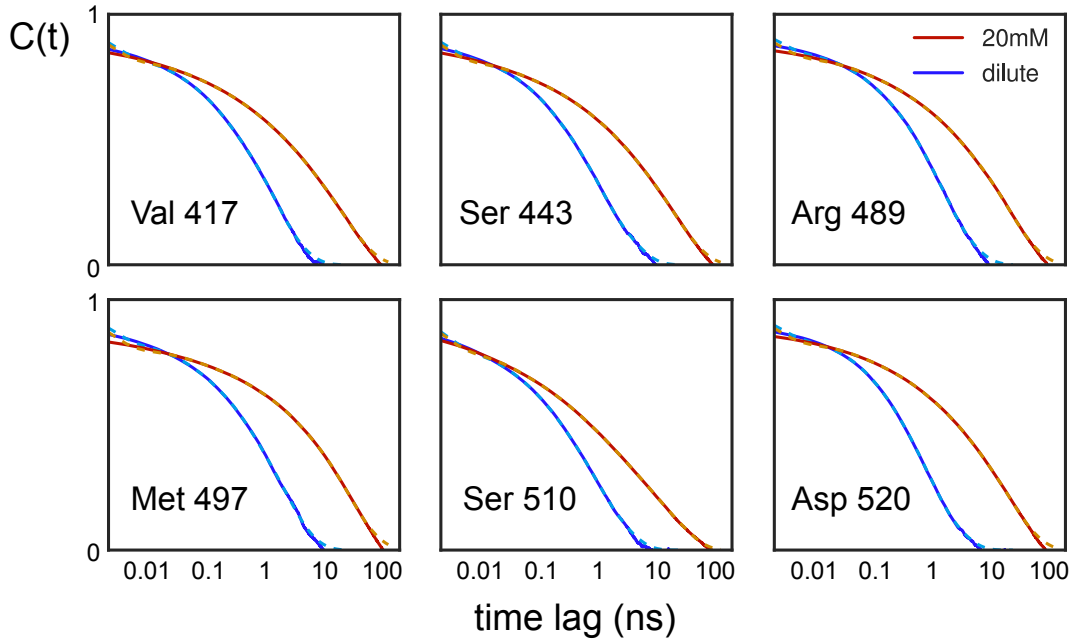
Radii of gyration were averaged over 100 conformers in the ASTEROIDS ensemble, and over 200ns and 125, 125 and 343 copies of the protein for the self-crowding simulations (5, 14 and 20mM respectively).

Figure S16



Evolution of average radii of gyration of the MD simulations of highly concentrated N_{TAIL}
Radii of gyration were averaged over 125, 125 and 343 copies of the protein for the self-crowding simulations (5, 14 and 20mM respectively).

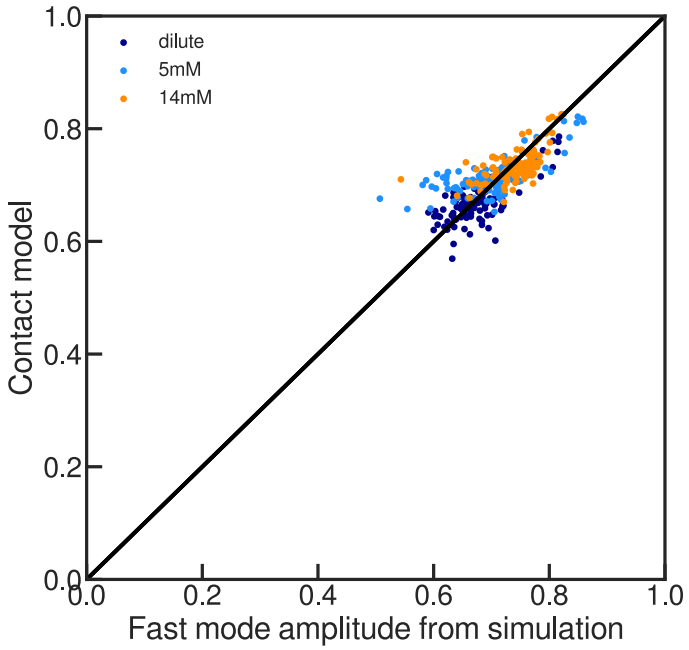
Figure S17



Comparison of autocorrelation functions calculated from dilute and concentrated phase simulations

Solid lines show correlation functions calculated from the ABSURD-derived ensemble of trajectories describing the dilute phase (blue) and the 20mM N_{TAIL} box (red), in this case the correlation function is averaged over 343 copies. Orange and light blue dashed lines show the fit to the correlation functions used to calculate the relaxation rates. Correlation functions (417, 443, 489, 497, 510 and 520) were randomly selected along the primary sequence.

Figure S18



Correlation of amplitudes of motions calculated from the contact model (see methods) and derived from model free analysis of simulated relaxation rates.

Bibliography

(1) Tanford, C.; Reynolds, J., *Nature's Robots: A History of Proteins*, 1st edition; Oxford University Press: New York, NY Oxford, 2004; 320 pp.

(2) Mulder, G. J. *Annalen der Pharmacie* **1838**, *28*, 73–82.

(3) Hartley, H. *Nature* **1951**, *168*, 244–244.

(4) Vickery, H. B. *The Yale Journal of Biology and Medicine* **1932**, *4*, 595–610.

(5) Fischer, E. *Zeitschrift für Untersuchung der Nahrungs- und Genußmittel, sowie der Gebrauchsgegenstände* **1902**, *5*, 1207–1208.

(6) Sanger, F. *Biochemical Journal* **1949**, *45*, 563–574.

(7) Pauling, L.; Corey, R. B.; Branson, H. R. *Proceedings of the National Academy of Sciences* **1951**, *37*, 205–211.

(8) Kauzmann, W. *Journal of Cellular and Comparative Physiology* **1956**, *47*, 113–131.

(9) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.

(10) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *Journal of Molecular Biology* **2004**, *337*, 635–645.

(11) Wright, P. E.; Dyson, H. J. *Journal of Molecular Biology* **1999**, *293*, 321–331.

(12) Dunker, A. K. et al. *Journal of Molecular Graphics and Modelling* **2001**, *19*, 26–59.

(13) Tompa, P. *Trends in Biochemical Sciences* **2002**, *27*, 527–533.

(14) Tompa, P.; Schad, E.; Tantos, A.; Kalmar, L. *Current Opinion in Structural Biology* **2015**, *35*, 49–59.

(15) Wright, P. E.; Dyson, H. J. *Nature Reviews Molecular Cell Biology* **2015**, *16*, 18–29.

