Hydration water dynamics of the tau protein in its native and amyloid states

Yann Fichou

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 7 Aout 2006

Présentée par

Yann FICHOU

Thèse dirigée par **Martin WEIK**

préparée au sein **de l'Institut de Biologie Structurale**
et de **l'école doctorale de physique**

Dynamique de l'eau d'hydratation de la protéine tau dans ses formes native et amyloïde

Thèse soutenue publiquement le **11 mars 2015**,
devant le jury composé de :

Pr Antonio Cupane
Université de Palerme, Palerme, Italie, Rapporteur
Pr Damien Laage
Ecole normale supérieure, Paris, France, Rapporteur
Dr Martin Blackledge
Institut de biologie structurale, Grenoble, France, Examinateur
Pr Marc Jamin
Institut de biologie structurale, Grenoble, France, Examinateur
Dr Véronique Receveur-Bréchot
Centre de recherche en cancérologie de Marseille, Marseille, France, Examinatrice
Pr Douglas Tobias
Université de Californie Irvine, Californie, Etats-unis, Examinateur
Dr Giuseppe Zaccai
Institut Laue-Langevin, Grenoble, France, Examinateur
Dr Martin Weik
Institut de biologie structurale, Grenoble, France, Directeur de thèse

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 7 Aout 2006

Présentée par

Yann FICHOU

Thèse dirigée par **Martin WEIK**

préparée au sein **de l'Institut de Biologie Structurale**
et de **l'école doctorale de physique**

Hydration water dynamics of the tau protein in its native and amyloid states

Thèse soutenue publiquement le **11 mars 2015**,
devant le jury composé de :

Pr Antonio Cupane
Université de Palerme, Palerme, Italie, Rapporteur
Pr Damien Laage
Ecole normale supérieure, Paris, France, Rapporteur
Dr Martin Blackledge
Institut de biologie structurale, Grenoble, France, Examinateur
Pr Marc Jamin
Institut de biologie structurale, Grenoble, France, Examinateur
Dr Véronique Receveur-Bréchot
Centre de recherche en cancérologie de Marseille, Marseille, France, Examinatrice
Pr Douglas Tobias
Université de Californie Irvine, Californie, Etats-unis, Examinateur
Dr Giuseppe Zaccai
Institut Laue-Langevin, Grenoble, France, Examinateur
Dr Martin Weik
Institut de biologie structurale, Grenoble, France, Directeur de thèse

Acknowledgements - Remerciements

First of all, I would like to thank Damien Laage and Antonio Cupane for reviewing this manuscript. I am also grateful to Véronique Receveur-Bréchot, Marc Jamin, Martin Blackledge, Douglas Tobias, Martin Weik and Giuseppe Zaccai to have been part of my defense committee.

My most sincere acknowledgment goes to my advisor, Martin Weik, who made this whole experience possible. He trained me to carry out proper research and initiated many projects on which I was honored to work. I want to thank Martin for having given me the freedom of research in the different projects, although he was always available for advices and discussions. It was scientifically and personally very enriching to work with such a great person. Martin, es hat mich sehr gefreut mit dir zu arbeiten.

I'm truly grateful to Douglas Tobias, who welcomed me in his group at UCI for about 7 months. He allowed me to learn a lot about MD simulations and to apply it to my research projects. Our discussions have always been very pleasant and fruitful.

Je ne vois pas comment remercier François-Xavier Gallat si ce n'est en utilisant la langue de Molière, si chère à son coeur. Merci FX de m'avoir appris ce qu'il fallait savoir de cette fameuse protéine tau, ainsi que de m'avoir initié à la diffusion de neutron. Ce fut toujours un réel plaisir d'échanger avec toi.

Je voudrais remercier Giorgio Schirò, pour nos discussions scientifiques sur la diffusion de neutron. Avec un sens de l'humour à la hauteur de sa connaissance des différentes techniques de diffusion, il m'a beaucoup appris pendant ma dernière année de thèse.

Je voudrais remercier tous les membres, passés et présents, de l'équipe DYNAMOP. Tout d'abord, Mathilde lethier, qui m'a enseigné les bases de la biochimie en faisant preuve de beaucoup de pédagogie. Mes interactions avec Antoine, Jacques, Virgile, Chady, Maryam, Virginia, Joyce, Damien et Céline furent toujours plaisantes. Merci également à Aude Vernet qui a beaucoup accompli lors de son stage dans notre équipe. Gianluca, ce fut un plaisir d'être à tes cotés du début à la fin de ma thèse. Enfin, je tiens à remercier Linda Ponet pour l'immense travail administratif qu'elle a accompli pour moi.

i

ACKNOWLEDGEMENTS - REMERCIEMENTS

Preface

Long before being a scientific quest, water had been a central substance throughout the history of human civilizations, as described in the book H_2O: *a biography of water*, by Philip Ball. In ancient Egypt, water was this flowing substance that one day could bring life and prosperity, and the other day could spread death and misery over the world. Water was once the ultimate element before dissociation, which could turn into essentially everything present on earth. For about two millenniums, water had remained one of the four fundamental elements from which derives everything in the universe. Then, No wonder that water has long been the object of many fantasies and mysteries. Lavoisier pioneered the rigorous work that provides the knowledge of water as a mere chemical compound, in the 18th century. Yet, even modern science has trouble to remain objective whenever water is involved, as evidenced by several passionate debates that have animated the scientific community in the 20th century. Nevertheless, it seems that water has not yet revealed all its secrets, as reviewed in chapter 1, and hopefully this work will contribute to a better understanding of the substance in a biological context.

The work on the dynamics of the tau protein and its hydration water follows up on the research carried out by Dr. François-Xavier Galla and Dr. Kathleen Wood, on the dynamical properties of different classes of proteins, using incoherent neutron scattering. The work presented in this manuscript was mostly achieved in the DYNAMOP group at the Institut de Biologie Structurale (IBS), under the supervision of Dr. Martin Weik. In addition, I had the honor to receive a Fulbright fellowship that allowed me spending almost seven months in the group of Prof. Douglas Tobias at the University of California Irvine (UCI). The MD simulation studies presented here were carried out at UCI under the supervision of Douglas Tobias. Thanks to the collaboration between these two groups, I was able to address the proposed issues by using a combination of computational and experimental techniques, which turned out to be a powerful approach. The work on THz spectroscopy was achieved in collaboration with the group of Prof. Martina Havenith at the University of Bochum, in Germany.

The results of the thesis work are presented as a gathering of manuscripts, which are submitted or in a final stage before submission. The first chapter sets the background

and motivations of the work, before introducing the different methods used in the second chapter. The five following sections present the different results in the form of journal articles, with the exception of chapter 6. Each chapter ends with a brief summary, which can all be read independently in order to obtain rapidly a complete overview of the thesis in a few minutes. A detailed summary of the manuscript is provided in French in appendix C.

Contents

Chapter 1

Introduction

This chapter sets the scientific background of the issues investigated in this manuscript. First, we argue that water is an essential substance in biology, so important that it has been named the *matrix of life* (Szent-Gyorgyi, 1979). We provide several examples demonstrating that water molecules should be considered as biomolecules on their own, which are required to interact with all biological entities, including proteins and nucleic acids. Second, the focus will be put on proteins and their dynamics. After recalling that the dynamical features of a protein are as important as its structural properties, we introduce the notion of energy landscape as a means of understanding protein stability and dynamics. Then, we focus on a recently-discovered class of proteins that is composed of the intrinsically disordered proteins (IDPs). Although research on IDPs has skyrocketed only about fifteen years ago, it is now established that they are ubiquitous in living systems and they fulfilled a large variety of roles. IDPs have drawn a particular attention, because their aggregation is involved in many diseases including neurodegenerative disorders. Finally we focus on a particular IDP called tau that has been the main object of this thesis. Present in neuronal cells where it regulates microtubule stability, the tau protein is of particular interest because its fibrillation is one of the hallmarks of the Alzheimer disease.

1.1   Water, the matrix of life

1.1.1   Molecular crowding

We have all been taught that 70% of our body mass is made of water. However it is somewhat hard to transpose this value at a molecular level and get an intuition of what the interior of a living cell is like. It is crowded, very crowded. The total concentration of macromolecules inside cells can be up to 400 mg/ml, meaning that they occupy up to 40% of the total volume (Ellis and Minton, 2003). The average distance between macromolecules is roughly 20 angstrom (Å; 1 Å = 10^{-10} meter), which corresponds to about six layers of water. Those numbers are well illustrated by David Goodsell (Goodsell, 2009) who made an elegant and realistic painting of the inside of the bacteria *Escherichia coli* (see figure 1.1).
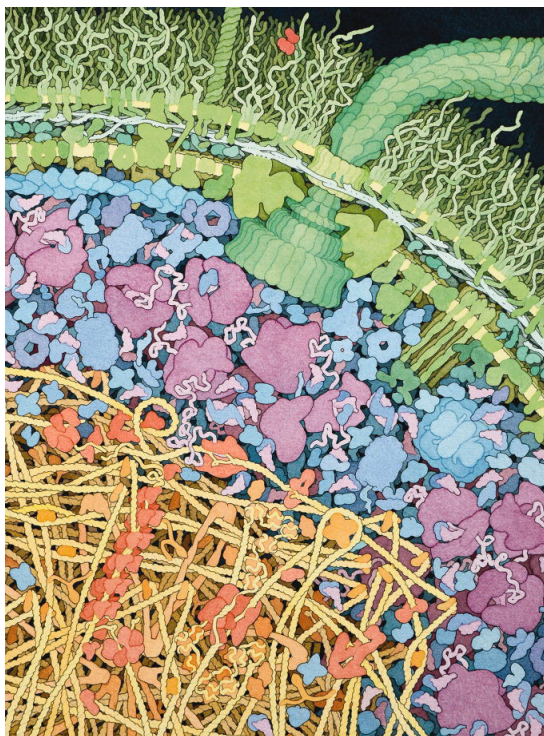
Figure 1.1: Schematic representation of the molecular crowd inside the bacteria *E. coli* extracted from (Goodsell, 2009).

While moving in a random-walk manner, supplied by thermal energy, macromolecules bump into each other, sometimes ricocheting, and sometimes interacting to accomplish their vital function. One might naively think of water as a biologically inert medium that only fills the gap between biomolecules. In reality, this view of water being a mere bystander in the play of life has been severely reviewed. Water molecules are actually

active components, ubiquitous in life, that have to be considered as biomolecules on their own. Macromolecular interactions are mediated by their surrounding water molecules, the so-called hydration shell, thereby enabling the complex machinery of life to function. Thus, in order to deepen our understanding of biological processes, scientists have studied how water influences biomolecules such as proteins and reciprocally how proteins influence water, so I have done in the present work.

1.1.2 Water and biological function

Water fulfills a variety of tasks in different biological processes, as thoroughly reviewed by Philip Ball (Ball, 2008). I report in this section a few examples that are directly or indirectly relevant to the work presented in the manuscript.

One essential role of water in biology is its participation in protein folding, mainly through the hydrophobic effect (Baldwin, 2014). Although the precise mechanisms remain incompletely understood, the basic principle is that proteins collapse upon their hydrophobic residues to hide them from the surrounding water, therefore maximizing water entropy. The folding process is probably twofold, the protein first collapse to a near-native conformation that retains water in its interior, followed by a slower expulsion of its residual water (Cheung et al., 2002; Garcia and Onuchic, 2003). Water molecules also act as temporary linkers between hydrophilic residues during the folding process, holding thereby the folded chain together until it is ready for final compaction (Papoian et al., 2004). Indeed, because one water molecule can form multiple hydrogen bonds, it can mediate protein-protein and protein-ligand interactions. Philip Ball (Ball, 2008) wrote "In other words, it is not simply the case that water molecules can bridge two proteins: such contacts can be imbued with significant information content that allows the interactions to be discriminating. Thus, protein surfaces in a sense extend the range of their influence via their hydration shells". And this does not only apply to proteins but also to nucleic acids (Robinson and Sligar, 1993). When proteins achieve their biological function, dynamical and structural changes in their hydration water are quite often observed (Barillari et al., 2007; Autenrieth et al., 2004; Royer et al., 1996; Kornblatt and Hoa, 1990; Colombo et al., 1992; Grossman et al., 2011). The importance of these changes for macromolecular functions is, however, hard to evaluate and often remains speculative. We add one piece to that statement in chapter 5 by reporting the enhancement of hydration water dynamics after the aggregation of a protein. Recently, it was proposed that a gradient in hydration water dynamics at the protein surface might drive a ligand to its right location, acting as a "hydration funnel" (Conti Nibali and Havenith, 2014).

Proteins involved in electron and proton transfer play a central role in many biological processes, including photosynthesis and protein fluorescence. Not only water molecules can participate in proton and electron transfer, but they can form organized water wires of several molecules. Electron transfer can be promoted by water chains of up to 10 Å in length (Francisco et al., 2004) while a water wire of 23 Å in *Rhodobacter sphaeroides* has been suggested to act as a proton pathway (Ermler et al., 1994). Finally, hydration water has an essential role in the properties of nucleic acids. Without water and its high

dielectric permittivity to shield charges, the DNA would not maintain the nice double helix that we know. Moreover some researchers (see for instance (Fuxreiter et al., 2005)) have proposed that a given DNA sequence has a "hydration fingerprint" that determines its binding properties.

1.1.3 Hydration water

After reporting the role of water in some biological processes, let us now focus on how water is influenced by the presence of biomolecules. At a molecular level, water and biomolecules interact with one another, mainly by forming hydrogen bonds, which disturb their structure and dynamics. The term hydration water refers to water that features dynamical or structural properties different from those of bulk water because of its proximity to a biomolecule. It is generally considered that water molecules are slowed down and adopt a structure distinct from bulk water within at most the two first water layers around proteins. The first hydration shell was shown by simulation, x-ray and neutron scattering to be 10-15% more dense than bulk water (Svergun et al., 1998; Seki et al., 2002; Merzel and Smith, 2002). Some water molecules, sometimes refered as "immobile" or crystallographic water, are slowed down by a minimum of 3 orders of magnitude (residence time superior to 1 ns), while others are almost not affected. However, there is no consensus on the average water slowdown at the protein surface. The retardation ranges in the literature from a mere twofold to several orders of magnitude (see (Halle, 2004) for a critical review of these values).

In addition to the individual-molecule dynamics mentioned above, the collective dynamics of water is influenced in the vicinity of proteins. In particular, recent experimental and theoretical research has shown that protein surface motions are correlated with water motions up to 10-20 Å away from the protein (Heyden and Tobias, 2013; Meister et al., 2013). These correlations are THz vibrations that propagate as sound wave through the water hydrogen bond network. In other words, the vibrational properties of water hydrogen bond network are modified in the vicinity of biomolecules.

1.1.4 Water and entropy

Because of its interaction with proteins, mainly through the formation of hydrogen bonds, hydration water has a lower entropy than bulk water. By modulating the water-protein hydrogen bond network, the protein can change the entropy of its hydration water. In the case where the hydration water entropy increases, this thermodynamics fuel can be used by the protein. This is known as entropy compensation, which can be either an entropy-entropy or an entropy-enthalpy compensation, depending on whether it compensates for an entropic or enthalpic cost. Thus, the notion of water being a reservoir of entropy for biomolecules has emerged. Using the terminology of Erwin Schrodinger in his book "what is life?", one can say that macromolecules feed upon the entropy of their hydration water. Several experimental and computational studies have quantified this effect in different

systems for protein aggregation (Thirumalai et al., 2012), protein folding (Harano and Kinoshita, 2004) and ligand binding (Breiten et al., 2013).

1.2 Protein dynamics

The discovery of protein tertiary structure (by Dorothy Wrinch in the late 1930's) had led the scientific community to think, for some time at least, that the function of a protein is strictly defined by its 3D structure. This picture was backed up by the lock-and-key theory first postulated in 1894 by Emil Fischer. It claims that the active side of a molecule, let's says an enzyme, has a unique geometric shape that is complementary to the geometric shape of the substrate molecule, similar to the fit of puzzle pieces. This structure-function relationship went along with the tremendous development of protein crystallography, made possible by new technologies developed in the second half of the twentieth century. This structure-function paradigm has now evolved to include a third component, the dynamics. As well as particular structural features, a protein requires a proper dynamics, or conformational motions, to be functional.