(16) Pelham, J. F.; Dunlap, J. C.; Hurley, J. M. *Cell Communication and Signaling* **2020**, *18*, 181.

(17) Boothby, T. C.; Tapia, H.; Brozena, A. H.; Piszkiewicz, S.; Smith, A. E.; Giovannini, I.; Rebecchi, L.; Pielak, G. J.; Koshland, D.; Goldstein, B. *Molecular Cell* **2017**, *65*, 975–984.e5.

(18) Huang, A.; Stultz, C. M. *Future Medicinal Chemistry* **2009**, *1*, 467–482.

(19) Spillantini, M. G.; Crowther, R. A.; Jakes, R.; Hasegawa, M.; Goedert, M. *Proceedings of the National Academy of Sciences* **1998**, *95*, 6469–6473.

(20) Hubin, E.; van Nuland, N. A. J.; Broersen, K.; Pauwels, K. *Cellular and Molecular Life Sciences* **2014**, *71*, 3507–3521.

(21) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. *Current Opinion in Structural Biology* **2011**, *21*, 432–440.

(22) Milles, S.; Jensen, M. R.; Lazert, C.; Guseva, S.; Ivashchenko, S.; Communie, G.; Maurin, D.; Gerlier, D.; Ruigrok, R. W. H.; Blackledge, M. *Science Advances* **2018**, *4*, eaat7778.

(23) Guseva, S.; Milles, S.; Blackledge, M.; Ruigrok, R. W. H. *Frontiers in Microbiology* **2019**, *10*, 1832.

(24) Camacho-Zarco, A. R.; Kalayil, S.; Maurin, D.; Salvi, N.; Delaforge, E.; Milles, S.; Jensen, M. R.; Hart, D. J.; Cusack, S.; Blackledge, M. *Nature Communications* **2020**, *11*, 3656.

(25) Bessa, L. M.; Guseva, S.; Camacho-Zarco, A. R.; Salvi, N.; Maurin, D.; Perez, L. M.; Botova, M.; Malki, A.; Nanao, M.; Jensen, M. R.; Ruigrok, R. W. H.; Blackledge, M. *Science Advances* **2022**, *8*, eabm4034.

(26) Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. *Physical Review* **1938**, *53*, 318–318.

(27) Purcell, E. M.; Torrey, H. C.; Pound, R. V. *Physical Review* **1946**, *69*, 37–38.

(28) Bloch, F.; Hansen, W. W.; Packard, M. *Physical Review* **1946**, *70*, 474–485.

(29) Abragam, A., *The Principles of Nuclear Magnetism*; Clarendon Press: 1961; 664 pp.

(30) Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1990; 640 pp.

(31) Cavanagh, J.; Fairbrother, W. J.; III, A. G. P.; Skelton, N. J., *Protein NMR Spectroscopy: Principles and Practice*; Elsevier: 1995; 915 pp.

(32) Levitt, M. H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd Edition; Wiley–Blackwell: Chichester, England ; Hoboken, NJ, 2008; 740 pp.

(33) Redfield, A. G. *Advances in Magnetic Resonance* **1965**, *1*, 1–32.

(34) Wangsness, R. K.; Bloch, F. *Physical Review* **1953**, *89*, 728–739.

(35) Hubbard, P. S. *Physical Review* **1969**, *180*, 319–326.

(36) Bolik-Coulon, N.; Kadeřávek, P.; Pelupessy, P.; Dumez, J.-N.; Ferrage, F.; Cousin, S. F. *Journal of Magnetic Resonance* **2020**, *313*, 106718.

(37) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. *Journal of Magnetic Resonance* **2003**, *161*, 258–264.

(38) Pelupessy, P.; Ferrage, F.; Bodenhausen, G. *The Journal of Chemical Physics* **2007**, *126*, 134508.

(39) Woessner, D. E. *The Journal of Chemical Physics* **1962**, *36*, 1–4.

(40) Wittebort, R. J.; Szabo, A. *Journal of Chemical Physics* **1978**, *69*, 1722–1736.

(41) King, R.; Jardetzky, O. *Chemical Physics Letters* **1978**, *55*, 15–18.

(42) Lipari, G.; Szabo, A. *Biophysical Journal* **1980**, *30*, 489–506.

(43) Richarz, R.; Nagayama, K.; Wuethrich, K. *Biochemistry* **1980**, *19*, 5189–5196.

(44) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.; (United States)* **1982**, *104:17*, DOI: 10.1021/ja00381a009.

(45) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *Journal of the American Chemical Society* **1990**, *112*, 4989–4991.

(46) Kay, L. E.; Torchia, D. A *Journal of Magnetic Resonance (1969)* **1991**, *95*, 536–547.

(47) Tjandra, N.; Szabo, A.; Bax, A. *Journal of the American Chemical Society* **1996**, *118*, 6986–6991.

(48) Daragan, V. A.; Mayo, K. H. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997**, *31*, 63–105.

(49) Korzhnev, D. M.; Bocharov, E. V.; Zhuravlyova, A. V.; Orekhov, V. Y.; Ovchinnikova, T. V.; Billeter, M.; Arseniev, A. S. *FEBS Letters* **2001**, *495*, 52–55.

(50) Gill, M. L.; Byrd, R. A.; Arthur G. Palmer, I. I. I. *Physical Chemistry Chemical Physics* **2016**, *18*, 5839–5849.

(51) Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R. W.; Jensen, M. R.; Blackledge, M. *Journal of the American Chemical Society* **2016**, *138*, 6240–6251.

(52) Adamski, W.; Salvi, N.; Maurin, D.; Magnat, J.; Milles, S.; Jensen, M. R.; Abyzov, A.; Moreau, C. J.; Blackledge, M. *Journal of the American Chemical Society* **2019**, *141*, 17817–17829.

(53) Guseva, S.; Schnapka, V.; Adamski, W.; Maurin, D.; Ruigrok, R. W. H.; Salvi, N.; Blackledge, M. *Journal of the American Chemical Society* **2023**, *145*, 10548–10563.

(54) Lankhorst, D.; Schriever, J.; Leyte, J. C. *Berichte der Bunsengesellschaft für physikalische Chemie* **1982**, *86*, 215–221.

(55) Tsan, P.; Hus, J.-C.; Caffrey, M.; Marion, D.; Blackledge, M. *Journal of the American Chemical Society* **2000**, *122*, 5603–5612.

(56) Halle, B.; Wennerström, H. *The Journal of Chemical Physics* **1981**, *75*, 1928–1943.

(57) Halle, B. *The Journal of Chemical Physics* **2009**, *131*, 224507.

(58) Brink, D. M.; Satchler, G. R.; Brink, D. M.; Satchler, G. R., *Angular Momentum*, Third Edition, Third Edition; Oxford University Press: Oxford, New York, 1994; 182 pp.

(59) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry* **1992**, *31*, 5269–5278.

(60) Mandel, A. M.; Akke, M.; Palmer Arthur G., I. *Journal of Molecular Biology* **1995**, *246*, 144–163.

(61) Mandel, A. M.; Akke, M.; Palmer, A. G. *Biochemistry* **1996**, *35*, 16009–16023.

(62) Malmendal, A.; Evenäs, J.; Forsén, S.; Akke, M. *Journal of Molecular Biology* **1999**, *293*, 883–899.

(63) Cordier, F.; Grzesiek, S. *Journal of Molecular Biology* **2002**, *317*, 739–752.

(64) Diehl, C.; Engström, O.; Delaine, T.; Håkansson, M.; Genheden, S.; Modig, K.; Leffler, H.; Ryde, U.; Nilsson, U. J.; Akke, M. *Journal of the American Chemical Society* **2010**, *132*, 14577–14589.

(65) Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L. *Science* **2015**, *348*, 578–581.

(66) Peng, J. W.; Wagner, G. *Biochemistry* **1992**, *31*, 8571–8586.

(67) Peng, J. W.; Wagner, G. *Biochemistry* **1995**, *34*, 16733–16752.

(68) Ishima, R.; Nagayama, K. *Biochemistry* **1995**, *34*, 3162–3171.

(69) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. *Journal of Biomolecular NMR* **1995**, *6*, 153–162.

(70) Kadeřávek, P.; Zapletal, V.; Rabatinová, A.; Krásný, L.; Sklenář, V.; Žídek, L. *Journal of Biomolecular NMR* **2014**, *58*, 193–207.

(71) Khan, S. N.; Charlier, C.; Augustyniak, R.; Salvi, N.; Déjean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F. *Biophysical Journal* **2015**, *109*, 988–999.

(72) Hsu, A.; Ferrage, F.; Palmer, A. G. *Biophysical journal* **2018**, *115*, 2301–2309.

(73) Dais, P.; Spyros, A. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1995**, *27*, 555–633.

(74) Beckmann, P. A. *Physics Reports* **1988**, *171*, 85–128.

(75) Redfield, A. G. *Magnetic Resonance in Chemistry* **2003**, *41*, 753–768.

(76) Cousin, S. F.; Kadeřávek, P.; Bolik-Coulon, N.; Gu, Y.; Charlier, C.; Carlier, L.; Bruschweiler-Li, L.; Marquardsen, T.; Tyburn, J.-M.; Brüschweiler, R.; Ferrage, F. *Journal of the American Chemical Society* **2018**, *140*, 13456–13465.

(77) Bolik-Coulon, N.; Languin-Cattoën, O.; Carnevale, D.; Zachrdla, M.; Laage, D.; Sterpone, F.; Stirnemann, G.; Ferrage, F. *Physical Review Letters* **2022**, *129*, 203001.

(78) McConnell, H. M. *The Journal of Chemical Physics* **1958**, *28*, 430–431.

(79) Abergel, D.; Palmer III, A. G. *Concepts in Magnetic Resonance Part A* **2003**, *19A*, 134–148.

(80) Abergel, D.; Palmer, A. G. *The Journal of Physical Chemistry B* **2005**, *109*, 4837–4844.

(81) Nageswara Rao, B. D. In *Methods in Enzymology*; Nuclear Magnetic Resonance Part A: Spectral Techniques and Dynamics, Vol. 176; Academic Press: 1989, pp 279–311.

(82) Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J. *Scientific Reports* **2016**, *6*, 1–8.

(83) Carr, H. Y.; Purcell, E. M. *Physical Review* **1954**, *94*, 630–638.

(84) Meiboom, S.; Gill, D. *Review of Scientific Instruments* **1958**, *29*, 688–691.

(85) Jen, J. *Journal of Magnetic Resonance (1969)* **1978**, *30*, 111–128.

(86) Loria, J. P.; Rance, M.; Palmer, A. G. *Journal of Biomolecular NMR* **1999**, *15*, 151–155.

(87) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *The Journal of Physical Chemistry B* **2008**, *112*, 5898–5904.

(88) Palmer, A. G.; Grey, M. J.; Wang, C. In *Methods in Enzymology*; Nuclear Magnetic Resonance of Biological Macromolecules, Vol. 394; Academic Press: 2005, pp 430–465.

(89) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *The Journal of Physical Chemistry B* **2011**, *115*, 14891–14900.

(90) Forsén, S.; Hoffman, R. A. *The Journal of Chemical Physics* **1963**, *39*, 2892–2901.

(91) Gupta, R. K.; Redfield, A. G. *Science* **1970**, *169*, 1204–1206.

(92) Cayley, P. J.; Albrand, J. P.; Feeney, J.; Roberts, G. C.; Piper, E. A.; Burgen, A. S. *Biochemistry* **1979**, *18*, 3886–3895.

(93) Alger, J. R.; Shulman, R. G. *Quarterly Reviews of Biophysics* **1984**, *17*, 83–124.

(94) Ward, K. M; Aletras, A. H; Balaban, R. S *Journal of Magnetic Resonance* **2000**, *143*, 79–87.

(95) Zhou, J.; Zijl, P. C. M. v. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48*, 109–136.

(96) Van Zijl, P. C. M.; Yadav, N. N. *Magnetic Resonance in Medicine* **2011**, *65*, 927–948.

(97) Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M. *Nature* **2011**, *480*, 268–272.

(98) Yuwen, T.; Kay, L. E.; Bouvignies, G. *ChemPhysChem* **2018**, *19*, 1707–1710.

(99) Bodenhausen, G.; Freeman, R.; Morris, G. A. *Journal of Magnetic Resonance* **1976**, *23*, 171–175.

(100) Palmer, A. G.; Massi, F. *Chemical Reviews* **2006**, *106*, 1700–1719.

(101) Rangadurai, A.; Szymaski, E. S.; Kimsey, I. J.; Shi, H.; Al-Hashimi, H. M. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2019**, *112-113*, 55–102.

(102) Montelione, G. T.; Wagner, G. *Journal of the American Chemical Society* **1989**, *111*, 3096–3098.