1.2.1 Energy landscape

The free energy landscape of a protein is a mapping of all possible conformations as a function of their free energy. In other words, the knowledge of this landscape gives the energetically most favorable conformations. The energy landscape theory is often used to describe protein folding and stability, as thoroughly reviewed by Plotkin and Onuchic (Plotkin and Onuchic, 2002a,b), as well as protein misfolding and aggregation (see figure 1.2 and 1.7). In their normal condition, globular proteins possess a global minima corresponding to their well-folded states (figure 1.2), while intrinsically disordered proteins (IDPs) lacks of such minima in their monomeric state (see the introduction to IDPs in section 1.3). Hans Frauenfelder and coworkers introduced the concept of conformational substates when studying the rebinding kinetics of carbon monoxide to myoglobin after photodissociation (Austin et al., 1975). This concept states that even globular proteins have no unique conformation, but rather multiple conformations that are energetically very close. In terms of energy landscape, this means that within a global energy minimum corresponding to a well-folded state, one finds many accessible substates (see figure 1.2), as if the ground state of a protein was highly degenerated. Transitions from one minimum to another correspond to dynamical changes in the structure of the protein that take it from one substate to another. The timescale of the conformational switching can range from the picosecond (ps), for energetically-close substates, to the second, for conformational changes involving ligand binding for instance. The relation between fast and slow conformational changes is nicely discussed by Henzler-Wildman and Kern (Henzler-Wildman and Kern, 2007). Although the concept of conformational substates was introduced in the mid-seventies, it is worth noticing that already in 1959 Linderstrom-Lang and Schellman stated

"...a protein cannot be said to have "a" secondary structure but exists mainly as a group of structures not too different from one another in free energy, but frequently differing considerably in energy and entropy. In fact the molecule must be conceived as trying out every possible structure..." (Linderstrom-Lang and Schellman, 1959).
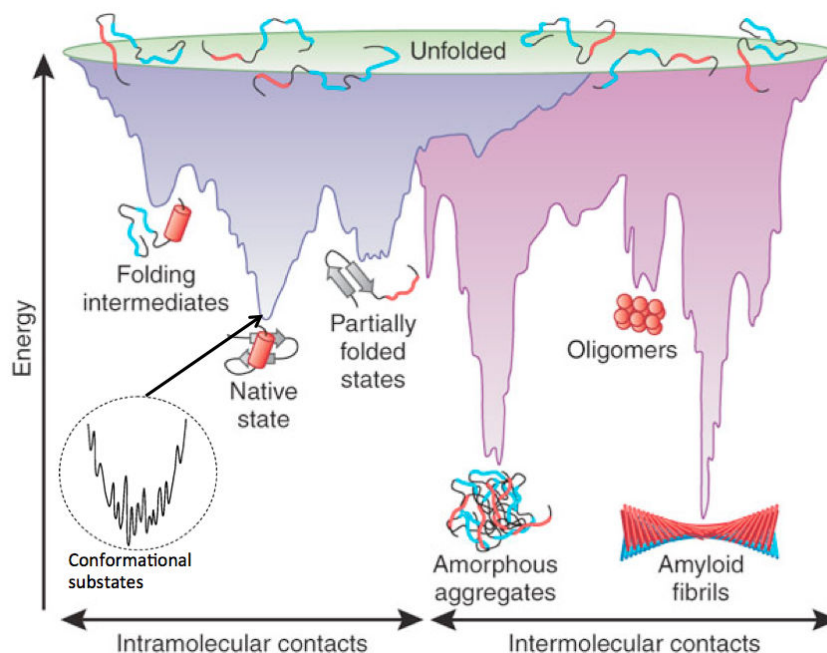
Figure 1.2: Energy landscape of globular proteins. The protein energy landscape is represented by the free energy of the protein as a function of some reaction coordinate. Transiently unfolded proteins are in a higher state of energy. When exploring their energy landscape proteins transit spontaneously through different folding states, before eventually ending up in their native state, which is the lowest energy state. This phenomenon is known as the folding funnel. Within the minimum of the native state, a multitude of substates, known as conformational substates, are constantly explored. In conditions where inter-protein interactions are possible, a new domain of the landscape opens up with new minima, including oligomers, amorphous and amyloid aggregates. This domain is further discussed in section 1.3.4. Figure adapted from (Jahn and Radford, 2005).

The energy landscape of a protein very much depends on its environment, especially on its hydration state. Many experiments, particularly using neutron scattering, have shown that protein dynamics is significantly reduced in the absence of water (Rupley and Careri, 1991). Therefore, the picture of hydration water acting on the protein as a plasticizer has emerged (Doster, 2008). Fenimore and colleagues (Fenimore et al., 2004) have used the following formulation : "The large number of conformational substates is essential; proteins cannot function without this reservoir of entropy, which resides mainly in the

6

hydration shell". It is worth noticing that this view has been recently challenged by a solvent-free polymer-protein hybrid that was found to preserve the function and dynamics of the protein in the absence of water (see (Perriman et al., 2010) and (Gallat et al., 2012a), which is included in appendix A).

1.2.2 Dynamics-function relationship in proteins

The loss of biological function when hindering protein motions, either by lowering the temperature or by removing hydration water (Rasmussen et al., 1992; Zaccai, 1987; Ferrand et al., 1993; Doster et al., 1993), provided strong evidence for a dynamics-function relationship. Since then, many examples of dynamical processes involved in protein functions have been identified, as reviewed for instance by (Henzler-Wildman and Kern, 2007). Slow dynamics, involving displacements of entire domains, are typically required for macromolecular activity such as in enzyme catalysis or signal transduction. Even conformational changes that occur very rarely (*i.e.* poorly-populated conformational states) have been shown to be essential for protein function, as shown for instance in lysozyme (Mulder et al., 2001) or the protein gpW (Sanchez-Medina et al., 2014). A very good illustration of functional large-amplitude motions is the class of IDPs, which constantly undergo large conformational changes and very often collapse upon binding to their biological partners (see section 1.3). Dynamics on shorter timescales, involving energetically close substates, have also been shown to take part in protein function. For instance the binding of carbon monoxyde to myoglobin requires flexibility of protein side chains (Schotte et al., 2003). More precisely, the concept of proteinquake proposed in (Ansari et al., 1985) illustrates that local allosteric changes are collectively triggered to allow the release of a ligand. Another example of functional dynamics is at play in photoactivable proteins that undergo global conformational rearrangements after their chromophore have absorbed photons (Lukyanov et al., 2005). A recent study on a photoactivable flavoprotein showed that the interruption of the proteinquake by mutageneis short-circuits the protein function (Brust et al., 2013). The notion of a proteinquake is a nice concept that illustrates how protein dynamics on a very broad range of timescales (from fs to second) can be correlated to enable biological function.

1.2.3 The dynamical transition

In the framework of the dynamics-function relationship, protein dynamics have been assessed over a wide temperature range, from a few kelvins, up to 400 K. A typical profile of temperature-dependent protein dynamic on the ns time scale is shown in figure 1.3. A first onset of anharmonic dynamics is observed around 100-150 K followed by a drastic increase around 220-240 K, the so-called dynamical transition (Doster et al., 1989). The first increase has been assigned to the rotation of methyl groups entering the instrumental time window (Roh et al., 2005; Wood et al., 2010; Krishnan et al., 2008; Schiro et al., 2010b,a), while the dynamical transition corresponds to the onset of large amplitude motions. This transition has been studied in particular by incoherent neutron scattering for

a large variety of biomolecules, including globular proteins (Wood et al., 2008), membrane proteins (Wood et al., 2007), intrinsically disordered proteins (Gallat et al., 2012b), entire cell proteomes (Tehei et al., 2004), amino acids (Schiro et al., 2011), nucleic acids (Caliskan et al., 2006) and sugars (Jasnin et al., 2010). It has been proposed that the large amplitude motions responsible for the dynamical transition are necessary for the protein biological activity (Rasmussen et al., 1992; Ferrand et al., 1993; Ostermann et al., 2000; Pieper et al., 2007), although some counter-examples have been reported (Daniel et al., 1998). In addition, the dynamical transition is suppressed in the absence of water (Roh et al., 2006), consistent with the picture outlined above of water promoting protein dynamics and allowing macromolecular biological activity. The role of water in this transition is discussed in detail in chapter 3. The origin of the dynamical transition remains controversial as evoked in chapters 3 and 7. Briefly, there are two main explanations of the observed dynamical onset at around 220-240 K (Becker et al., 2004). The first hypothesis is that the protein responds to a change of dynamics occurring in the solvent at the temperature of the dynamical transition. In particular, the confined hydration water would undergo a fragile-to-strong crossover originating from the transition between a high density liquid and a low density liquid phase (Chen et al., 2006; Lagi et al., 2008). This hydration water transition would correspond to the crossing of the so-called Widom line (Xu et al., 2005) at atmospheric pressure. On the other hand, the dynamical transition has been interpreted as an effect purely originating from the instrumental resolution (Khodadadi et al., 2008; Doster et al., 2010). The transition seen at 220-240 K would correspond to the temperature at which thermally activated protein motions enter the space-time window of the spectrometer.

Figure 1.3: Mean square displacements (MSDs) of a dry and a hydrated protein measured by neutron scattering (Gallat et al., 2012b). The MSDs presented here are representative of the average protein dynamics on the ns timescale. The onset of dynamics at around 140 K originates from the rotation of methyl groups entering the instrumental time window and does not depend on the hydration level. The so-called dynamical transition is visible at around 240 K in hydrated proteins and is suppressed in the absence of water.

1.3 Intrinsically disordered protein (IDPs)

For most of the twentieth century, biologists had considered that every proteins had to have a well-defined structure. The primary origin of this paradigm was the 'lock and key' hypothesis, formulated in 1894 by Emil Fischer to explain the astonishing specificity of the enzymatic hydrolysis of glucoside multimers by different types of similar enzymes. Although the notion of disordered proteins emerged in the 1990's, it is not before the beginning of this millennium that they became popular among the scientific community and thus intensively studied. A breakthrough in this field was to be expected when looking at the situation in the early 2000's:

- The biology community realized that disordered proteins were very abundant in the proteome (in particular due to the work of Dunker and coworkers (Dunker et al., 2000)),

- IDPs were found to fulfill a large variety of biological functions,

- many diseases were found to involve natively disordered proteins,

- yet, this new class of protein was almost uncharacterized.

Therefore, it is no surprise that research efforts on IDPs have grown exponentially in the past decade, as shown in figure 1.4.
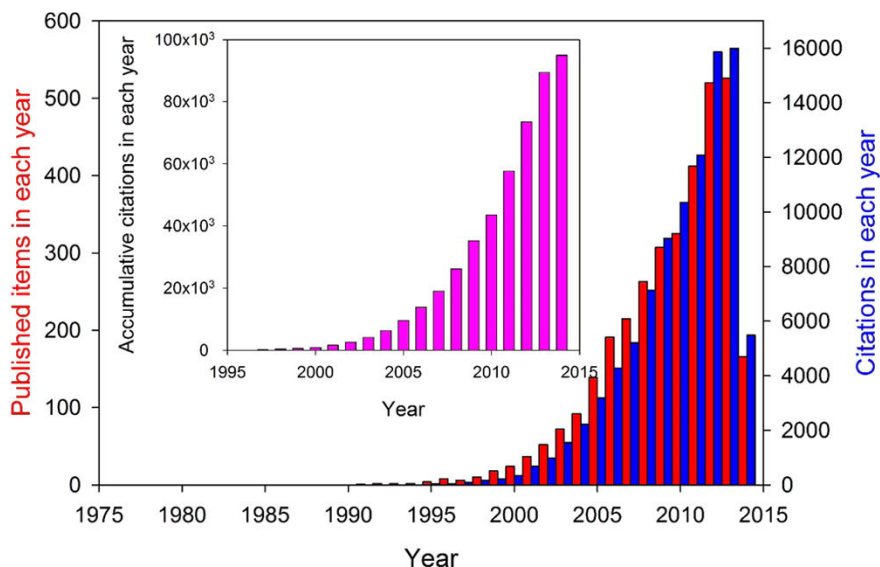
Figure 1.4: Growing interest of researchers in intrinsically disordered proteins. Number of publications (red bars) and corresponding citations (blue bars) related to IDPs by year, from 1991 to the beginning of 2014. The inset represents accumulative citations in each year. Figure extracted from (Uversky, 2014)

1.3.1 Generalities

Definition There is no exact definition of intrinsically disordered proteins (IDPs), because no strict delimitation between disordered and ordered classes proteins exists. Proteins rather fall onto a structural continuum, from tightly folded single domains, to multi-domain proteins that might have flexible or disordered regions, to compact but disordered entities and, finally, to highly extended unstructured states (Dyson and Wright, 2005). A general definition of IDPs would be any protein that exists in different structures, either at the secondary or tertiary level. This lack of precise definition is reflected in the variety of names that can be found in the literature. A list of these terms would include floppy, pliable, rheomorphic, flexible, mobile, partially folded, natively denatured, natively unfolded, natively disordered, intrinsically unstructured, intrinsically denatured, intrinsically unfolded, intrinsically disordered, vulnerable, chameleon, malleable, 4D, protein clouds, dancing proteins, proteins waiting for partners, 3^2 proteins (Uversky, 2014). Several bioinformatic studies have used an arbitrary cutoff of 30 to 50 consecutive disordered amino

acids to define the class of intrinsically disordered proteins. The notion of unfoldome and unfoldomics has been recently introduced to refer to the realm of IDPs and its field of study (Uversky et al., 2009). Rather than with a unique conformation, as one can be used to see in the Protein DataBank (PDB), the "structure" of an IDP is described by an ensemble of conformations. The conformations composing this ensemble are selected to be the smallest group of conformations that is able to describe some measured observable, such as small-angle scattering (SAS) spectra (Bernado et al., 2007). Thus, there is not a unique ensemble describing an IDP. The combination of nuclear magnetic resonance (NMR) spectroscopy and SAS has proven to be a valuable tool to provide structural characterizations of IDPs (Sibille and Bernadó, 2012).

IDP classification   Despite the blurred line between ordered and disordered proteins, a classification was proposed by Dunker, Dyson, and coworkers (Dunker et al., 2001; Dyson and Wright, 2005), as shown in figure 1.5 extracted from (Uversky and Dunker, 2010). Briefly, four different states of disorder are proposed:

- The entirely disordered proteins that possess neither tertiary nor secondary (or almost not) structures and are essentially composed of extended random coils. They are almost like polymers.

- The pre-molten globules, which have some residual elements of secondary structure and are more compact that pure random coil proteins.

- The molten globules, which are disordered but compact. The proteins in this intermediate state have no (or only a trace of) rigid tertiary structure. However, they are characterized not only by a well-developed secondary structure, but also by the presence of some topology, i.e., relatively fixed mutual positioning of the secondary structure elements.

- The ordered proteins, where no or only a few residues are disordered.

Collapsed globules is another term sometimes found in the literature and that can be identified to molten or pre-molten globules.

Natural occurrence   The occurrence of intrinsically disordered protein in the living kingdom is strikingly high. A bioinformatics analysis from Dunker and coworkers (Dunker et al., 2000) estimated that:

- 9-37% of proteins from archea

- 6-33% of proteins from bacteria

- 35-51% of proteins from eukaryotes

contain disordered segments longer than 40 amino acids (see table 1.6). Moreover 3-18% of proteins in eukaryotic organisms were predicted to be entirely disordered.
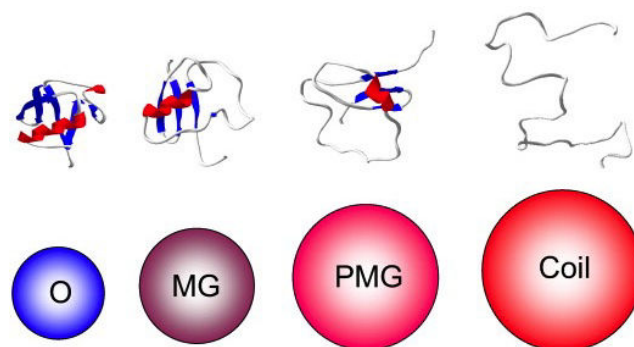
Figure 1.5: Illustratition of a classification of order among proteins. Top line: Order (O), collapsed (molten globule-like, MG) disorder; extended (pre-molten globule-like, PMG) disorder; and (coil-like, coil) disorder. The figures represent model structures of a 100 residue-long polypeptide chain. Middle line: Relative hydrodynamic volumes occupied by a 100 residue-long polypeptide chain in these four conformations. The figure is extracted from (Uversky and Dunker, 2010).

1.3.2 Biological functions and advantages of IDPs

IDPs are ubiquitous among the proteome and, therefore, it is interesting to investigate what are the advantages of being unfolded in the biomolecular world. This topic was nicely reviewed by H. Jane Dyson in (Dyson, 2011) and is briefly evoked here :

- **Promiscuous activity :** One feature that comes to mind is the possibility that a disordered domain could bind in different conformations to different partners. One example is the carboxyl-terminal activation domain (CAD) of the hypoxia-inducible factor α (Hif-1α) domain. In complex with the CREB-binding protein, the CAD of Hif-1α takes up a largely helical configuration, but when bound to the enzyme that catalyzes the hydroxylation of Hif-1α, this same sequence is present as an extended structure.

- **Enhanced specificity :** IDPs that fold upon binding can have very complex interaction with their partner, thus making the interaction highly specific (Sugase et al., 2007). Several examples have been observed for the interaction between transcription factors and DNA where some disordered part of the transcription factors participate in high-affinity binding.

- **Complex connection :** Many central functions of the cell, such as gene transcription, require the assembly of multi-domain complexes. Flexible linkers are essential to bind all the domains together despite their different architectures.

- **Post-translational modifications :** After translation, the sites of phosphorylation, methylation, hydroxylation, etc., remain accessible, thus allowing for facilitated post-translational modifications.

Kingdom	Species	# seqs	L ≥ 30		L ≥ 40		L ≥ 50		CDF*	
Archaea	*Methanococcus jannaschii*	1714	367	21%	155	9%	71	4%	26	2%
Archaea	*Pyrococcus horikoshii*	2062	660	32%	330	16%	164	8%	70	3%
Archaea	*Pyrococcus abyssi*	1764	641	9%	338	19%	157	9%	62	4%
Archaea	*Archaeoglobus fulgidus*	2402	867	36%	492	20%	244	10%	93	4%
Archaea	*Methanobacterium thermoautotrophicum*	1869	971	52%	643	34%	365	20%	140	7%
Archaea	*Halobacterium sp. NRC-1*	2057	1096	53%	724	35%	484	24%	233	11%
Archaea	*Aeropyrum pernix K1*	2694	1547	57%	1010	37%	637	24%	490	18%
Bacteria	*Ureaplasma urealyticum*	611	87	14%	44	7%	14	2%	9	1%
Bacteria	*Rickettsia prowazekii*	834	129	15%	54	6%	23	3%	5	1%
Bacteria	*Borrelia burgdorferi*	845	110	13%	57	7%	26	3%	14	2%
Bacteria	*Campylobacter jejuni*	2309	328	14%	148	6%	80	3%	21	1%
Bacteria	*Mycoplasma genitalium*	480	77	16%	39	8%	20	4%	10	2%
Bacteria	*Helicobacter pylori*	1532	280	18%	140	9%	69	5%	24	2%
Bacteria	*Aquifex aeolicus*	1522	482	32%	234	15%	94	6%	29	2%
Bacteria	*Haemophilus influenzae*	1708	456	27%	227	13%	126	7%	27	2%
Bacteria	*Bacillus subtilis*	4093	1214	30%	622	15%	323	8%	87	2%
Bacteria	*Escherichia coli*	4281	1396	33%	731	17%	363	8%	107	2%
Bacteria	*Vibrio cholerae*	3815	1160	30%	595	16%	333	9%	93	2%
Bacteria	*Mycoplasma pneumoniae*	675	160	24%	95	14%	60	9%	14	2%
Bacteria	*Xylella fastidiosa*	2761	858	31%	463	17%	246	9%	103	4%
Bacteria	*Thermotoga maritima*	1842	670	36%	340	18%	165	9%	53	3%
Bacteria	*Neisseria meningitidis MC58*	2015	653	32%	351	17%	190	9%	64	3%
Bacteria	*Chlamydia pneumoniae*	1052	351	33%	185	18%	100	10%	40	4%
Bacteria	*Synechocystis sp*	3167	1106	35%	624	20%	338	11%	104	3%
Bacteria	*Chlamydia trachomatis*	894	314	35%	169	19%	99	11%	42	5%
Bacteria	*Treponema pallidum*	1028	392	38%	222	22%	115	11%	37	4%
Bacteria	*Pseudomonas aeruginosa*	5562	2314	42%	1310	24%	702	13%	183	3%
Bacteria	*Mycobacterium tuberculosis*	3916	2004	51%	1219	31%	747	19%	293	7%
Bacteria	*Deinococcus radiodurans chr 1*	2580	1335	52%	864	33%	534	21%	212	8%
Eukaryota	*Plasmodium falciparum chr II, III*	422	203	48%	147	35%	107	25%	11	3%
Eukaryota	*Caenorhabditis elegans*	17049	8304	49%	6156	36%	4636	27%	1322	8%
Eukaryota	*Arabodiopsis thaliana*	7849	4465	57%	3206	41%	2248	29%	653	8%
Eukaryota	*Saccharomyces cerevisiae*	6264	3373	54%	2527	40%	1858	30%	356	6%
Eukaryota	*Drosophila melanogaster*	13885	8771	63%	7031	51%	5651	41%	2403	17%

Figure 1.6: Prediction of disorder from 34 different genomes. *L* represents the number of consecutive disordered amino acids. The right column represents the predicted percentage of entirely disordered proteins according to a cumulative distribution function analysis. The figure was extracted from (Dunker et al., 2000).

- **Regulation by proteolysis :** The abundance of IDPs in the cell, which may precipitate into pathological states if perturbed, has been shown to be tightly regulated by several cellular processes, including mRNA transcript clearance, translational rate and proteolytic clearance. IDPs can undergo ubiquitin-independent degradation that allow for rapid clearance of potentially harmful proteins, without the necessity for specific intact ubiquitination sites, which could potentially be inactivated by mutation.

- **Higher capture radius :** The concept of "fly-casting" suggests that disorder in a polypeptide chain enhances the capture radius of the chain, with consequences on the kinetics of complex formation.

- **Bulk physicochemical effects :** Since the solvent-accessible surface of a disordered ensemble is greater than that of a globular folded protein of the same size, it might be expected to have a proportionally greater effect on the bulk physical chemistry of

the solution. This effect is partially discussed in this manuscript through the coupling of the tau protein with its hydration water.

The functions carried out by IDPs are manifold and take part in the regulation of key cellular processes, such as transcription, translation, signal transduction and the cell cycle. Using the Swiss-Prot database, a thorough statistical analysis of order- and disorder-associated functions showed that the latter cover essentially all functional categories (Xie et al., 2007).

Nevertheless, it has been proposed that disordered proteins functions fall into five main categories (Tompa, 2002). The first one is that of entropic chains, with functions that directly stem from the flexibility, pliability and plasticity of the protein. They rely entirely on an extended random coil conformation that has to remain in constant motion during functioning. The second class, the effectors, modify the activity of a single partner protein or assembled proteins. The third class is that of scavengers, which store and/or neutralize small ligands. The fourth class are the assemblers, which assemble, stabilize and regulate large multi-protein complexes, such as the ribosome, cytoskeleton, transcription preinitiation complex. The fifth class, the display sites, mediate regulatory post-translational modifications, such as phosphorylation or limited proteolysis. Such modifications often require intrinsic disorder, which enables high specificity/low affinity interactions with the active site of the modified enzyme. This classification is summarized and illustrated with examples in the table 1.1 extracted from (Tompa, 2002).

1.3.3 IDPs and diseases

It has been shown that IDPs are assiociated with a large range of diseases, including inherited and acquired maladies. The connection between disorder and diseases such as cancer, cardiovascular diseases, amyloidoses, neuro-degenerative diseases and diabetes has been established (Uversky et al., 2009). A detailed review of the different diseases related to IDPs, as well as a discussion on what makes IDPs so common in human diseases was written by Uversky and coworkers (Uversky et al., 2008). The authors concluded that these diseases, which have been loosely grouped as "conformational diseases", are characterized not only by protein misfolding but also by failures of post-translational modification and inability for proteins to interact correctly with their physiological partners. The largest group of conformational diseases, including numerous neurodegenerative disorders, originates from the conversion of specific proteins from their soluble functional states into stable, highly ordered, filamentous protein aggregates, known as amyloid fibrils, and from the deposition of these aggregated materials into a variety of organs and tissues (Chiti and Dobson, 2006).

1.3.4 Protein aggregation and IDPs

IDPs' predisposition in the aggregation pathway Although there is no consensus on a unique protein aggregation pathway, a general process could be the following (Ross

IDP	Target/partner	Function/action
Entropic chains		
MAP2	Not applicable	Entropic bristle (spacing in microtubule architecture)
Titin	Not applicable	Entropic spring (passive contractile force in muscle)
Effectors		
Calpastatin	Ca^{2+}-activated protease (calpain)	Inhibitor of calpain
Stathmin	Tubulin dimers	Microtubule disassembly
Scavengers		
Caseins	Calcium phosphate	Nanocluster formation, inhibition of precipitation in milk
Desiccation stress protein 16	Water	Retention of water to prevent desiccation of plants
Assemblers		
MAP2	Tubulin dimers	Microtubule polymerization, bundling
Tau	Tubulin	Microtubule polymerization, bundling
Caldesmon	Ca^{2+}/calmodulin, F-actin, myosin, tropomyosin	Actin polymerization, bundling
Display sites		
MAP2	Protein kinases	Regulation by phosphorylation
Tau	Protein kinases	Regulation by phosphorylation

Table 1.1: Functional classification of IDPs.

and Poirier, 2004). First the protein reaches a pathological partially unfolded (or partially folded) state. This step can be triggered by several factors including covalent modifications (such as phosphorylation, partial cleavage, etc...), and environmental changes (such as pH, temperature, concentration of osmolyte, etc...). This structurally altered state, prone to self-assemble, forms intermediate oligomers, before eventually ending up in stable aggregates. The reason why IDPs eagerly aggregate was proposed to be found in the first step of this pathway (Uversky and Fink, 2004). Because of their unfolded nature and their high conformational flexibility, IDPs can readily switch to the partially unfolded states required to trigger the aggregation. This is in contrast to stable gobular proteins that need to undergo more significant conformational changes to reach the same partially unfolded state. The predisposition of IDPs for aggregation is illustrated in the protein energy landscape scheme shown in figure 1.7.

Amyloid and amorphous aggregates   Protein aggregates fall into two structural categories : amorphous aggregates and structured, β-rich amyloid aggregates. The latter have attracted a particular attention because of their implication in many diseases. They are structurally characterized by intermolecular β-sheets, so-called cross-β, and their presence
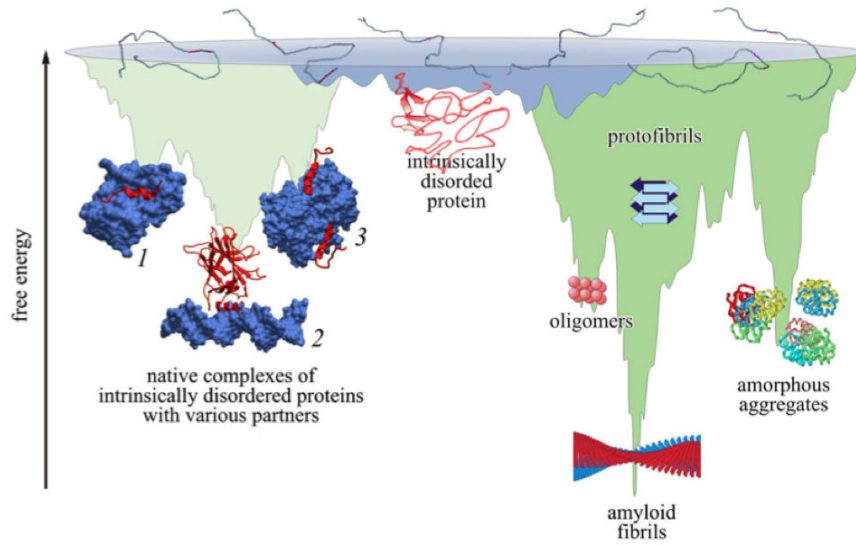
Figure 1.7: The protein energy landscape is represented by the free energy of the protein as a function of some reaction coordinate. When folding upon binding their partners, IDPs (center of the scheme) fall into lower states of energy. The propensity of IDPs to interact with various partners (labels 1, 2 and 3) determines their biological functions in recognition of various binding partners (ligands, nucleic acids and other proteins), in regulation of almost all cellular processes, and in signal transduction. In contrast to the folded globular proteins that have to unfold in order to aggregate (see figure 1.2), IDPs seem to be always ready for such intermolecular interactions, as illustrated by the small energy barrier separating them from stable aggregates. Amyloid aggregates, as the most stable form of aggregates, possess the lowest free energy. The figure is adapted from (Turoverov et al., 2010).

is often demonstrated by powder diffraction, featuring an inter β-sheet distance of 4.7 Å (Astbury et al., 1935; Sunde et al., 1997). A schematic view of an amyloid structure is shown in figure 1.8. Amyloid aggregates are thought to be the most stable form that a protein can adopt (Baldwin et al., 2011), as it is illustrated in its energy landscape in figure 1.7. In this view, the native state of a protein (unfolded or globular) would be just a metastable state. It is worthy to note that although amyloids are found in many diseases, the so-called functional amyloids are amyloid aggregates that play a biological role (Turoverov et al., 2010). An example is the storage of protein hormones in some secretary cells under the form of high density protein clusters, so-called granules. These granules, which were recently found to be amyloid aggregates (Maji et al., 2009), allow storing a large quantity of hormone proteins until a signal triggers their release, at which point the cell can secrete hormones at a fast rate. The main advantage of this process is that the rate of hormones secretion is no longer limited by the rate of peptide synthesis.
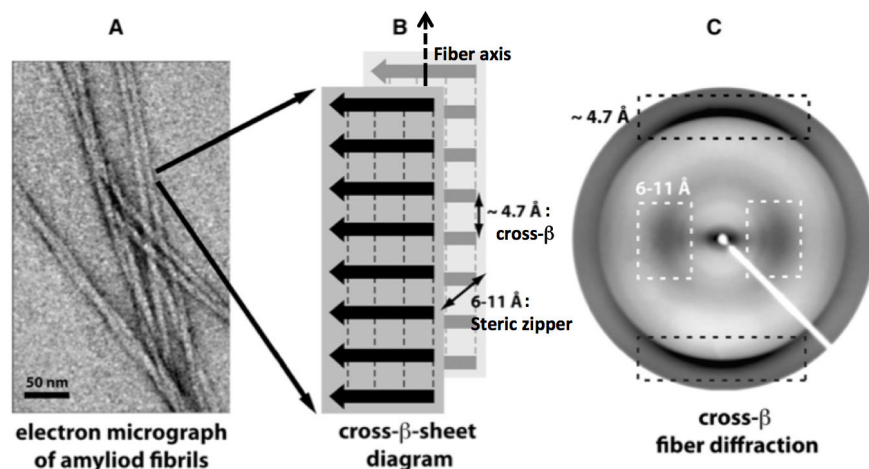
Figure 1.8: (A) Amyloid aggregates are fibrillar-like and are visible in negatively stained transmission electron micrographs. (B) Schematic diagram of the amyloid structure : the intermolecular β sheets along the fiber axis, so-called cross-β sheets, are spaced by 4.7 Å. The so-called steric zippers stabilized two parallel β sheets, separated by 6 to 11 Å perpendicularly to the fiber axis. (C) The diffraction pattern reflecting the characteristic distances of the cross-β sheets (black dashes) and the steric zippers (white dashes). Figure modified from (Greenwald and Riek, 2010).

1.4 The tau protein

1.4.1 Generalities

The tau protein, essentially present in neurons, is a protein that modulates the stability of axonal microtubules. Thereby, It belongs to the class of microtubule associated proteins (MAPs). Tau is absent in dendrites and is active primarily in the distal portions of axons where it provides microtubule stabilization by binding to them. The microtubule-binding domain is composed of three or four repeat domains, as well as the proline-rich flanking domains on each side of the repeat domains (Amos, 2004). The N-terminal fragment does not bind to microtubules, but projects away from their surface and hence is termed "projection domain". Tau is thought to have other marginal binding partners, whose role is not clearly understood (Mandelkow and Mandelkow, 2012). Although the tau protein is essentially disordered, a few transient secondary structure elements have been identified (see figure 1.9).
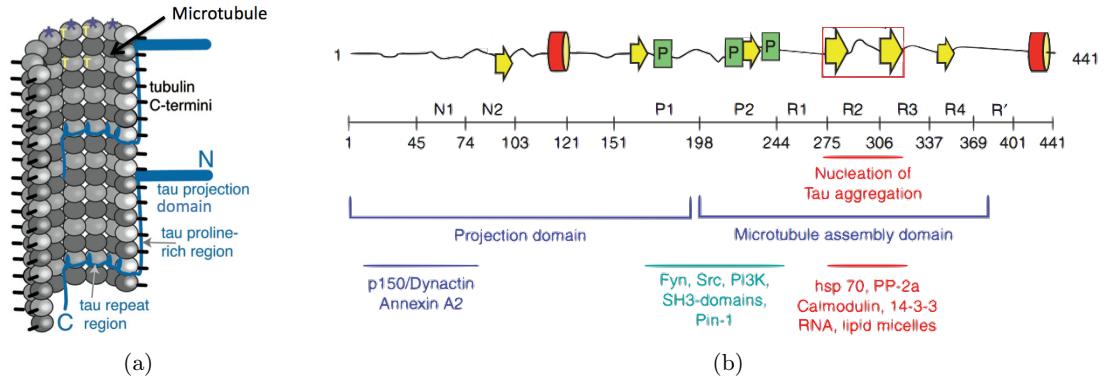
Figure 1.9: (a) Schematic view of tau bound to a microtubule, adapted from (Amos, 2004). (b) Top : structural elements in htau40. Most of the protein is unfolded (black lines), with a few short and transient elements of secondary structures (α-helix in red, β-strand in yellow, poly-proline helix in green). The red frame indicates the region of the two hexapeptide motifs ($^{275}VQIINK^{280}$ and $^{306}VQIVYK^{311}$) essential for tau fibrillation. Middle : domain subdivision. R1-R4 represent the repeat domains, which essentially coincide with the microtubule interaction domain. Bottom: approximate location of interaction sites with other proteins. Figure adapted from (Mandelkow and Mandelkow, 2012).