(103) Wider, G.; Neri, D.; Wüthrich, K. *Journal of Biomolecular NMR* **1991**, *1*, 93–98.

(104) Meier, B. H.; Ernst, R. R. *Journal of the American Chemical Society* **1979**, *101*, 6441–6442.

(105) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *The Journal of Chemical Physics* **1979**, *71*, 4546–4553.

(106) Wishart, D. S.; Sykes, B. D. *Journal of Biomolecular NMR* **1994**, *4*, 171–180.

(107) Oldfield, E. *Journal of Biomolecular NMR* **1995**, *5*, 217–225.

(108) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Journal of Molecular Biology* **1991**, *222*, 311–333.

(109) Wishart, D. S.; Sykes, B. D. *Journal of biomolecular NMR* **1994**, *4*, 171–180.

(110) Kragelj, J.; Ozenne, V.; Blackledge, M.; Jensen, M. R. *ChemPhysChem* **2013**, *14*, 3034–3045.

(111) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Science* **2006**, *15*, 2795–2804.

(112) Tamiola, K.; Acar, B.; Mulder, F. A. A. *Journal of the American Chemical Society* **2010**, *132*, 18000–18003.

(113) Tamiola, K.; Mulder, F. *Biochemical Society Transactions* **2012**, *40*, 1014–1020.

(114) Emsley, J.; Lindon, J. *Molecular Physics* **1975**, *29*, 531–538.

(115) Blackledge, M. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005**, *46*, 23–61.

(116) Salmon, L.; Blackledge, M. *Reports on Progress in Physics* **2015**, *78*, 126601.

(117) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proceedings of the National Academy of Sciences* **2005**, *102*, 13099–13104.

(118) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proceedings of the National Academy of Sciences* **2005**, *102*, 17002–17007.

(119) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M. *Structure* **2009**, *9*, 1169–1185.

(120) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *Journal of the American Chemical Society* **2009**, *131*, 17908–17918.

(121) Bouvignies, G.; Bernadó, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proceedings of the National Academy of Sciences* **2005**, *102*, 13885–13890.

(122) Tolman, J. R.; Ruan, K. *Chemical Reviews* **2006**, *106*, 1720–1736.

(123) Bouvignies, G.; Markwick, P. R. L.; Blackledge, M. *ChemPhysChem* **2007**, *8*, 1901–1909.

(124) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *Journal of the American Chemical Society* **2009**, *131*, 16968–16975.

(125) Clore, G. M.; Iwahara, J. *Chemical Reviews* **2009**, *109*, 4108–4139.

(126) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *Journal of the American Chemical Society* **2010**, *132*, 8407–8418.

(127) Karplus, M. *The Journal of Chemical Physics* **1959**, *30*, 11–15.

(128) Pardi, A.; Billeter, M.; Wüthrich, K. *Journal of Molecular Biology* **1984**, *180*, 741–751.

(129) Vuister, G. W.; Bax, A. *Journal of the American Chemical Society* **1993**, *115*, 7772–7777.

(130) Serrano, L. *Journal of Molecular Biology* **1995**, *254*, 322–333.

(131) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *Journal of Molecular Biology* **1996**, *255*, 494–506.

(132) Penkett, C. J.; Redfield, C.; Jones, J. A.; Dodd, I.; Hubbard, J.; Smith, R. A. G.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1998**, *37*, 17054–17067.

(133) Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. *The FEBS Journal* **2005**, *272*, 3938–3949.

(134) Tang, Y.; Goger, M. J.; Raleigh, D. P. *Biochemistry* **2006**, *45*, 6940–6946.

(135) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. *Journal of the American Chemical Society* **2010**, *132*, 540–551.

(136) Oh, K.-I.; Jung, Y.-S.; Hwang, G.-S.; Cho, M. *Journal of Biomolecular NMR* **2012**, *53*, 25–41.

(137) Xiang, S.; Gapsys, V.; Kim, H.-Y.; Bessonov, S.; Hsiao, H.-H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Lührmann, R.; de Groot, B.; Zweckstetter, M. *Structure* **2013**, *21*, 2162–2174.

(138) Demarco, A.; LlináS, M.; Wuthrich, K. *Biopolymers* **1978**, *17*, 2727–2742.

(139) Karimi-Nejad, Y.; Schmidt, J. M.; Rueterjans, H.; Schwalbe, H.; Griesinger, C. *Biochemistry* **1994**, *33*, 5481–5492.

(140) West, N. J.; Smith, L. J. *Journal of Molecular Biology* **1998**, *280*, 867–877.

(141) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *Journal of Molecular Biology* **1999**, *288*, 705–723.

(142) Vajpai, N.; Gentner, M.; Huang, J.-r.; Blackledge, M.; Grzesiek, S. *Journal of the American Chemical Society* **2010**, *132*, 3196–3203.

(143) Sziegat, F.; Silvers, R.; Hähnke, M.; Jensen, M. R.; Blackledge, M.; Wirmer-Bartoschek, J.; Schwalbe, H. *Biochemistry* **2012**, *51*, 3361–3372.

(144) Gomes, G.-N. W.; Krzeminski, M.; Namini, A.; Martin, E. W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J. D.; Gradinaru, C. C. *Journal of the American Chemical Society* **2020**, *142*, 15697–15710.

(145) Naudi-Fabra, S.; Tengo, M.; Jensen, M. R.; Blackledge, M.; Milles, S. *Journal of the American Chemical Society* **2021**, *143*, 20109–20121.

(146) Jeschke, G.; Pannier, M.; Spiess, H. W. In *Distance Measurements in Biological Systems by EPR*, Berliner, L. J., Eaton, G. R., Eaton, S. S., Eds.; Biological Magnetic Resonance; Springer US: Boston, MA, 2000, pp 493–512.

(147) Lovett, J. E.; Bowen, A. M.; Timmel, C. R.; Jones, M. W.; Dilworth, J. R.; Caprotti, D.; Bell, S. G.; Wong, L. L.; Harmer, J. *Physical Chemistry Chemical Physics* **2009**, *11*, 6840–6848.

(148) Murphy, D. M.; Chechik, V., *Electron Paramagnetic Resonance: Volume 26*; Royal Society of Chemistry: 2018; 186 pp.

(149) Sicoli, G.; Kress, T.; Vezin, H.; Ledolter, K.; Kurzbach, D. *The Journal of Physical Chemistry Letters* **2020**, *11*, 8944–8951.

(150) Kachala, M.; Valentini, E.; Svergun, D. I. In *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, Felli, I. C., Pierattelli, R., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, 2015, pp 261–289.