There are mainly six isoforms of tau expressed in the adult human brain, all of which are derived from a single gene by alternative splicing. They differ from each other in the number of repeat domain (either three or four) and in the presence or absence of either one or two 29 amino-acid-long inserts at the N-terminal portion of the protein. The longest isoform, which is studied in this work, is refered to as htau40 and possesses 441 amino acids. It is composed of four repeat domains and has an average radius of gyration of 62 Å in solution, being thereby considerably more extended than a globular protein of the same molecular weight (Gallat et al., 2012b). Although the six isoforms appear to be broadly functionally similar, each is likely to have precise, and to some extent distinctive, roles. The tau protein is overall positively charged and possesses a high heterogeneity of charge distribution along its surface that is important for its interaction with other proteins as well as for aggregation. The repeat domains bind to specific pockets in β-tubulin at the inner surface of the microtubules, whereas the positively charged proline-rich regions are tightly bound to the negatively charged microtubule-surface, and the negatively charged projection domain branches away from the microtubule surface (see figure 1.9a). When tau is bound to a microtubule it stabilizes the structure and promotes its growth. It is interesting to note that tau is in a constant dynamic binding equilibrium, on and off the microtubules, constantly provoking polymerization and depolymerization. Each tau protein remains on average 4 seconds on a macrotubule surface. The main mechanism that modulates the affinity of tau for microtubule is its phosphorilation state, which is regulated by kinase and phosphatase enzymes. Interestingly, as an IDP, tau has been identified to bind marginally to many other partners such as actin filaments or nucleic acids. However the biological importance of the alternative binding to other partners is hardly known.

1.4.2 Dynamical coupling with hydration water

As an IDP, the tau protein lacks a tertiary structure and thus possesses a solvent-accessible surface greater than a globular protein of a similar molecular weight. It implies a larger protein-water hydrogen bond network and, therefore, one might expect the coupling between the tau protein and its hydration water to differ from that of globular proteins.

Previous research carried out in our laboratory has addressed the question of the dynamical coupling between the tau protein and its hydration water (Gallat et al., 2012b), and compared it to that of a globular protein of the same molecular weight, the maltose binding protein (MBP). The MBP is located in the periplasm and binds maltose, being involved in the transport process of maltodextrin. When comparing the dynamics of proteins and hydration water, as a function of temperature, they found that both follow each other up to 250 K for tau, while they separate at 220 K for MBP. François-Xavier Gallat and colleagues concluded that an enhanced coupling exists between the tau protein and its hydration water as compared to the globular protein. They suggested an extension of this conclusion to the entire class of IDPs and proposed a gradient of protein-hydration water coupling across different protein classes. From looser to tighter coupling, the membrane proteins would come first, followed by the globular proteins, and finally the IDPs with the tightest coupling (see figure 1.10).
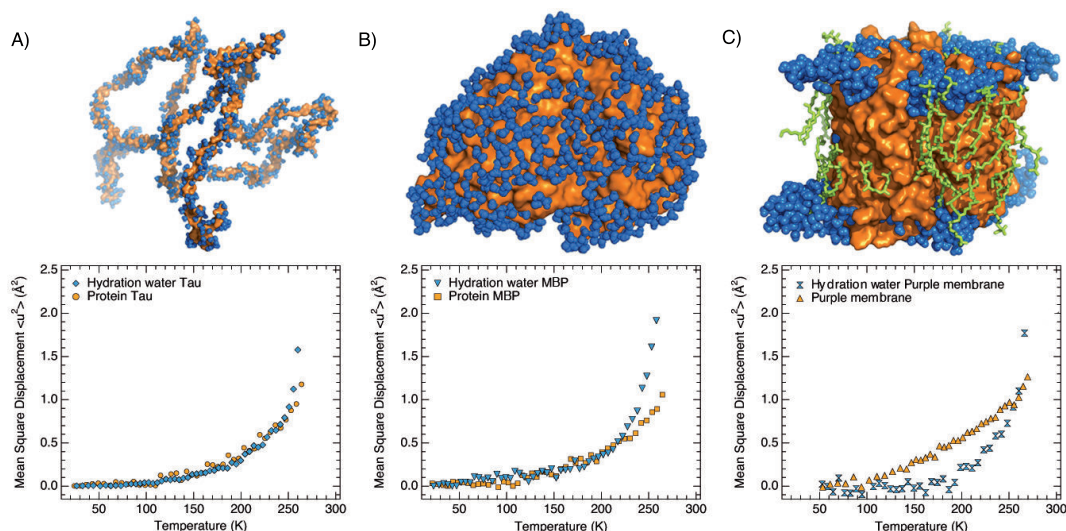
Figure 1.10: Protein-water dynamical coupling for different protein classes. The lower panels show mean square displacements (representative of the dynamics of the considered atoms) of hydration water and proteins. The upper panel represents a schematic view of three protein classes : A) the tau protein as an IDP, B) the maltose binding protein (MBP) as a globular protein and C) the bacteriorhodopsin as a membrane protein. A decreased dynamical coupling between hydration water and proteins is observed from the class A) to C). The figure is extracted from the thesis of François-Xavier Gallat (Gallat, 2011).

1.4.3 Tau amyloid fibers and diseases

Tau has been much more studied than some other MAPs because of its implication in many diseases. Indeed the neurofibrillary tangles (NTFs) found in several neurodegenerative diseases are composed of the tau protein in the form of straight or paired helical filaments (SFs or PHFs, respectively). Diseases involving tau filaments have been grouped under the term of tauopathies and include Alzheimer disease (AD) and the frontotemporal dementias. A list of these diseases is shown in table 1.2.

Alzheimer disease
Amyotrophic lateral sclerosis/parkinsonism-dementia complex
Argyrophilic grain disease
Autosomal-dominant Parkinson's disease
Autosomal-dominant parkinsonism
Autosomal-recessive juvenile parkinsonism
Corticobasal degeneration
Dementia pugilistica
Diffuse neurofibrillary tangles with calcification
Down syndrome
Familial British dementia
Frontotemporal dementia and parkinsonism linked to chromosome 17
Gerstmann-Straussler-Scheinker disease
Guadeloupean parkinsonism
Hallervorden-Spatz disease
Myotonic dystrophy
Niemann-Pick disease, type C
Non-Guamanian motor neuron disease with neurofibrillary tangles
Pick's disease
Postencephalitic parkinsonism
Prion protein cerebral amyloid angiopathy
Progressive subcortical gliosis
Progressive supranuclear palsy
Subacute sclerosing panencephalitis
Tangle only dementia

Table 1.2: Diseases in which tau deposits have been described (tauopathies), extracted from (Goedert et al., 2006).

PHFs were first discribed as twisted fibrillar structures, composed of two β-rich amyloid subfibers. Although it has now been shown that the term "tau fibers" is probably more appropriate than "tau PHFs" (see (Mandelkow and Mandelkow, 2012) for more details), we use indifferently both terms throughout this manuscript as it is common practice in the literature. Although the precise structure of tau PHFs remains unknown, it can be divided into two different regions, a rigid β-rich core (denoted as the fiber core), which is

essentially composed of the four repeat domains, and the remainder, so-called fuzzy coat, which is highly flexible and remains disordered. An essential factor for tau fibrillation is its propensity to form β-structures, encoded in the short hexapeptides $^{275}VQIINK^{280}$ and $^{306}VQIVYK^{311}$, present in the repeat domains (see figure 1.9b). Disruption of these motifs abrogates tau's tendency to aggregate. Another important feature that promotes tau amyloid formation is the charge compensation of the basic middle part of tau by polyanions. *In vitro*, this can be achieved by the presence of sulfated glycosaminoglycans (for instance, heparin was used in this work to trigger fiber formation), which helps to overcome the nucleation barrier.

Very little is known about what causes the tau protein to aggregate. It is likely that a higher cytosolic concentration of unbound tau increases the chances of pathogenic conformational changes that, in turn, lead to the aggregation and fibrillation of tau. However, the potential origins of such an increase are multiple (figure 1.11 illustrate a proposed pathway of tau aggregation). Throughout the tauopathies, the tau proteins included in NFTs are always hyperphosphorylated, which is likely to be related to the detachment of tau from the microtubules. Whether this feature is the direct cause of tau fibrillation or a consequence of an early disruption remains to be elucidated. One step of the fibrillation process that has recently attracted considerable attention is the formation of transient soluble oligomers, also called pretangles. They have been proposed to be the toxic species in the aggregation process, while the mature PHFs would merely be a means of storing the excess of proteins. For further reading about the tau protein and the associated tauopathies, I recommend a review from Ballatore and coworkers published in Nature Reviews Neuroscience in 2007 (Ballatore et al., 2007) and a review written by Eva-Maria and Eckhard Mandelkow in 2012 (Mandelkow and Mandelkow, 2012).
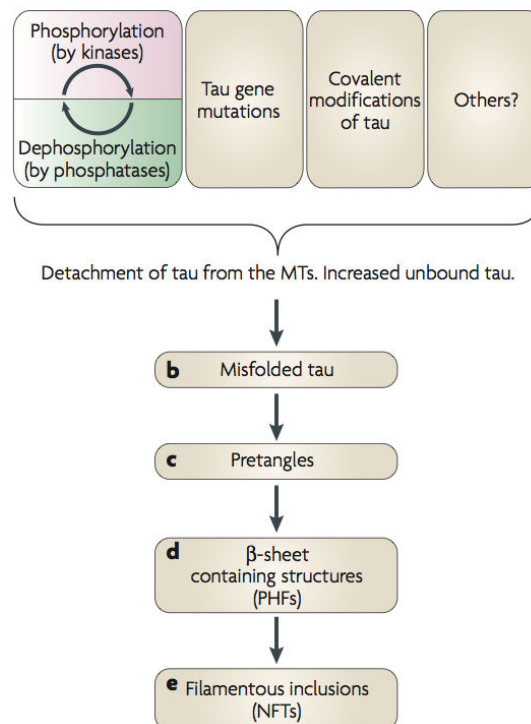
Figure 1.11: Schematic view of a possible aggregation pathway, extracted from (Ballatore et al., 2007). Several factors can cause the detachment of tau from the microtubules, including genetic mutation and the disturbance of the phosphorylation/dephosphorylation equilibrium. These modifications lead to a misfolding of the protein, which will aggregate in pre-tangles (soluble oligomers) before forming the PHFs incorporated in the NFTs.

1.5 Summary of the introduction and issues addressed in this thesis

Water is an essential component in any biological system known today and, for that reason, has been named as the "matrix of life". At the molecular level, water molecules play a key role in a large variety of situations, such as ligand binding, DNA and protein stability, electron and proton transfer, etc. Water and macromolecules, including proteins, are in a constant give-and take relationship. On the one hand, water in the vicinity of macromolecules, so-called hydration water, is perturbed and possesses properties distinct from those of bulk water. On the other hand, the presence of water is essential for proteins to acquire their proper structure and dynamics, which they require to achieve their function. Among the proteome, intrinsically disordered proteins (IDPs) composed a recently-discovered class of proteins that lack a well-defined three dimensional structure. Their "spaghetti-like" structures make their dynamics and their interaction with water quite different from globular proteins. IDPs have attracted particular attention, because their aggregation is involved in a large variety of diseases, including neurodegenerative disorders. The intrinsically disordered protein tau, which is the main object of this thesis, regulates the microtubule activity in neuronal cells. Tau has been particularly studied because its fibrillation into amyloid aggregates is one of the hallmarks of Alzheimer disease.

In this thesis, we intend to reveal whether or not the dynamical properties of protein hydration water are biologically relevant and could be used to further understand and modulate biological processes. In order to do so, we have used incoherent neutron scattering, THz spectroscopy, small angle X-ray scattering, and all-atom molecular dynamics simulations to investigate water dynamics around the intrinsically disordered tau protein, compared to the structured maltose binding protein (MBP), as well as around functional and aggregated proteins.

After introducing in **chapter 2** the biophysical methods used throughout this work, the manuscript presents an attempt to answer the following questions :

Chapter 3 What are the dynamical features of hydration water that allow a protein to be biologically active? The question was addressed by combining quasi-elastic neutron scattering and molecular dynamics (MD) simulations on the IDP tau as well as on the globular MBP.

Chapter 4 Does the collective dynamics of hydration water differ significantly for different types of proteins? This point was addressed by applying THz spectroscopy on tau and MBP.

Chapter 5 What changes occur in the hydration water dynamics when a protein forms amyloid fibers? This issue was investigated by carrying out neutron scattering and MD simulations on the tau protein in its native and amyloid states.

Chapter 6 Is there a dynamical signature of protein aggregates. If yes, is this signature related to the protein-water interactions? We address this question using elastic

neutron scattering on different types of aggregates, including amorphous and amyloid aggregates.

Chapter 7 Can we use MD simulations to model a protein powder of an IDP, and what can one learn from it? This issue was investigated by comparing MD simulations of a IDP tau powder model and an MBP powder model.

Chapter 2

Biophysical methods

This chapter presents the different biophysical methods used in this work. We examine the main features of the different techniques, as well as their main applications. A particular attention is paid to neutron scattering (NS) and all-atom molecular dynamics (MD) simulation, on which relies a large part of this PhD work. These two techniques were used in a complementary manner to evaluate dynamical properties of proteins and water on the picosecond-to-nanosecond timescale. THz spectroscopy is also described as a means of probing collective water dynamics. The descriptions remain general and the theoretical background of specific data analysis procedure will be detailed in the relevant chapters.

2.1 Neutron scattering and molecular dynamics

2.1.1 Introduction

Scattering is a general physical process where waves or particles are deviated after crossing an obstacle. In scattering experiments, the sample constitutes the obstacle and the properties of the scattered beam, such as its angle and energy, reflects the properties of this sample. In order to obtain information at the atomic level, the wavelength of the incoming particle must be on the order of inter-atomic distances, *i.e.* a few Angstrom (Å; 1 Å = 10^{-10} m). This criterium can be fulfilled by several particles such as photons, electrons and neutrons.

Neutron beams possess several advantages over other particle beamss. As neutrons are neutral particles, they interact with atom nuclei. It results in a weak interaction with matter and, therefore, a deep penetration into it. Moreover, it implies that the strength of interaction with an atom, characterized by its cross section σ, is not linear with its atomic number. Especially σ will be sensitive to different isotopes, allowing for isotopic labeling, as detailed in section 2.1.5. Another interesting property lies in the low energy of few-Å-wavelength neutrons, which therefore preserves the samples. That is why neutron scattering is of such a high interest for biological samples, which are particularly sensitive to radiation damages. For instance, a neutron of 1 Å wavelength has an energy of 82 meV, while an X-ray photon of the same wavelength has an energy of 12.4 keV, *i.e.* six orders of magnitude higher. Furthermore, the energy of a neutron associated with a wavelength of atomic distances (Å) is on the order of thermal energy (meV), thus allowing to measure molecular dynamics. In particular, one can study the energy exchange between neutrons and the sample in quasi-elastic and inelastic scattering experiments. Finally, neutrons possess a non-zero spin (either 1/2 or -1/2), which allows to exploit incoherent and coherent scattering, as detailed in section 2.1.4.

The main disadvantage of neutrons over X-ray has to do with their production. Neutron sources are considerably less intense than X-ray sources, requiring therefore a much longer sample exposure time and larger sample amounts.