(151) Kikhney, A. G.; Svergun, D. I. *FEBS Letters* **2015**, *589*, 2570–2577.

(152) Sibille, N.; Bernadó, P. *Biochemical Society Transactions* **2012**, *40*, 955–962.

(153) Hermann, M. R.; Hub, J. S. *Journal of Chemical Theory and Computation* **2019**, *15*, 5103–5115.

(154) Bernadó, P.; I. Svergun, D. *Molecular BioSystems* **2012**, *8*, 151–167.

(155) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *Journal of the American Chemical Society* **2007**, *129*, 5656–5664.

(156) Lessing, J.; Roy, S.; Reppert, M.; Baer, M.; Marx, D.; Jansen, T. L. C.; Knoester, J.; Tokmakoff, A. *Journal of the American Chemical Society* **2012**, *134*, 5032–5035.

(157) Sethi, A.; Anunciado, D.; Tian, J.; Vu, D. M.; Gnanakaran, S. *Chemical Physics* **2013**, *422*, 143–155.

(158) Weiss, S. *Nature Structural Biology* **2000**, *7*, 724–729.

(159) Schuler, B.; Müller-Späth, S.; Soranno, A.; Nettels, D. In *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, Uversky, V. N., Dunker, A. K., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2012, pp 21–45.

(160) Haas, E. In *Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools*, Uversky, V. N., Dunker, A. K., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012, pp 467–498.

(161) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. *Biophysical Journal* **2012**, *9*, 1940–1949.

(162) LeBlanc, S. J.; Kulkarni, P.; Weninger, K. R. *Biomolecules* **2018**, *8*, 140.

(163) Tuckerman, M.; Tuckerman, M., *Statistical Mechanics: Theory and Molecular Simulation*, New Edition, Second Edition, New Edition, Second Edition; Oxford Graduate Texts; Oxford University Press: Oxford, New York, 2023; 880 pp.

(164) Brooks, B. R. et al. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

(165) Lee, S.; Tran, A.; Allsopp, M.; Lim, J. B.; Hénin, J.; Klauda, J. B. *The Journal of Physical Chemistry B* **2014**, *118*, 547–556.

(166) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. *Nature Methods* **2017**, *14*, 71–73.

(167) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(168) Yu, L.; Li, D.-W.; Brüschweiler, R. *Journal of Chemical Theory and Computation* **2020**, *16*, 1311–1318.

(169) Mu, J.; Liu, H.; Zhang, J.; Luo, R.; Chen, H.-F. *Journal of Chemical Information and Modeling* **2021**, *61*, 1037–1047.

(170) Levitt, M. *Journal of Molecular Biology* **1976**, *104*, 59–107.

(171) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531–1548.

(172) Tirion, M. M. *Physical Review Letters* **1996**, *77*, 1905–1908.

(173) Takada, S. *Current Opinion in Structural Biology* **2012**, *22*, 130–137.

(174) Tozzini, V. *Current Opinion in Structural Biology* **2005**, *15*, 144–150.

(175) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. *Chemical Reviews* **2016**, *116*, 7898–7936.

(176) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(177) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *The Journal of Physical Chemistry B* **2004**, *108*, 750–760.

(178) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.

(179) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2013**, *9*, 687–697.

(180) Vitalis, A.; Pappu, R. V. In *Annual Reports in Computational Chemistry*, Wheeler, R. A., Ed.; Elsevier: 2009; Vol. 5, pp 49–76.

(181) Bottaro, S.; Lindorff-Larsen, K. *Science* **2018**, *361*, 355–360.

(182) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-r.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463–1470.

(183) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. *Current Opinion in Structural Biology* **2017**, *42*, 106–116.

(184) Thomasen, F. E.; Lindorff-Larsen, K. *Biochemical Society Transactions* **2022**, *50*, 541–554.

(185) Ravera, E.; Sgheri, L.; Parigi, G.; Luchinat, C. *Physical Chemistry Chemical Physics* **2016**, *18*, 5686–5701.

(186) Chan-Yao-Chong, M.; Durand, D.; Ha-Duong, T. *Journal of Chemical Information and Modeling* **2019**, *59*, 1743–1758.

(187) Jensen, M. R.; Zweckstetter, M.; Huang, J.-r.; Blackledge, M. *Chemical Reviews* **2014**, *114*, 6632–6660.

(188) Pfaendtner, J.; Bonomi, M. *Journal of Chemical Theory and Computation* **2015**, *11*, 5062–5067.

(189) Choy, W.-Y.; Forman-Kay, J. D. *Journal of Molecular Biology* **2001**, *308*, 1011–1032.

(190) Marsh, J. A.; Forman-Kay, J. D. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 556–572.

(191) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *Journal of the American Chemical Society* **2010**, *132*, 1270–1272.

(192) Huang, A.; Stultz, C. M. *PLOS Computational Biology* **2008**, *4*, e1000155.

(193) Fisher, C. K.; Huang, A.; Stultz, C. M. *Journal of the American Chemical Society* **2010**, *132*, 14919–14927.

(194) Ullman, O.; Fisher, C. K.; Stultz, C. M. *Journal of the American Chemical Society* **2011**, *133*, 19536–19546.

(195) Levy, R. M.; Karplus, M.; Wolynes, P. G. *Journal of the American Chemical Society* **1981**, *103*, 5998–6011.

(196) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *Journal of the American Chemical Society* **1992**, *114*, 2289–2302.

(197) Brüschweiler, R. *Journal of the American Chemical Society* **1992**, *114*, 5341–5344.

(198) Schmidt, J. M.; Brüschweiler, R.; Ernst, R. R.; Dunbrack, R. L.; Joseph, D.; Karplus, M. *Journal of the American Chemical Society* **1993**, *115*, 8747–8756.

(199) Bremi, T.; Brüschweiler, R.; Ernst, R. R. *Journal of the American Chemical Society* **1997**, *119*, 4272–4284.

(200) Brüschweiler, R.; Wright, P. E. *Journal of the American Chemical Society* **1994**, *116*, 8426–8427.

(201) Bremi, T.; Brüschweiler, R. *Journal of the American Chemical Society* **1997**, *119*, 6672–6673.

(202) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *Journal of the American Chemical Society* **1998**, *120*, 9870–9879.

(203) Prompers, J. J.; Scheurer, C.; Brüschweiler, R. *Journal of Molecular Biology* **2001**, *305*, 1085–1097.

(204) Prompers, J. J.; Brüschweiler, R. *Journal of the American Chemical Society* **2001**, *123*, 7305–7313.

(205) Prompers, J. J.; Brüschweiler, R. *Journal of the American Chemical Society* **2002**, *124*, 4522–4534.

(206) Xue, Y.; Skrynnikov, N. R. *Journal of the American Chemical Society* **2011**, *133*, 14614–14628.

(207) Showalter, S. A.; Brüschweiler, R. *Journal of Chemical Theory and Computation* **2007**, *3*, 961–975.

(208) Salvi, N.; Abyzov, A.; Blackledge, M. *The Journal of Physical Chemistry Letters* **2016**, *7*, 2483–2489.

(209) Salvi, N.; Abyzov, A.; Blackledge, M. *Angewandte Chemie (International Ed. in English)* **2017**, *56*, 14020–14024.

(210) Salvi, N.; Abyzov, A.; Blackledge, M. *Science Advances* **2019**, *5*, eaax2348.

(211) Abascal, J. L. F.; Vega, C. *The Journal of Chemical Physics* **2005**, *123*, 234505.

(212) Bolik-Coulon, N.; Ferrage, F. *The Journal of Chemical Physics* **2022**, *157*, 125102.

(213) Li, D.-W.; Brüschweiler, R. *Angewandte Chemie* **2010**, *122*, 6930–6932.

(214) Hoffmann, F.; Mulder, F. A. A.; Schäfer, L. V. *The Journal of Physical Chemistry B* **2018**, *122*, 5038–5048.

(215) Cayley, S.; Lewis, B. A.; Guttman, H. J.; Record, M. T. *Journal of Molecular Biology* **1991**, *222*, 281–300.

(216) Zimmerman, S. B.; Trach, S. O. *Journal of Molecular Biology* **1991**, *222*, 599–620.

(217) Zeskind, B. J.; Jordan, C. D.; Timp, W.; Trapani, L.; Waller, G.; Horodincu, V.; Ehrlich, D. J.; Matsudaira, P. *Nature Methods* **2007**, *4*, 567–569.

(218) Kang, H.; Pincus, P. A.; Hyeon, C.; Thirumalai, D. *Physical Review Letters* **2015**, *114*, 068303.

(219) Parsegian, V. A.; Rand, R. P.; Rau, D. C. *Proceedings of the National Academy of Sciences* **2000**, *97*, 3987–3992.

(220) Sarkar, M.; Li, C.; Pielak, G. J. *Biophysical Reviews* **2013**, *5*, 187–194.

(221) Kim, Y. C.; Mittal, J. *Physical Review Letters* **2013**, *110*, 208102.

(222) Kuznetsova, I. M.; Zaslavsky, B. Y.; Breydo, L.; Turoverov, K. K.; Uversky, V. N. *Molecules* **2015**, *20*, 1377–1409.

(223) Miller, C. M.; Kim, Y. C.; Mittal, J. *Biophysical Journal* **2016**, *111*, 28–37.

(224) Balu, R.; Wanasingha, N.; Mata, J. P.; Rekas, A.; Barrett, S.; Dumsday, G.; Thornton, A. W.; Hill, A. J.; Roy Choudhury, N.; Dutta, N. K. *Science Advances* **2022**, *8*, eabq2202.

(225) Kalwarczyk, T.; Kwapiszewska, K.; Szczepanski, K.; Sozanski, K.; Szymanski, J.; Michalska, B.; Patalas-Krawczyk, P.; Duszynski, J.; Holyst, R. *The Journal of Physical Chemistry B* **2017**, *121*, 9831–9837.

(226) Makuch, K.; Hołyst, R.; Kalwarczyk, T.; Garstecki, P.; F. Brady, J. *Soft Matter* **2020**, *16*, 114–124.

(227) Agasty, A.; Wisniewska, A.; Kalwarczyk, T.; Koynov, K.; Holyst, R. *ACS Applied Polymer Materials* **2021**, *3*, 2813–2822.

(228) Gennes, P.-G. d., *Scaling Concepts in Polymer Physics*; Cornell University Press: 1979; 336 pp.

(229) Brochard Wyart, F.; de Gennes, P. *The European Physical Journal E* **2000**, *1*, 93–97.

(230) Holyst, R.; Bielejewska, A.; Szymański, J.; Wilk, A.; Patkowski, A.; Gapiński, J.; Żywociński, A.; Kalwarczyk, T.; Kalwarczyk, E.; Tabaka, M.; Ziębacz, N.; Wieczorek, S. A. *Physical Chemistry Chemical Physics* **2009**, *11*, 9025–9032.

(231) Kalwarczyk, T.; Ziębacz, N.; Bielejewska, A.; Zaboklicka, E.; Koynov, K.; Szymański, J.; Wilk, A.; Patkowski, A.; Gapiński, J.; Butt, H.-J.; Hołyst, R. *Nano Letters* **2011**, *11*, 2157–2163.

(232) Kalwarczyk, T.; Tabaka, M.; Holyst, R. *Bioinformatics* **2012**, *28*, 2971–2978.

(233) Tabaka, M.; Kalwarczyk, T.; Szymanski, J.; Hou, S.; Holyst, R. *Frontiers in Physics* **2014**, *2*.

(234) Kalwarczyk, T.; Sozanski, K.; Ochab-Marcinek, A.; Szymanski, J.; Tabaka, M.; Hou, S.; Holyst, R. *Advances in Colloid and Interface Science* **2015**, *223*, 55–63.

(235) Sekhar, A.; Latham, M. P.; Vallurupalli, P.; Kay, L. E. *The Journal of Physical Chemistry B* **2014**, *118*, 4546–4551.

(236) Verkman, A. S. *Trends in Biochemical Sciences* **2002**, *27*, 27–33.

(237) Kühn, T.; Ihalainen, T. O.; Hyväluoma, J.; Dross, N.; Willman, S. F.; Langowski, J.; Vihinen-Ranta, M.; Timonen, J. *PLOS ONE* **2011**, *6*, e22962.

(238) König, I.; Soranno, A.; Nettels, D.; Schuler, B. *Angewandte Chemie* **2021**, *133*, 10819–10824.

(239) König, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B. *Nature Methods* **2015**, *12*, 773–779.

(240) Plitzko, J. M.; Schuler, B.; Selenko, P. *Current Opinion in Structural Biology* **2017**, *46*, 110–121.

(241) Kozer, N.; Schreiber, G. *Journal of Molecular Biology* **2004**, *336*, 763–774.