2.1.2 Neutron sources and instrumentation

Neutron sources

Two types of technologies are used today to produce high-flux neutron sources : nuclear reactors and spallation sources. The first one is based on the chain fission of ^{235}U and produces a continuous flux of neutrons. The second uses an accelerated and pulsed proton beam that hits a heavy metal target. The high-energy protons strike the target nuclei and break them apart, resulting in the emission of high-energy neutrons. Neutrons produced by either spallation or reactor sources are very fast, having energies in the MeV regime. To be useful in condensed matter research, neutrons must have their energy reduced by

many orders of magnitude. Neutrons behave like a gas and are slowed down by thermal equilibrium through successive collisions in a material called moderator. Most of the time, for the energies used in biology, the moderator is light or heavy water.

A few dozens of neutron sources exist worldwide, most of these in Europe and North America. At present, the world-leading sources are the ILL in Grenoble, ISIS in the UK and SNS in Tennessee. An updated list of neutron sources can be found on the NMI3 website (www.nmi3.eu/neutron-research/where.html). It is worth mentioning the project of the european spallation source (ESS) that aims at becoming the most powerful neutron source in the world. The construction started in 2014 and should be completed in 2019, but the installation is foreseen to be fully operational only in 2025. This facility promises to provide performances never reached before.

Instruments

There is a large variety of neutron spectrometers and diffractometers, each of them being optimized to obtain certain precise information. They can be sorted under four different layouts : triple-axis, spin echo, time-of-flight, and backscattering spectrometers. The last type has been mainly used in this work and will be detailed below.

The operational principle of a backscattering spectrometer is the following (see the schematic view in figure 2.1): a neutron beam is brought onto a moving monochromator. For an elastic experiment, the monochromator is kept fixed and the monochromatic beam with an energy E_0 is directed to the sample. In a quasi-elastic experiment, the monochromator oscillates and, by Doppler effect, modulates the energy of the neutrons. The energy range of the outgoing neutron beam is directly related to the oscillation speed. The neutrons are then scattered by the sample, according to its dynamics, at a given angle and with a given energy variation. Neutrons with an energy E_0 are backscattered (scattering at 180°) by the analyzers, sent back through the sample and finally measured by the detectors. In order to discount neutrons that are scattered from the sample directly to the detectors (without being analyzed) a chopper is used to pulse the incoming beam. Because the speed of a neutron depends on its energy, the time of flight between the Doppler and the detector is used to trace back the energy exchange between the sample and each counted neutron. The backscattering condition allows these spectrometers to reach resolutions down to the μeV range, which corresponds to motions on the nanosecond (ns) timescale (see below for more details on the instrument resolution). The accessible range of energy transfer is on the order of tens of μeV.

In this study, two backscattering spectrometers have been used : IN16 (Frick and Gonzalez, 2001) at the Institut Laue-Langevin (ILL) in Grenoble, France, and SPHERES (Wuttke et al., 2012) at the Forschungs-Neutronenquelle Heinz Maier-Leibnitz (MLZ), in Garching, Germany. These two spectrometers have similar characteristics (resolution, momentum transfer range, energy exchange range...). Their main features are reported in table 2.1.
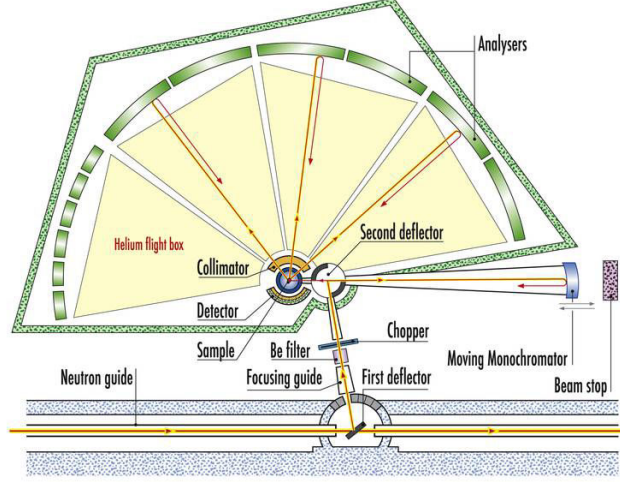
Figure 2.1: Schematic representation of the backscattering spectrometer IN16 at the ILL.
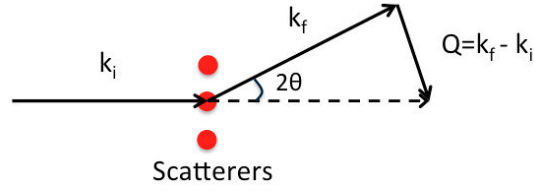
Figure 2.2: Scattering by a single nucleus. The incoming neutron has a wavevector k_i, and k_f after being scattered. The momentum transfer $\mathbf{Q}$ is defined as $\mathbf{Q} = \mathbf{k}_f - \mathbf{k}_i$. Q is the reciprocal quantity corresponding to the scattering angle θ.

	IN16	SPHERES
Wavelength λ_0	6.27 Å	6.27 Å
Resolution (FWHM)	0.9 μeV	0.65μeV
Characteristic times	$\sim$ns	$\sim$ns
Resolution shapes	Gaussian	Gaussian with Lorentzian tail
Q_{max}	1.9 Å^{-1}	1.9 Å^{-1}
Energy range	-15/+15 μeV	-30/+30 μeV

Table 2.1: Wavelength, characteristic time, resolution shape, momentum transfer range and energy range of the backscattering spectrometers IN16 (ILL) and SPHERES (MLZ).

2.1.3 Scattering theory

We introduce the neutron scattering cross section σ, which represents the ability of a system to scatter neutrons. The angular dependence of the scattered neutrons is a most

important aspect in neutron scattering experiments.In order to describe this dependence, we employ the differential scattering cross section:

$$\frac{\partial \sigma}{\partial \Omega} = \frac{1}{\phi} \frac{\text{number of neutrons scattered per second into solid angle } d\Omega}{d\Omega} = b^2 , \qquad (2.1)$$

where ϕ is the incident neutron flux, $d\Omega$ is a solid angle, and b is the scattering length, which represents an effective linear dimension of the scatterer. b is on the order of 10^{-15} m, much smaller than the neutron wavelength, so that the nucleus of an atom can be considered as a point and the scattering is isotropic. The total scattering cross section (usually given in units of barns, where 1 barn is 10^{-24} cm^2) is the integral of the differential scattering cross section in all directions:

$$\sigma_{tot} = \int_{4\pi} \frac{\partial \sigma}{\partial \Omega} d\Omega = 4\pi b^2 \qquad (2.2)$$

The scattering of a neutron being dependent, in general, on the energy exchange occurring during the scattering event, the concept of the differential scattering cross section can be extended to the double-differential cross section $\partial^2 \sigma / \partial \Omega \partial E$:

$$\frac{\partial \sigma}{\partial \Omega} = \int \frac{\partial^2 \sigma}{\partial \Omega \partial E} dE , \qquad (2.3)$$

where E is the energy exchange between the nuclei and the neutron. In practice, E is often replaced by ω, defined as $E = \hbar\omega$, with $\hbar$ the reduced Planck constant.

The wavevectors of a neutron before and after a scattering event are denoted $\mathbf{k}_i$ and $\mathbf{k}_f$, respectively. Their difference is a momentum transfer $\mathbf{Q}$, *i.e.* $\mathbf{Q} = \mathbf{k}_f - \mathbf{k}_i$ (see figure 2.2). Q is the reciprocal quantity corresponding to the scattering angle θ and their relation can be written as $Q = \frac{4\pi}{\lambda} sin(\theta/2)$ if there is no energy exchange with the scatterer. Information on the sample dynamics is contained in the scattering function, also called the dynamic structure factor, $S(\mathbf{Q}, \omega)$. The scattering function is related *via* a spatio-temporal Fourier transform to the positions of the sample scattering nuclei as a function of time. For a system of N identical atoms, this function can be written as a function of the double-differential cross section :

$$\frac{\partial^2 \sigma}{\partial \Omega \partial \omega} = N \frac{k_f}{k_i} b^2 S(\mathbf{Q}, \omega) \qquad (2.4)$$

The probability of measuring scattered neutrons is dependent upon constructive interferences of scattered waves from nuclei located at certain positions in space and time. The scattering function can be written as the Fourier transform in time of the intermediate scattering $I(\mathbf{Q}, t)$ function:

$$S(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} I(\mathbf{Q}, t) e^{-i\omega t} dt \qquad (2.5)$$

with

$$I(\mathbf{Q}, t) = \frac{1}{N} \sum_{m,n} \langle e^{i\mathbf{Q}\mathbf{r}_m(t)} e^{-i\mathbf{Q}\mathbf{r}_n(0)} \rangle \qquad (2.6)$$

where $\langle \rangle$ stands for ensemble averaging. The intermediate scattering function is a correlation function in time and can also be expressed as the Fourier transformation in space of the pair correlation function $g(\mathbf{r}, t)$:

$$I(\mathbf{Q}, t) = \int_{-\infty}^{+\infty} g(\mathbf{r}, t) e^{i\mathbf{Q}\mathbf{r}} d\mathbf{r} \tag{2.7}$$

The correlation function $g(\mathbf{r}, t)$ describes the mutual positions of nuclei in the sample as a function of time, and can be expressed as:

$$g(\mathbf{r}, t) = \frac{1}{N} \sum_{m,n} \langle \delta(\mathbf{r} + \mathbf{r}_m(0) - \mathbf{r}_n(t)) \rangle \tag{2.8}$$

It is important to note that in this formalism, two approaches can be taken. One can measure the dynamic structure factor (or scattering function) and track down to the atomic scale the sample structure and dynamics (top-down approach, going from $S(\mathbf{Q}, \omega)$ to $g(\mathbf{r}, t)$), as it is typically done for any scattering experiment. On the other hand, one can also calculate the scattering function, knowing the atomic distribution of the sample (bottom-up approach, going from $g(\mathbf{r}, t)$ to $S(\mathbf{Q}, \omega)$), as it is commonly done in MD simulations.

2.1.4 Incoherent and coherent scattering

When a neutron of spin $1/2$ encounters a single atom with a nuclear spin I, the spin of the neutron-nucleus system can assume two values $I \pm 1/2$. The scattering lengths of the two systems are denotes b^+ and b^-, and the number of spin states associated with each are $2(I + 1/2) + 1 = 2I + 2$ and $2(I - 1/2) + 1 = 2I$, respectively. The total number of states is $4I + 2$. If the neutrons are unpolarized and the nuclear spins are randomly oriented, each state has the same probability. Thus the probability of b^+ and b^-, named n^+ and n^-, are $\frac{I + 1}{2I + 1}$ and $\frac{I}{2I + 1}$, respectively.

Furthermore, the scattering length of an atom can be decomposed as

$$b = \bar{b} + \Delta b \tag{2.9}$$

where $\bar{b}$ is the averaged scattering length given by $n^+ b^+ + n^- b^-$, and Δb is the deviation from the averaged. The following development for two atoms can be further made:

$$\langle b_m b_n \rangle = \overline{b_m} \, \overline{b_n} + \langle \Delta b_m \Delta b_n \rangle = \overline{b_m} \, \overline{b_n} + (\overline{b_{m=n}^2} - (\overline{b_{m=n}})^2) \tag{2.10}$$

Thus, we can develop equation 2.4 by using the equations 2.5, 2.6 and 2.10:

$$
\begin{aligned}
\frac{\partial^2 \sigma}{\partial \Omega \partial \omega} &= \frac{1}{2\pi N} \frac{k_f}{k_i} \sum_{m,n} \langle b_m b_n \rangle \int_{-\infty}^{\infty} \left\langle e^{i\mathbf{Q}\mathbf{r}_m(0)} e^{-i\mathbf{Q}\mathbf{r}_n(t)} \right\rangle e^{-i\omega t} dt \\
&= \frac{1}{2\pi N} \frac{k_f}{k_i} \sum_{m,n} (\overline{b_m}\,\overline{b_n}) \int_{-\infty}^{\infty} \left\langle e^{i\mathbf{Q}\mathbf{r}_m(0)} e^{-i\mathbf{Q}\mathbf{r}_n(t)} \right\rangle e^{-i\omega t} dt \\
&+ \frac{1}{2\pi N} \frac{k_f}{k_i} \sum_{m} (\overline{b_m^2} - (\overline{b_m})^2) \int_{-\infty}^{\infty} \left\langle e^{i\mathbf{Q}\mathbf{r}_m(0)} e^{-i\mathbf{Q}\mathbf{r}_m(t)} \right\rangle e^{-i\omega t} dt \\
&= \frac{1}{2\pi N} \frac{k_f}{k_i} \sum_{m,n} (b_{coh,m} b_{coh,n}) \int_{-\infty}^{\infty} \left\langle e^{i\mathbf{Q}\mathbf{r}_m(0)} e^{-i\mathbf{Q}\mathbf{r}_n(t)} \right\rangle e^{-i\omega t} dt \\
&+ \frac{1}{2\pi N} \frac{k_f}{k_i} \sum_{m} b_{inc,m}^2 \int_{-\infty}^{\infty} \left\langle e^{i\mathbf{Q}\mathbf{r}_m(0)} e^{-i\mathbf{Q}\mathbf{r}_m(t)} \right\rangle e^{-i\omega t} dt
\end{aligned}
\tag{2.11}
$$

From the equation above, we can define the coherent scattering length b_{coh} as the the mean value of the scattering length of a nucleus

$$
b_{coh} := \langle b \rangle = \frac{I+1}{2I+1} b^+ + \frac{I}{2I+1} b^-,
\tag{2.12}
$$

whereas the incoherent scattering length b_{inc} of a nucleus is the deviation from this mean value :

$$
b_{inc} := \sqrt{\langle b^2 \rangle - \langle b \rangle^2}, \quad \langle b^2 \rangle = \frac{I+1}{2I+1} (b^+)^2 + \frac{I}{2I+1} (b^-)^2
\tag{2.13}
$$

Table 2.2 reports the scattering lengths and cross sections of the main elements present in biological samples.

Thus, from equation 2.11 it results that the measured intensity consists of an incoherent part, resulting from nuclei auto-correlation, and a coherent part, coming from nuclei cross-correlation:

$$
\frac{\partial^2 \sigma}{\partial \Omega \partial \omega} = \frac{\partial^2 \sigma_{coh}}{\partial \Omega \partial \omega} + \frac{\partial^2 \sigma_{inc}}{\partial \Omega \partial \omega}
\tag{2.14}
$$

$$
\sigma S_{exp}(\mathbf{Q}, \omega) = \sigma_{coh} S(\mathbf{Q}, \omega)_{coh} + \sigma_{incoh} S(\mathbf{Q}, \omega)_{incoh}
\tag{2.15}
$$

	Z	σ_{coh} (barns) ($=10^{-24}$ cm^2)	σ_{incoh} (barns)	b_{coh} (pm)	b_{incoh} (pm) ($=10^{-12}$ m)
Atome					
H	1	1.57	**80.26**	-0.37	2.52
D	1	5.59	**2.05**	0.67	0.4
C	6	5.55	0	0.65	0
N	7	11.03	0.5	0.94	0.20
O	8	4.23	0	0.58	0
P	15	3.31	0.01	0.28	0.02
S	16	0.98	0	0.51	0

Table 2.2: Incoherent and coherent scattering cross sections and scattering lengths. Bold text indicates the values of particular interest for labeling biological samples.

2.1.5 Deuteration is a powerful tool

In order to probe the dynamical properties of a sample, one seeks to obtain the incoherent structure factor. For non-crystalline biological samples, it has been shown that the scattered intensity is mainly incoherent scattering coming from the hydrogen nuclei, which themselves reflect the dynamics of larger goups to which they are bound. This is due to the fact that hydrogen atoms scatter at least two orders of magnitude more than any other naturally abundant element, as shown in table 2.2. Moreover, the hydrogen atom has an incoherent cross section about 40 times higher than its isotope ^{2}H, also called deuterium (D).

This H/D isotope effect is widely employed to label and measure the scattering signal originating from specific parts of a given sample. In practice, one "hides" any part of the sample by replacing hydrogen atoms by deuterium, which to some extent have similar chemical properties. For protein samples, isotope labeling is sometimes used to focus on specific parts of the protein, such as methyl groups or solvent-exposed residues. Another common practice, which has been widely used in this work, aims to dissociate the protein dy-

Figure 2.3: Illustration of isotopic labeling. (top) A deuterated protein hydrated in H_2O allows to study the hydration water dynamics. (bottom) A hydrogenated protein hydrated in heavy water (D_2O) is used to measure protein dynamics.

namics from its hydration water dynamics in a system of hydrated proteins. On the one hand, the protein dynamics can be studied by using hydrogenated proteins in heavy water (D_2O), on the other hand, the hydration water dynamics is assessed by hydrating a perdeuterated protein in H_2O (see figure 2.3).