(242) Zhou, Y.-L.; Liao, J.-M.; Chen, J.; Liang, Y. *The International Journal of Biochemistry & Cell Biology* **2006**, *38*, 1986–1994.

(243) Kozer, N.; Kuttner, Y. Y.; Haran, G.; Schreiber, G. *Biophysical Journal* **2007**, *92*, 2139–2149.

(244) Kim, Y. C.; Best, R. B.; Mittal, J. *The Journal of Chemical Physics* **2010**, *133*, 205101.

(245) Doi, M.; Edwards, S. F.; Edwards, S. F., *The Theory of Polymer Dynamics*; Clarendon Press: 1988; 420 pp.

(246) Asakura, S.; Oosawa, F. *The Journal of Chemical Physics* **1954**, *22*, 1255–1256.

(247) Vrij, A. *Pure and Applied Chemistry* **1976**, *48*, 471–483.

(248) Lekkerkerker, H. N.; Tuinier, R., *Colloids and the Depletion Interaction*; Lecture Notes in Physics, Vol. 833; Springer Netherlands: Dordrecht, 2011.

(249) Berezhkovskii, A. M.; Szabo, A. *The Journal of Physical Chemistry B* **2016**, *120*, 5998–6002.

(250) Zosel, F.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Schuler, B. *Proceedings of the National Academy of Sciences* **2020**, *117*, 13480–13489.

(251) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 285–298.

(252) Shin, Y.; Brangwynne, C. P. *Science* **2017**, *357*, eaaf4382.

(253) Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. *Nature Reviews Molecular Cell Biology* **2023**, *24*, 288–304.

(254) Gomes, E.; Shorter, J. *Journal of Biological Chemistry* **2019**, *294*, 7115–7127.

(255) Darling, A. L.; Liu, Y.; Oldfield, C. J.; Uversky, V. N. *PROTEOMICS* **2018**, *18*, 1700193.

(256) Flory, P. J. *The Journal of Chemical Physics* **1942**, *10*, 51–61.

(257) Huggins, M. L. *Journal of the American Chemical Society* **1942**, *64*, 1712–1719.

(258) Overbeek, J. T. G.; Voorn, M. J. *Journal of Cellular and Comparative Physiology* **1957**, *49*, 7–26.

(259) Perry, S. L. *Current Opinion in Colloid & Interface Science* **2019**, *39*, 86–97.

(260) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. *Cell* **2018**, *174*, 688–699.e16.

(261) Rubinstein, M.; Dobrynin, A. V. *Trends in Polymer Science* **1997**, *6*, 181–186.

(262) Semenov, A. N.; Rubinstein, M. *Macromolecules* **1998**, *31*, 1373–1385.

(263) Sherrill, C. D. *Accounts of Chemical Research* **2013**, *46*, 1020–1028.

(264) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E. *Proceedings of the National Academy of Sciences* **2017**, *114*, E8194–E8203.

(265) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. *Nature Structural & Molecular Biology* **2019**, *26*, 637–648.

(266) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. *eLife* **2018**, *7*, ed. by Shan, Y., e31486.

(267) Monahan, Z.; Ryan, V. H.; Janke, A. M.; Burke, K. A.; Rhoads, S. N.; Zerze, G. H.; O'Meally, R.; Dignon, G. L.; Conicella, A. E.; Zheng, W.; Best, R. B.; Cole, R. N.; Mittal, J.; Shewmaker, F.; Fawzi, N. L. *The EMBO Journal* **2017**, *36*, 2951–2967.

(268) Lin, Y.; Currie, S. L.; Rosen, M. K. *Journal of Biological Chemistry* **2017**, *292*, 19110–19120.

(269) Qamar, S. et al. *Cell* **2018**, *173*, 720–734.e15.

(270) Krainer, G. et al. *Nature Communications* **2021**, *12*, 1085.

(271) Taylor, N. O.; Wei, M.-T.; Stone, H. A.; Brangwynne, C. P. *Biophysical Journal* **2019**, *117*, 1285–1300.

(272) Abyzov, A.; Blackledge, M.; Zweckstetter, M. *Chemical Reviews* **2022**, *122*, 6719–6748.

(273) Wei, M.-T.; Elbaum-Garfinkle, S.; Holehouse, A. S.; Chen, C. C.-H.; Feric, M.; Arnold, C. B.; Priestley, R. D.; Pappu, R. V.; Brangwynne, C. P. *Nature Chemistry* **2017**, *9*, 1118–1125.

(274) Murakami, T. et al. *Neuron* **2015**, *88*, 678–690.

(275) Mitrea, D. M.; Chandra, B.; Ferrolino, M. C.; Gibbs, E. B.; Tolbert, M.; White, M. R.; Kriwacki, R. W. *Journal of Molecular Biology* **2018**, *430*, 4773–4805.

(276) Zhang, H.; Elbaum-Garfinkle, S.; Langdon, E. M.; Taylor, N.; Occhipinti, P.; Bridges, A. A.; Brangwynne, C. P.; Gladfelter, A. S. *Molecular Cell* **2015**, *60*, 220–230.

(277) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; Tompa, P.; Fuxreiter, M. *Trends in Cell Biology* **2018**, *28*, 420–435.

(278) Elbaum-Garfinkle, S.; Kim, Y.; Szczepaniak, K.; Chen, C. C.-H.; Eckmann, C. R.; Myong, S.; Brangwynne, C. P. *Proceedings of the National Academy of Sciences* **2015**, *112*, 7189–7194.

(279) Hofweber, M.; Hutten, S.; Bourgeois, B.; Spreitzer, E.; Niedner-Boblenz, A.; Schifferer, M.; Ruepp, M.-D.; Simons, M.; Niessing, D.; Madl, T.; Dormann, D. *Cell* **2018**, *173*, 706–719.e13.

(280) Lin, Y.; Protter, D. S. W.; Rosen, M. K.; Parker, R. *Molecular cell* **2015**, *60*, 208–219.

(281) Ray, S. et al. *Nature Chemistry* **2020**, *12*, 705–716.

(282) Burke, K. A.; Janke, A. M.; Rhine, C. L.; Fawzi, N. L. *Molecular Cell* **2015**, *60*, 231–241.

(283) Reichheld, S. E.; Muiznieks, L. D.; Keeley, F. W.; Sharpe, S. *Proceedings of the National Academy of Sciences* **2017**, *114*, E4408–E4415.

(284) Kim, T. H.; Payliss, B. J.; Nosella, M. L.; Lee, I. T. W.; Toyama, Y.; Forman-Kay, J. D.; Kay, L. E. *Proceedings of the National Academy of Sciences* **2021**, *118*, DOI: 10.1073/pnas.2104897118.