2.1.6 Elastic, quasi-elastic and inelastic scattering

In the scattering function S($\mathbf{Q}$,ω), ω represents the energy exchange (or transfer) between the scattered neutron and the sample. The energy is associated with time through a Fourier transformation. A common way to visualize the scattering function is to plot the scattered intensity as a function of the energy exchange at a given momentum transfer $\mathbf{Q}$. In a typical plot (see figure 2.4), three types of scattering can be distinguished. The elastic scattering refers to neutrons that did not exchange energy with the sample, quasi-elastic scattering refers to small energy exchanges around the elastic peak, and inelastic scattering arises from larger energy exchanges.
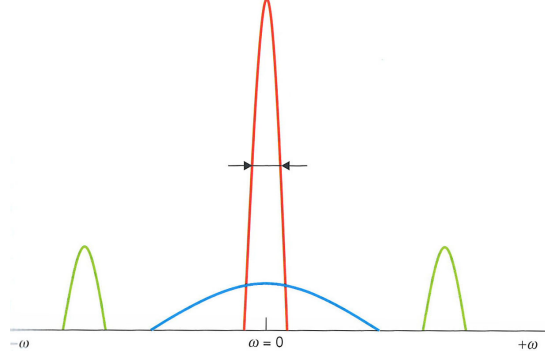
Figure 2.4: Schematic scattering intensities as a function of energy transfer observed at a given value of Q. The elastic peak is represented in red, with the energy resolution marked as the full width at half maximum (FWHM), the quasi-elastic scattering in blue and the inelastic scattering in green. Figure extracted from (Serdyuk et al., 2007).

Elastic neutron scattering   The elastic intensity in a scattering experiment is the peak of radiation scattered without energy exchange, *i.e.* $\omega = 0$. In theory, it should be a Dirac δ-function, but in practice, there is a threshold where the energy is too small to be measured by the instrument. This ω_{min} is called the energy resolution and is represented by the elastic peak drawn in red in figure 2.4. For incoherent scattering, the **Q** dependence of the elastic intensity reflects the averaged dynamics of the probed atoms. It is common to extract the mean square displacements (MSDs) from the Gaussian approximation of $S(\mathbf{Q},\omega = 0)$. This quantity provides an information about the amplitude of atomic motions, without presupposing their spacial distribution. The procedure to extract such a quantity is detailed in the methods of the different chapters.

Quasi-elastic neutron scattering   As we start to measure beyond the elastic peak, we gain information on the time constants of different dynamical processes. The quasi-elastic scattering intensity is shown in blue in figure 2.4. It is centered on a zero energy exchange and appears as a broadening of the elastic peak (it is therefore sometimes referred as the quasi-elastic broadening or quasi-elastic tails). Quasi-elastic scattering arises from diffusive processes, where the energy transfer width of the curve is inversely proportional to the relaxation time of the motions. Here, the data treatment consists in modeling a 2D data set, with Q values representing the spatial scale of the diffusion, and ω values representing the time scale of the diffusion. Many theoretical models have been developed, because, in most cases, one simple diffusion process is not sufficient to describe the quasi-elastic spectra. A thorough description of quasi-elastic neutron scattering analysis is given in the book of Marc Bee (Bee, 1988). The model used in this work is described in the methods of chapter 3.

Inelastic neutron scattering   The inelastic scattering intensity is shown in green in figure 2.4. The inelastic peaks arise from the atomic vibrations in the sample. Coherent

inelastic scattering probes the collective vibrational modes, which can be seen as waves propagating through the samples, or phonons. For liquids, it has been used to characterize collective density fluctuations, such as sound waves, propagating within the liquid. In incoherent inelastic scattering, the peaks arise from the vibration of individual atoms. It is possible to extract the vibrational density of states of the sample after proper data treatment.

In summary, four types of neutron scattering experiments can be distinguished:

- Coherent elastic scattering is used to obtain structural information on crystalline samples or protein solutions (see section 2.2 on small angle scattering for an example of a coherent scattering technique using protein solution samples).

- Coherent inelastic scattering probes the collective vibrations of the sample lattice and can be used for example to measure phonon propagation.

- Incoherent inelastic scattering probes individual atom vibrations and can be used to measure vibrational normal modes.

- Incoherent (quasi-)elastic neutron scattering is used to probe the diffusional properties of individual atoms.

It should be noted that, while elastic and inelastic scattering can be discriminated in the instrumental setup by measuring neutron energies, neutrons that are coherently or incoherently scattered are indistinguishable. One or the other is assumed to be predominant in the measured intensity after proper data treatments (*e.g.* background subtraction) or by the knowledge of the sample (chemical composition and structural arrangement, *i.e.* crystal-like or disordered).

2.1.7   Time and space windows

Each spectrometer has a defined time and space window, beyond which motions cannot be resolved by the instrument. The energy resolution, which defines the width of the elastic peak, corresponds to the longest lapse of time t_{max} over which the experiment is sensitive. If an atom has moved and reached a new position during a time longer than t_{max}, then the interference between waves scattered by the atoms in these two positions won't be detected. For instance, typical energy resolutions of neutron scattering spectrometers are 1 μeV and 10 μeV, which correspond to time resolutions of about 1 ns and 0.1 ns, respectively. In the same idea, a neutron spectrometer opens a window in space defined by the minimum and maximum momentum transfer (or scattering vector) Q_{min} and Q_{max} respectively, accessible by the instrument. Analogously to the relation between energy transfer ω and time, the momentum transfer Q is the reciprocal quantity of a distance in the direct space, and they are therefore inversely proportional. Small-amplitude motions will appear at large Q vectors and large-amplitude motions at low Q vectors. Thus, the space-time window of an instrument can be delilimited at its high-dynamics boundary by a displacement $2\pi/Q_{min}$ achieved within $2\pi/\omega_{min}$, and at its low-dynamics boundary by a

displacement $2\pi/Q_{max}$ achieved within $2\pi/\omega_{max}$. Here, ω_{max} refers to the highest energy exchange measurable by the spectrometer.

2.2 Small angle scattering

Small angle scattering (SAS), which can be light, x-ray, or neutron scattering, is a broad area that have many different applications. Because this technique was only marginally used in this work, it won't be presented in detailed here. Rather, I would like to provide a very intuitive introduction on the principle of small angle scattering experiments.

Small angles refer to small Q vectors, defined in section 2.1 as the momentum transfer or the scattering vector. This quantity has an inverse distance unit and is inversely proportional to the characteristic length that is probed (see section 2.1.7). In other words, small angle scattering experiments focuses on long-range correlations. While the neutron scattering techniques described in section 2.1 investigate inter-atomic distances, *i.e.* on the Å lengthscale, SAS focuses on the 1-100 nm length scale. In structural biology it means that it can be used to characterize the shapes of macromolecules (or of their complexes) or inter-protein interactions (see figure 2.5). One of the main advantages of SAS is its ability to probe the properties of particles in solution. This is major alternative to protein crystallography for proteins that cannot be crystallized.

Briefly, the measured intensity can be divided into two parts:

$$I(\mathbf{Q}) = F(\mathbf{Q}) \times S(\mathbf{Q}) \tag{2.16}$$

where $F(\mathbf{Q})$ is called the form factor and $S(\mathbf{Q})$ the structure factor. The form factor is the contribution most often studied and reflects the structure, or the *form* of the particle. The exploitation of form factors has been successful in obtaining low-resolution structures of large complexes as well as ensembles of disordered proteins. The structure factor is of interest when investigating inter-protein interactions, such as attraction or repulsion. It has been used in particular to study crowded solutions in this work.
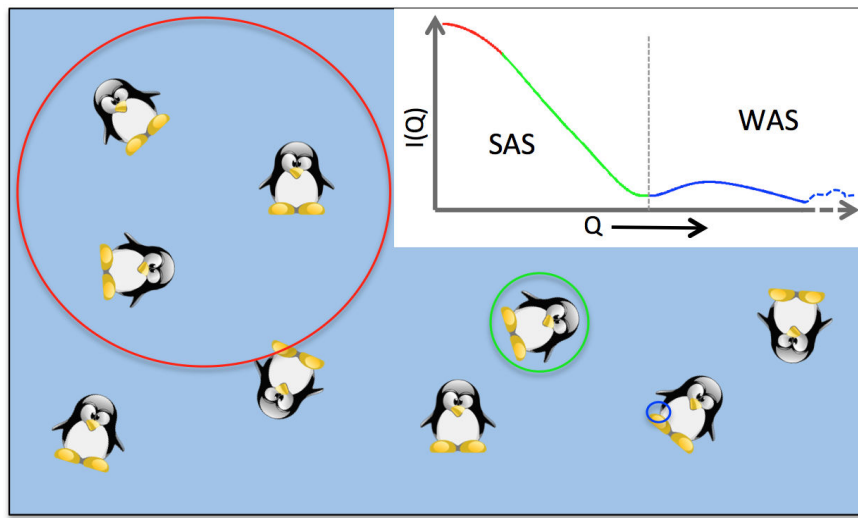
Figure 2.5: Illustration of a solution of structured particles. In the scattering curve (inset), the smaller are the Q, the larger correlation it reflects in solution (colored circles). A rough estimate of the correspondence between a characteristic distance and its Q vector can be obtained by the relation $D = \frac{2\pi}{Q}$. At low Q values, in the small angle scattering (SAS) region (red and green), inter-particle ordering (carried by the structure factor $S(\mathbf{Q})$) and particle shapes (contained in the form factor $F(\mathbf{Q})$) can be measured. At high Q values (blue), in the so-called wide angle scattering (WAS) region, inter-atomic distances are probed.

2.3 THz spectroscopy

The terahertz (THz) frequencies, which correspond to sub-millimetric wavelengths and around a few hundreds of wavenumbers, lie between the microwave and the infrared domains (see figure 2.6). Because water is highly absorbent in the range 1-5 THz, THz absorption spectroscopy is particularly suitable to probe water vibrational properties. It is important to note that the nature of motions responsible for the probed absorption cannot be directly identified by absorption spectroscopy. However, significant computational research has been very useful in revealing the underlying molecular mechanisms of the water THz spectrum. In particular, Matthias Heyden and coworkers have dissected the water motions fluctuating in the THz frequencies and have shown that they are essentially inter-molecular vibrations (Heyden et al., 2010). These motions that originate from the vibrations of the hydrogen bond network result in a fluctuation of collective dipole moments in the THz frequencies. Thus, THz spectroscopy can probe collective dynamics, *i.e.* collective modes that are delocalized over several atoms.

The collective water dynamics (*i.e.* the THz spectrum) has been shown to be perturbed by the presence of solutes in solution (Leitner et al., 2008). Indeed the solute-water interaction disrupts the water-water hydrogen bond network, thus changing its vibrational properties in the THz range. In particular this perturbation results in an increased density of modes around the probed frequency (2-3 THz), resulting in an increased absorption coefficient. Therefore, the water in the vicinity of a solute (including proteins) is perturbed and possesses absorption properties in the THz domain that are different from those of bulk water. This effect was shown to extend to several hydration layers away from the solute (Ebbinghaus et al., 2007; Meister et al., 2013). Heyden and Tobias recently showed *in silico* a long-range correlation between the THz fluctuations of atoms at the protein surface and water molecules that are up to 10 Å away from the protein surface (Heyden and Tobias, 2013). They concluded

> "...the protein solute can act as a synchronizing element for the vibrations of water molecules in its hydration shell, with consequences, for example, for the frequency dependent dipole fluctuations in the protein hydration water compared to independent collective oscillators in the bulk. The latter provides the connection to the experimentally observed changes of THz absorption coefficients in hydration shells of several globular proteins up to 10 Å thickness...".

Thus, the absorption of a protein solution is interpreted as the combination of three components : The absorption from the bulk water, from the perturbed water (hydration water) and from the protein (Leitner et al., 2008). Bulk water has a high absorption coefficient that is even enhanced for the hydration water, whereas proteins absorbed poorly in the THz domain.

A typical experiment consists in measuring the absorption of a solution as a function of protein concentration(Ebbinghaus et al., 2007). Assuming a model where only protein and bulk water contribute to the total absorption, an increase in protein concentration is expected to result in a linear decrease of the absorption, as water gets replaced by proteins.
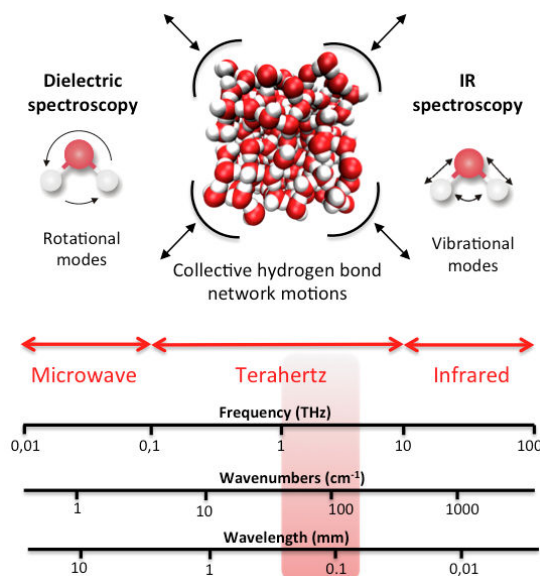
Figure 2.6: The THz regime, which corresponds to inter-molecular vibrational modes, is flanked by the dielectric and the infrared regimes. The red frame indicates the frequency range accessible to the table top p-germanium laser used in this work.

However, at low protein concentrations, a non-linear regime is observed, which originates from the contribution of the protein hydration water (see figure 2.7). A proper analysis of the non-linearity allows to extract the properties of the hydration water, such as its absorption coefficient and extension from the protein surface.

The experiment was carried out on a table top p-Ge THz laser difference spectrometer (Bergner et al., 2005), set up by Martina Havenith's team at the University of Bochum in Germany. The laser beam is divided into two paths, one hitting the sample and the other hitting a reference, before being recombined and measured (see figure 2.8). This setup allows the detection of small changes in absorption, on the order of few percents, which is required to probe the subtle changes due to the hydration water absorption.
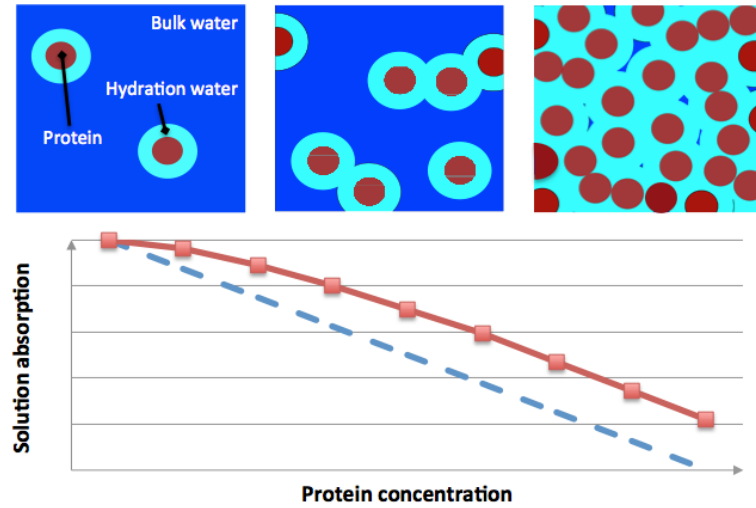
Figure 2.7: Bottom : Schematic representation of the absorption of a protein solution, as a function of protein concentration, assuming two (blue dash line) or three (red full line) absorbing components. Top : Illustration of a protein solution showing the three absorbing components. As the concentration increases (from left to right), the hydration shells of different protein molecules start to overlap, giving rise to a non-linear absorption as a function of protein concentration. The analysis of the non-linearity allows to obtain the volume of hydration water per protein.
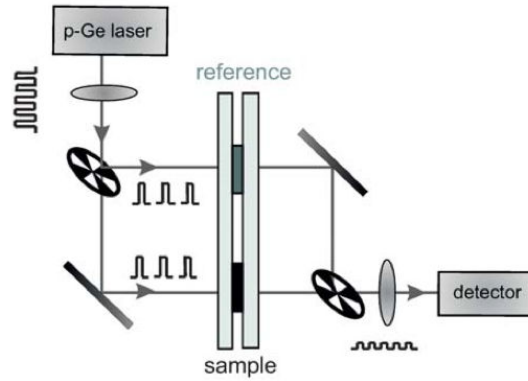
Figure 2.8: Schematic representation of the setup used for a precise measurement of THz absorption coefficients. The mirrored-chopper splits the radiation into two distinct paths. One part of the beam penetrates the buffer solution (reference), the second the protein solution. The radiation from both paths is then focused onto the detector and the difference between the sample and reference absorption coefficients is calculated. Figure extracted from (Niehues et al., 2011)

2.4 Molecular dynamics (MD) simulations

2.4.1 Introduction and equation of motion

Molecular dynamics (MD) simulation is a computational method that aims at calculating the time-dependent behavior of a molecular system. It is routinely used to investigate the structure, dynamics and thermodynamics of biological molecules. Nowadays, this method, run on state-of-the-art calculation centers, can simulate several hundreds of thousand of atoms on the nanosecond to microsecond timescale. In contrast to quantum mechanical approaches, MD simulation is based exclusively on classical physics. Though it can be quite complicated when it comes to details, the basic concept of MD simulations is quite simple. The key principle is to divide the time line into discrete time steps at which calculations are run to predict the state of the system at the following time step. Given a system of particles, one can determine how the system evolves over time (*i.e.* determine velocities and positions) by calculating the forces that act on each particle. This is done by resolving the second Newton's law :

$$\sum F = ma = m\frac{dv}{dt} = m\frac{d^2x}{dt^2} \tag{2.17}$$

where F are the forces, m the mass of the particle and a, v, x, the acceleration, velocity and position of the particle, respectively. The time step Δt must be small enough so that the forces can be assumed to remain constant, and thus that equation 2.17 can be integrated using, for exemple, the Verlet algorithm. In practice, Δt should be small enough to sample the fastest oscillations in the system and is generally set to a few femtoseconds. In this method, new positions are first calculated at time $t + \Delta t$ according to

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}\Delta t^2 a(t) \tag{2.18}$$

The forces (and therefore accelerations) are then calculated at time $t+\Delta t$, and the velocities are updated to their values at $t + \Delta t$ using

$$v(t + \Delta t) = v(t) + \frac{1}{2}a(t + \Delta t)\Delta t + \frac{1}{2}\Delta t^2 a(t). \tag{2.19}$$

The challenge of MD simulations is to find the best trade off between an accurate determination of the forces and the computational cost. The forces are computed according to an empirical force field that defines both the form and the intensity of the interaction potentials between atoms. The quality of the data from an MD simulation is only as good as the force field that is used to obtain them.