(285) Yuwen, T.; Brady, J. P.; Kay, L. E. *Journal of the American Chemical Society* **2018**, *140*, 2115–2126.

(286) Thakur, A.; Chandra, K.; Dubey, A.; D'Silva, P.; Atreya, H. S. *Angewandte Chemie (International ed in English)* **2013**, *52*, 2440–2443.

(287) Yuwen, T.; Bah, A.; Brady, J. P.; Ferrage, F.; Bouvignies, G.; Kay, L. E. *The Journal of Physical Chemistry B* **2018**, *122*, 11206–11217.

(288) Zhang, X.; Lin, Y.; Eschmann, N. A.; Zhou, H.; Rauch, J. N.; Hernandez, I.; Guzman, E.; Kosik, K. S.; Han, S. *PLOS Biology* **2017**, *15*, e2002183.

(289) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. *PLOS Computational Biology* **2018**, *14*, e1005941.

(290) Kapcha, L. H.; Rossky, P. J. *Journal of Molecular Biology* **2014**, *426*, 484–498.

(291) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. *ACS Central Science* **2019**, *5*, 821–830.

(292) Alshareedah, I.; Moosa, M. M.; Raju, M.; Potoyan, D. A.; Banerjee, P. R. *Proceedings of the National Academy of Sciences* **2020**, *117*, 15650–15658.

(293) Regy, R. M.; Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. *Nucleic Acids Research* **2020**, *48*, 12593–12603.

(294) Perdikari, T. M.; Jovic, N.; Dignon, G. L.; Kim, Y. C.; Fawzi, N. L.; Mittal, J. *Biophysical Journal* **2021**, *120*, 1187–1197.

(295) Wohl, S.; Jakubowski, M.; Zheng, W. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6684–6691.

(296) Garaizar, A.; Espinosa, J. R. *The Journal of Chemical Physics* **2021**, *155*, 125103.

(297) Latham, A. P.; Zhang, B. *Journal of Chemical Theory and Computation* **2020**, *16*, 773–781.

(298) Dannenhoffer-Lafage, T.; Best, R. B. *The Journal of Physical Chemistry B* **2021**, *125*, 4046–4056.

(299) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2111696118.

(300) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. *Protein Science* **2021**, *30*, 1371–1379.

(301) Choi, J.-M.; Dar, F.; Pappu, R. V. *PLOS Computational Biology* **2019**, *15*, e1007028.

(302) Zeng, X.; Holehouse, A. S.; Chilkoti, A.; Mittag, T.; Pappu, R. V. *Biophysical Journal* **2020**, *119*, 402–418.

(303) McCarty, J.; Delaney, K. T.; Danielsen, S. P. O.; Fredrickson, G. H.; Shea, J.-E. *The Journal of Physical Chemistry Letters* **2019**, *10*, 1644–1652.

(304) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *The Journal of Physical Chemistry B* **2012**, *116*, 8494–8503.

(305) Wu, H.; Wolynes, P. G.; Papoian, G. A. *The Journal of Physical Chemistry. B* **2018**, *122*, 11115–11125.

(306) Hyman, A. A.; Simons, K. *Science* **2012**, *337*, 1047–1049.

(307) Alberti, S.; Gladfelter, A.; Mittag, T. *Cell* **2019**, *176*, 419–434.

(308) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. *Nature Reviews Molecular Cell Biology* **2021**, *22*, 215–235.

(309) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. *Nature Physics* **2015**, *11*, 899–904.

(310) Fawzi, N. L.; Parekh, S. H.; Mittal, J. *Current Opinion in Structural Biology* **2021**, *70*, 78–86.

(311) Persson, E.; Halle, B. *Proceedings of the National Academy of Sciences* **2008**, *105*, 6266–6271.

(312) Boeynaems, S.; Holehouse, A. S.; Weinhardt, V.; Kovacs, D.; Van Lindt, J.; Larabell, C.; Van Den Bosch, L.; Das, R.; Tompa, P. S.; Pappu, R. V.; Gitler, A. D. *Proceedings of the National Academy of Sciences* **2019**, *116*, 7889–7898.

(313) Zhang, F.; Brüschweiler, R. *Journal of the American Chemical Society* **2002**, *124*, 12654–12655.

(314) Guseva, S.; Milles, S.; Jensen, M. R.; Schoehn, G.; Ruigrok, R. W.; Blackledge, M. *Current Opinion in Virology* **2020**, *41*, 59–67.

(315) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977–8991.

(316) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. *The Journal of Physical Chemistry B* **2020**, *124*, 11671–11679.

(317) Cai, M.; Huang, Y.; Yang, R.; Craigie, R.; Clore, G. M. *Journal of Biomolecular NMR* **2016**, *66*, 85–91.

(318) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *Journal of Biomolecular NMR* **1995**, *6*, 277–293.

(319) Lakomek, N.-A.; Ying, J.; Bax, A. *Journal of Biomolecular NMR* **2012**, *53*, 209–221.

(320) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *Journal of the American Chemical Society* **2012**, *134*, 8148–8161.

(321) Flory, P., *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953; 688 pp.

(322) Wang, M.; Song, X.; Chen, J.; Chen, X.; Zhang, X.; Yang, Y.; Liu, Z.; Yao, L. *Science Advances* **2023**, *9*, eadg9141.

(323) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P. *Chemical Reviews* **2014**, *114*, 6661–6714.

(324) Theillet, F.-X. *Chemical Reviews* **2022**, *122*, 9497–9570.

(325) Guseva, S.; Milles, S.; Jensen, M. R.; Salvi, N.; Kleman, J.-P.; Maurin, D.; Ruigrok, R. W. H.; Blackledge, M. *Science Advances*, *6*, eaaz7095.

(326) Jensen, M. R.; Communie, G.; Ribeiro, E. A.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W. H.; Blackledge, M. *Proceedings of the National Academy of Sciences* **2011**, *108*, 9839–9844.

(327) Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. *Proceedings of the National Academy of Sciences* **2004**, *101*, 8301–8306.

(328) Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D. F.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M. *Journal of the American Chemical Society* **2015**, *137*, 1220–1229.

(329) Breeze, A. *Progress in Nuclear Magnetic Resonance Spectroscopy - PROG NUCL MAGN RESON SPECTROS* **2000**, *36*, 323–372.

(330) Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M. F.; Nettels, D.; Best, R. B.; Schuler, B. *Nature* **2023**, *619*, 876–883.