2.4.2 Force fields

The term force field is a general term used to describe the set of equations and parameters that defines how the various atoms interact with each other in a molecular simulation. The

forces being the derivative of the potential energy, such as

$$F = -\frac{\partial V}{\partial x} \quad , \tag{2.20}$$

a force field represents the total potential energy of a system of particles, and can be expressed as a sum of terms representing the potential energy of one or several atoms. Many force fields have been developed, proposing different approximations to encompass the basic physical interactions in a molecular system. One general approach is to separate between intra-molecular (bonded) interactions and inter-molecular (non-bonded) interactions. The CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field, employed in the present work, is widely used for biomolecular simulations and is a good example of a typical molecular mechanics force field. The total potential energy is given by

$$
\begin{aligned}
V = & \sum_{ij}^{bonds} \frac{1}{2} k_{ij} (r_{ij} - r_{ij}^{eq})^2 \\
& + \sum_{ijk}^{angles} \frac{1}{2} k_{ij} (\theta_{ijk} - \theta_{ijk}^{eq})^2 \\
& + \sum_{ijkl}^{dihedrals} k_{\phi,ijkl} (1 + cos(n\phi_{ijkl} - \phi_0)) \\
& + \sum_{ijkl}^{improper} \frac{1}{2} k_{\xi,ijkl} (\xi_{ijkl} - \xi_0)^2 \\
& + \sum_{i,j>i} \left[\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + (\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6}) \right]
\end{aligned}
\tag{2.21}
$$

The first four terms are the bonded interactions, of which geometrical illustrations are given in figure 2.9. The last line represents the two non-bonded terms : the electrostatic interactions and the van der Waals interactions represented by a Lennard-Jones potential. Further computational tricks, which won't be detailed here, can be used to optimize the computational cost of these different terms. At each time step, the potential energy is evaluated and integrated for every atoms, and plugged into equation 2.17 to predict the evolution of the system at the next time step.

2.4.3 Water models

As we have seen in chapter 1, water and protein dynamics are intimately coupled. Although the simulation of proteins used to be done in vacuum, it is now beyond doubts that a good representation of water is an essential factor to obtain realistic MD simulations. Despite the chemical triviality of water, H_2O, its representation and parametrization has remained a challenge for the past 40 years. Astonishingly, none of the models available today provide a complete representation of physico-chemical properties of water. However, each model
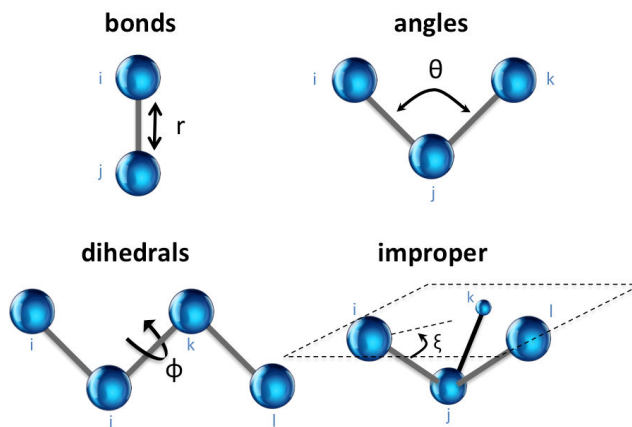
Figure 2.9: Illustration of the bonded interactions included in the CHARMM force field as described in equation 2.21.

is usually optimized to reproduce a precise property, such as water density, diffusion, ice nucleation, etc...

The cheapest way to treat the solvent (which is most of the time an aqueous solution) around the system of interest is to use the so-called implicit models, where the solvent is considered as a continuous medium. These models use a functional approach to calculate the influence of the solvent on each atom, in particular the contribution of water to the free energy of the system. Although this approach has been shown to be useful, especially for simulations of large systems, its accuracy is limited and it is anyway not suited to study the solvent properties.

In the explicit solvent approach, each atom of the solvent is explicitly represented. Many models have been proposed to represent water molecules. In general, they can be classified by following three points; (i) the number of interaction points, called sites, (ii) whether the model is rigid or flexible, (iii) whether or not the model includes polarization effects. Once again, one has to find a trade off between complexity of the model and computational cost. In the present work, we used rigid and non-polarizable models. A water molecule can be modeled by 2 to 6 sites, representing either real or dummy atoms, with defined charges and relative positions. Because three-site models achieve a high computational efficiency, they have been quite popular in MD simulations. Among them, the three site transferable interaction potential (TIP3P) (Jorgensen et al., 1983) and the extended simple point charge (SPC/E) (Berendsen et al., 1987) have been mainly used in this work. The parameters of these two water models are presented in table 2.3. TIP3P is the water model implemented with the CHARMM force field, used in this work, but SPC/E reproduced more accurately the experimental dynamical properties of bulk water. A thorough comparison of different water models was published in 2011 (Vega and Abascal, 2011) and ranked the four-site model TIP4P/2005 as the best model, followed by SPC/E.

	TIP3P	SPC/E
O-H bond length [Å]	0.9572	1.0
H-O-H angle [deg]	104.52	109.47
Lennard-Jones parameter A $[kcal.\mathring{A}^{12}/mol]$	582.0	629.4
Lennard-Jones parameter B $[kcal.\mathring{A}^{6}/mol]$	595.0	625.5
Partial charge of oxygen atom	-0.834	-0.8476
Partial charge of hydrogen atom	+0.417	+0.4238

Table 2.3: Parameters of the two water models used in this study. The parameters A and B represent the Lennard Jones parameters, A_{ij} and B_{ij}, in equation 2.21.

2.4.4 MD simulation and neutron scattering

As seen in chapter 2.1, neutron scattering typically probes molecular dynamics from the sub-ps time scale (vibrational motions, inelastic scattering) up to the ns time scale (diffusional motions, elastic scattering). Because the time scale accessible by MD simulations is very similar, the two techniques can be used in a complementary manner. A general approach is to compute *in silico* neutron scattering observables in order to evaluate the accuracy of the simulation, which is then used to further interpret the experimental results at a molecular level.

From the coordinates of each atom as a function of time (so-called trajectories), one can compute the intermediate scattering function, expressed in equation 2.6, for any Q value. A numerical Fourier transformation allows then to obtain the dynamical structure factor, directly comparable with the experiment. Note that one can compute independently coherent or incoherent structure factors.

The vibrational density of states (VDOS) $g(E)$ is the frequency distribution of vibrational modes in a system. It can be computed for any atoms by Fourier transforming their velocity autocorrelation function, *i.e.*

$$g(E) = \frac{1}{2} \int_{0}^{+\infty} \langle v_j(0)v_j(t) \rangle e^{-iEt} dt \tag{2.22}$$

where $v_j(t)$ is the velocity of an atom j at a time t, and E the energy. The VDOS can also be obtained by inelastic neutron scattering, thus representing another point of overlap between neutron scattering and MD simulations.

2.5 Time scale diagram : methods and biological processes
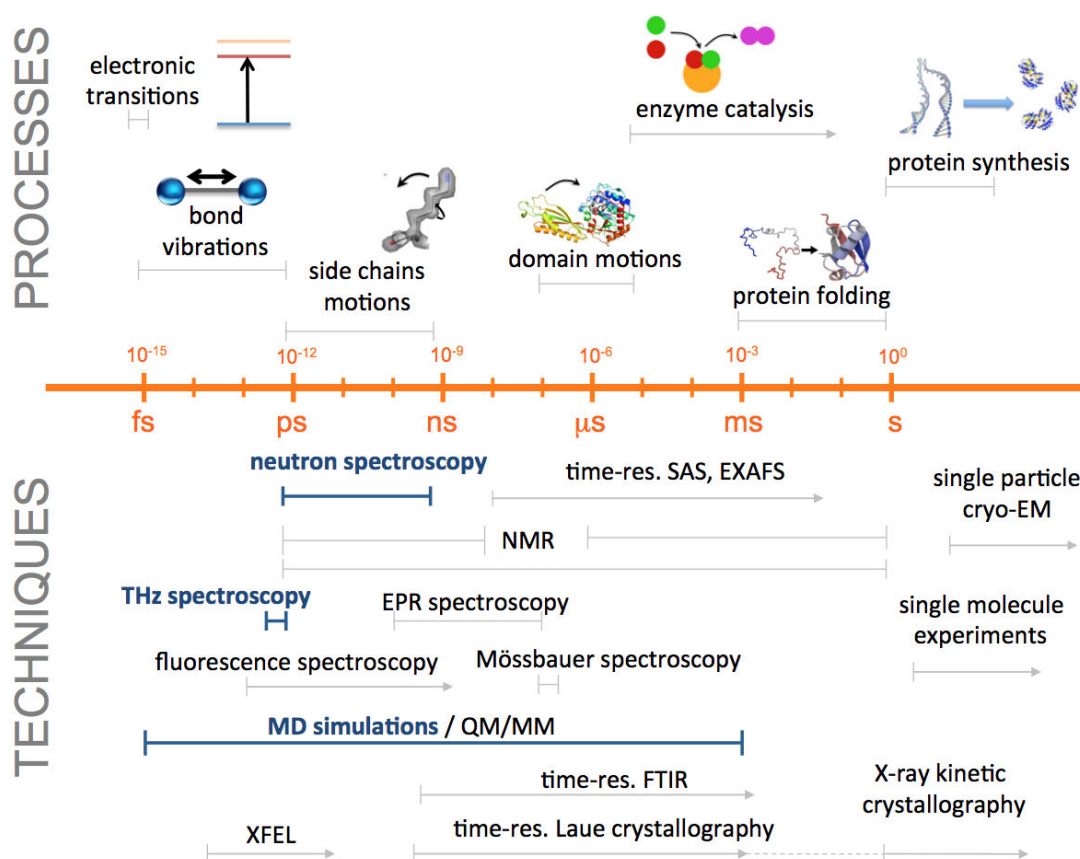
Figure 2.10: Temporal resolution of different biophysical methods, and time scale of different biological processes. The techniques used in this work are highlighted in blue. Diagram based on personal communications with Dr. Martin Weik.

Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins

The manuscript presented herein has been published in *Nature Communications* (Schiro et al., 2015). Its formatted version is available online.

3.1 Manuscript

Giorgio Schiro[1,2,3,*], Yann Fichou[1,2,3,10,*], Francois-Xavier Gallat[1,2,3,4,*], Kathleen Wood[5], Frank Gabel[1,2,3], Martine Moulin[4,8], Michael Härtlein[4,8], Matthias Heyden[6], Jacques-Philippe Colletier[1,2,3], Andrea Orecchini[9], Alessandro Paciaroni[9], Joachim Wuttke[7], Douglas J. Tobias[10], Martin Weik[1,2,3]

[1]Univ. Grenoble Alpes, IBS, F-38044 Grenoble, France; [2]CNRS, IBS, F-38044 Grenoble, France; [3]CEA, IBS, F-38044 Grenoble, France; [4]Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France; [5]Australian Nuclear Science and Technology Organisation Bragg Institute, Menai, Australia; [6]Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany; [7]Forschungszentrum Jülich, JCNS at MLZ, Lichtenbergstrasse 1, 85747 Garching, Germany; [8]ILL-EMBL Deuteration Laboratory, Partnership for Structural Biology, Grenoble; [9]Dipartimento di Fisica e Geologia, Università di Perugia, Via Pascoli, 06123 Perugia, Italy; [10]Department of Chemistry, University of California, Irvine, California, USA.

*Equal contribution

Abstract

Hydration water is the natural matrix of biological macromolecules and is essential for their activity in cells. The coupling between water and protein dynamics has been intensively studied, yet remains controversial. Here we combined protein perdeuteration, neutron scattering and molecular dynamics simulations to explore the nature of hydration water motions at temperatures between 200 and 300 K, across the so called protein dynamical transition, in the intrinsically disordered human protein tau and the globular maltose binding protein. Quasi-elastic broadening is fitted with a model of translating, rotating and immobile water molecules. In both experiment and simulation, the translational component markedly increases at the protein dynamical transition (around 240 K), regardless of whether the protein is intrinsically disordered or folded. Thus, we generalize the notion that the translational diffusion of water molecules on a protein surface promotes the large-amplitude motions of proteins that are required for their biological activity.

Main text

Hydration water surrounds biological macromolecules such as soluble proteins. Rather than being a mere bystander, hydration water actively partakes in macromolecular activity, such as protein-protein and protein-DNA recognition, protein allostery, electron and proton transfer and enzyme reactions[1]. The first hydration layer on the protein surface is of particular importance for biological activity that is lost in most completely dry proteins[2]. A protein together with its first hydration layer thus forms the biologically active entity. Protein and water molecules are connected by an extended hydrogen bonded network, the fluctuations of which lead to breakage and formation of water-protein hydrogen bonds that eventually allow for functionally important protein motions at physiological temperatures[3]. The coupling between water and protein dynamics has been, and remains, a matter of extensive debate. The relation has been compared to the one between a master (water) and a slave (protein dynamics)[4], yet a mutual give-and-take analogy probably best accounts for the experimentally observed differential influence of different classes of biological macromolecules (globular, membrane and intrinsically disordered proteins[5], DNA, RNA[6]) on the dynamics of their hydration water. The complex ensemble of water-protein motions can be teased apart by extending experiments and simulations down to cryo-temperatures; hydrated proteins undergo a transition to large-amplitude anharmonic motions on the picosecond timescale as the

temperature is increased above about 240 K. The origin of this transition, known as the protein dynamical transition[7], is controversially debated in the biophysical community[8-14]. A transition is also apparent in the picosecond motions of hydration water at a similar temperature[15]. The hydration water transition has been suggested to either be of kinetic origin and the manifestation of a glass transition at around 180 K[16-18] or due to a thermodynamic transition between a high- and a low-density form of water[19]. Despite the various controversies, there is a general agreement about the primary role played by hydration water dynamics in protein dynamics and function, since decreasing the hydration level significantly below monolayer coverage effects in suppressing both the dynamical transition[20,21] and the biological activity[2,22] of a protein. Molecular dynamics (MD) simulations have suggested that the onset of water translational diffusion on the protein surface correlates with the protein dynamical transition, so that water and protein motions emerge in parallel on the same time scale[23,24]. Yet, experimental evidence for such a correlation has remained elusive and its universal validity for different protein classes, including intrinsically disordered proteins, unexplored.

In contrast to folded proteins, intrinsically disordered proteins (IDP) lack a single, well defined three-dimensional structure, and are best described by an ensemble of rapidly interconverting conformations[25]. The unfolded nature of IDPs originates from a large content of hydrophilic amino acid residues compared to hydrophobic ones, preventing the hydrophobic collapse that would lead to a folded globular protein[26]. IDPs fulfill specific roles in biological cells, where they are mainly involved in signaling and regulatory processes[27-29] and in many cases fold upon binding to their target partner[30]. About 30% of eukaryotic proteins are thought to be either partially or fully disordered[31]. IDPs are more flexible than folded proteins, as shown, e.g., by neutron[5,32] and NMR spectroscopies[33]. Yet, like folded proteins IDPs undergo a dynamical transition[5] that is suppressed if hydration water is absent. Due to their unfolded nature, IDPs are characterized by a solvent accessible surface exceeding several times the one of folded globular proteins. One might thus conjecture that the interaction with hydration water is different for IDPs than for globular proteins.

Neutron scattering is a valuable tool to study the nano- to picosecond dynamics of biological macromolecules and of hydration water[34]. The incoherent scattering cross section of a hydrogen atom exceeds by two orders of magnitude the ones of all other atoms in a protein, including its isotope deuterium, and thus dominates the neutron scattering signal in

non-crystalline samples. Therefore, fully deuterating a protein (perdeuteration) is an elegant way to focus on hydration water dynamics by minimizing the protein's contribution to the incoherent scattering signal. Yet, perdeuterating a protein in the large quantities needed for energy-resolved incoherent neutron scattering experiments (100 – 200 mg) is time-consuming, costly, and has been carried out for only a handful of proteins so far, including C-phycocyanin[35], purple membrane[36], maltose binding protein[15], myoglobin[37], tau[5] and GFP[18]. For powders of perdeuterated proteins hydrated at a level of 0.4 g H_2O per g protein, at least 70% of the incoherent neutron scattering signal is from hydration water that corresponds roughly to monolayer coverage of the protein surface[36]. The remaining ~30% of the signal originates essentially from exchanged hydrogen atoms of the protein.

All-atom MD simulations are complementary to neutron scattering experiments. MD simulations routinely cover the same time scale (ps-ns) of scattering experiments, and it is straightforward to compute neutron scattering spectra from MD trajectories. Comparison of MD and neutron scattering data is useful for validating the simulations and, once validated, MD simulations can scrutinize models used to interpret the scattering data. MD simulations have been widely used to characterize the vibrational and diffusional properties of protein hydration water[15,23,24,38-43].

Here we studied hydration water dynamics on the surface of the folded maltose binding protein (MBP) and of the intrinsically disordered human protein tau that is, like many IDPs, involved in a neurodegenerative disease, *viz*. in Alzheimer disease[44]. A combination of protein perdeuteration, quasi-elastic neutron scattering and MD simulations provided evidence for an activation of translational water motions on tau and MBP at the temperature where the proteins undergo a dynamical transition. We establish a general connection between the translational diffusion of water molecules on a protein surface and the promotion of the large amplitude motions of proteins required for their biological activity. This connection is thus independent of the protein folding state.

Results

Neutron scattering reveals a change in water dynamics at 240K

The htau40 protein (the longest and most commonly found isoform of human tau, referred to as tau throughout the manuscript) and the maltose binding protein (MBP) are of similar molecular mass (corresponding to 441 and 380 amino acids, respectively), yet belong

to different protein families; the former is an intrinsically disordered protein, whereas the latter is fully folded. Hydrated powders of both proteins have been prepared and characterized earlier[5]. Elastic incoherent neutron scattering experiments on D_2O-hydrated, non-deuterated tau[5] and non-deuterated MBP[15] samples have indicated that a protein dynamical transition (fig. 1) takes place at a similar temperature in both proteins. This temperature can be estimated by subtracting the mean-squared displacements (MSD) of hydrated tau from those of dry tau (see inset in fig. 1) and calculating the second derivative that shows a sharp peak close to 240 K (Supplementary Figure 1).

In the present study, quasi-elastic neutron scattering (QENS) measurements were carried out on H_2O-hydrated, perdeuterated tau (denoted D-tau-H_2O) and H_2O-hydrated, perdeuterated MBP (D-MBP-H_2O) in an ample temperature range across the protein dynamical transition, *i.e.* between 200 and 300 K, to monitor hydration water dynamics. The data were recorded on the backscattering spectrometer SPHERES[45] (Jülich Centre for Neutron Science at MLZ, Garching, Germany) with an energy resolution of $\Delta E = 0.7$ μeV (full width at half maximum (FWHM)) that allows monitoring of motions in the nano- to picosecond (ns-ps) time scale. The neutron scattering signal shows quasi-elastic broadening at temperatures above 200 K, indicating the presence of ns-ps motions in the hydration shell. A representative description of the observed temperature dependence is given in figure 2, where the spectra of tau hydration water are reported at all measured temperatures, normalized for the spectral area measured at 20 K and at the scattering wavenumber $q=0.78$ Å^{-1} ($q=\lambda^{-1}4\pi\sin(\theta/2)$, where λ is the incident neutron wavelength and θ the scattering angle). An abrupt change in the spectral lineshape occurs between 240 and 250 K (fig. 2; an analogous behavior is revealed in MBP and at all the measured q values (not shown)). Such a change reveals that a modification in the dynamics of hydration water takes place on the ns-ps time scale in this temperature region, which corresponds to the protein dynamical transition temperature revealed by elastic data on D_2O-hydrated samples (fig. 1).

Data analysis reveals an onset of water translation at 240K

Water diffusional motions can be described as a translation of the center-of-mass and a rotation around the center-of-mass. Here we employ a model for analyzing QENS data in which water molecules either translate, rotate or remain immobile (see Methods section). Furthermore, we adopted a global fitting approach, in which the difference between

experimental data and the fitting function is minimized in the two-dimensional space defined by energy transfer (E) and scattering wavenumber (q). This procedure minimizes the number of free parameters and allows the identification of contributions to the QENS spectrum arising from confined motions (such as rotational motions). Indeed, since these contributions have a typical length scale (e.g. the radius of a rotational motion), their intensity is q-dependent and can vary significantly in the accessible q-range. In figure 3, we show typical results of the fitting procedure at two different q values, illustrating the excellent agreement of the model with the experimental data (fitting results at other temperatures and q values are reported in Supplementary Figure 2). Data shown in the upper panels of figure 3 highlight the importance of fitting the spectra by simultaneously taking into account their q-dependence. Indeed, the intensity of the rotational contribution cannot be distinguished from the immobile contribution (contained in the elastic peak in fig. 3) when analyzing single spectra at low q values.

From the global fit analysis, we can extract for each temperature point fractions of the total scattering intensity originating from water molecules that undergo translational or rotational diffusion or that remain immobile. These fractions are displayed in figure 4 for both tau and MBP hydration water as a function of temperature. For both tau and MBP, the translational fraction increases steeply above 240 K (fig. 4a). The rotational fraction has emerged at 200 K and increases up to 250 K, where it reaches a plateau (fig. 4b). Below 200 K, nearly all water molecules are thus immobile (fig. 4c) on the time scale determined by the SPHERES spectrometer (motions faster than 2 ns). Additional QENS experiments were carried out at various temperatures on natural-abundance tau and MBP powders, hydrated in D_2O (results to be published in detail elsewhere), and a contribution from the resulting protein dynamics included in eq. 1. The resulting temperature dependence of translational, rotational and immobile fractions of water molecules was qualitatively similar to the ones shown in figure 4. The quasi-elastic data (fig. 2) were collected in the energy range +/-15.8 μeV (see Methods section). In order to uncover a possible effect of the limited energy range on data analysis, we also collected a set of QENS spectra of D-tau-H_2O at four selected temperatures (240, 260, 280, and 300 K) in the full energy range available on SPHERES (+/-30.9 μeV). The extracted translational, rotational and immobile fractions are almost identical to the ones displayed in figure 4 (see Supplementary Figure 3). The fitting model (eq. 1) also contains a background (*i.e.* k(q)) originating from an instrumental background and fast motions outside

the dynamic range of the spectrometer. k(q) appears to display a temperature dependence with a transition at 240 K for all q values (Supplementary Figure 4), yet the rate of increase above 240 K is q-dependent. A q-dependent background with a transition at 240 K is compatible with a fast translational water component (detected by SPHERES as a flat energy-independent contribution; see Supplementary Figure 5 and Supplementary Information for a complete discussion) that shows a similar temperature dependence than the slower component shown in figure 4a. The other dynamical parameters (*i.e.*, rotational correlation rate and translational diffusion coefficient) reveal a perturbation of protein hydration water with respect to bulk water as already reported in the literature (see Supplementary Figure 6).

MD simulations confirm the onset of water translation

In order to complement the QENS results and to confirm their connection with previous computational work[23], we carried out MD simulations on tau and MBP. Models of protein powders hydrated at 0.4 g H_2O / g protein were prepared and the simulations were carried out at constant pressure and temperature for about 50 ns and analyzed. The agreement between simulation and experiment of the temperature evolution of the water MSD was found to be excellent for both MBP and tau (Supplementary Figure 7), thus validating our simulations and justifying further *in silico* analyis of water dynamics. We evaluated the temperature dependence of two types of protein-water hydrogen bond (HB) lifetimes computed from the simulations: the continuous HB relaxation time (also known as the fast HB relaxation time), τ_{HBC}, and the intermittent HB relaxation time (also referred as the slow HB network relaxation time), τ_{HBI}. τ_{HBC} is the average time before a protein-water HB breaks, while τ_{HBI} represents the timescale for relaxation of the protein-water HB network[46]. Relaxation rates, defined as $1/\tau$, are plotted as a function of temperature in figures 5a and 5b. The continuous HB relaxation rates (fig. 5a) show a smooth variation over the entire temperature range for both protein powders. In contrast, the intermittent HB relaxation rate (fig. 5b) exhibits a sharp increase at around 240-260 K. Furthermore, we computed for both protein powders the MSD of the hydration water oxygen atoms, assessing thereby the translational dynamics of water molecules. The MSD, plotted at 100 ps as a function of temperature in figure 5c, also show a sharp increase at around 240 K, signaling the onset of water translational diffusion.

Discussion

We addressed the relation between hydration water motions and the dynamics of both an intrinsically disordered protein (human htau40) and a folded protein (MBP) by combining protein perdeuteration, neutron scattering and MD simulations. Quasi-elastic neutron scattering of perdeuterated proteins hydrated in H_2O, which enables the isolation of hydration water motions, was measured at several temperatures between 200 and 300 K. The quasi-elastic signal broadens abruptly at 240 K, indicating that hydration water motions become activated on the ns time scale at a similar temperature where the underlying protein undergoes a dynamical transition that corresponds to the onset of anharmonic motions. Fitting the broadening with a model in which water molecules either translate, rotate or remain immobile provided evidence for an onset of water translational motions at the protein dynamical transition of both MBP and tau. The freezing of all water-diffusion degrees of freedom around 200 K on the ns time scale is in agreement with the presence of a second order-like glass transition of hydration water at a lower temperature on a longer time scale (100 s), as revealed by dielectric spectroscopy and calorimetry[47].

A previous MD simulation study of the globular protein RNase A established a correlation between the relaxation of the protein-water hydrogen-bond network and the onset of anharmonic motions that occurs at the protein dynamical transition[23]. In the present study, MD simulations of tau and MBP at various temperatures below and above the dynamical transition showed that the lifetimes (τ_{HBl}) of the protein-water hydrogen-bond networks abruptly decrease at the dynamical transition, concomitant with the onset of translational water diffusion. The observation of this coincidence for two globular proteins (RNase A[23] and MBP), as well as for an entirely disordered protein (tau), suggests a universality of the correlation between an abrupt increase in the hydrogen-bond network relaxation rate and the dynamical transition.

The hydration water properties of intrinsically disordered proteins are thought to have specific biological relevance, such as ensuring protein activity in desiccated cells[48], but they remain poorly understood. Solid-state NMR revealed larger activation energies for IDP than for globular-protein hydration water motions and indicated that IDPs bind more hydration water than globular proteins[49]. ^{17}O magnetic relaxation dispersion experiments suggested that IDPs with extended conformations display weaker dynamical perturbations of their hydration water than folded proteins[50]. Elastic neutron scattering has provided evidence of smaller

mean-square displacements for the hydration water of an IDP compared to a globular protein[5]. Despite these differences that might be due to the conformational heterogeneity and the amino acid composition of IDPs that differ from those of folded proteins, our results show that water translational motions couple to large-amplitude protein motions in both globular and intrinsically disordered proteins. Among different classes of soluble proteins, the coupling between hydration water and protein motions thus seems to be of similar nature.

Concerning the water diffusive model used here, we underline that a much better fitting quality is obtained with a function where the translational and rotational terms are added (see eq.1) rather than convoluted (*viz.* chi-squared at least 20% higher on average, but even 50% higher at temperatures above 250 K; see Supplementary Figure 8). This indicates the presence of two populations of water molecules with different dynamical properties within the dynamical window probed by the SPHERES neutron spectrometer (ns-ps motions). The existence of two populations is supported by water displacement distributions calculated from the MD simulations (Supplementary Figure 9). The rotational term accounts mainly for water molecules strongly interacting with the protein surface and/or inside protein docking sites, while the translational term accounts for less hindered water molecules whose rotational motion is faster than the time window accessible by SPHERES. This finding is in agreement with the previously observed heterogeneity of water motions in the first hydration shell[51] and is in agreement with the experimental evidence that a certain protein hydration level (typically h > 0.15 g water / g protein) is needed for the dynamical transition to take place: at lower hydration levels water molecules are probably so tightly bound to the protein surface that only rotational motions are energetically accessible, while translational motions are not allowed and, as a consequence, the full set of protein motions required for the dynamical transition to take place can not be activated. Extending the present study to lower hydration levels would further clarify this point.

In summary, the combination of simulation and experimental data reported herein confirms the notion that the protein dynamical transition is connected to the onset of water translational diffusion *via* relaxation of the protein-water hydrogen-bond network. A key aspect of our approach was a line width analysis of QENS data that yielded insight into the nature of water motions not provided by mean-squared displacements deduced from elastic scattering. The similarity of the results obtained for the globular protein MBP and the IDP tau suggests, unexpectedly, that this connection does not depend on the degree of order in a

protein structure, and is a generic feature of the dynamics of soluble proteins. The protein dynamical transition has been shown to be correlated with the onset of biological activity[2,22,52]. Our data thus suggest that the ability of hydration water to undergo translational motions on a protein surface is a necessary requirement for protein activity. Extrapolated to the context of a biological cell, we conjecture that macromolecular activity might be modulated *via* a modification in the translational diffusion properties of hydration water. Such a modification can, e.g., be brought about by geometric confinement, macromolecular crowding or the presence of small solutes such as sugar molecules.

Methods

Sample preparation for neutron scattering experiments

We used the same perdeuterated htau40[5] (D-tau-H_2O) and perdeuterated maltose binding protein[15] (D-MBP-H_2O) samples as employed in our previous studies. Details about sample preparation and characterization and chemicals used can be found in Gallat et al.[5] (tau) and Wood et al.[15] (MBP). Briefly, 204 mg of perdeuterated htau40 and 221 mg of perdeuterated MBP were hydrated to a level of 0.38 g H_2O/g D-tau (D-tau-H_2O sample; corresponding to 1040 water molecules per tau molecule) and 0.37 g H_2O/g D-MBP (D-MBP-H_2O sample; corresponding to 910 water molecules per MBP molecule), respectively, in a 4×3 cm^2 flat aluminum sample holder. If one assumes that all exchangeable deuterons in the D-tau (H-MBP) protein, *i.e.* 24% (22%) of all deuterons in the protein, exchange, 71% (72%) of the total incoherent scattering cross section originates from hydration water (H_2O) and 29% (28%) from the protein.

Quasi-elastic incoherent neutron scattering experiments

QENS data were recorded on the backscattering spectrometer SPHERES (Jülich Centre for Neutron Science at the Heinz Maier-Leibnitz-Zentrum Garching (MLZ), Garching, Germany) with a scattering wavenumber range q=0.46–1.66 A^{-1}, an energy resolution $\Delta E = 0.7$ μeV (FWHM)[45] and in the energy range $E= \pm 15.8$ μeV, corresponding to a ns-ps time window. The samples were inserted at room temperature into a Janis cryostat at 135° with respect to the incoming neutron beam. QENS spectra were collected at several temperatures in the following sequence: 20, 200, 210, 220, 230, 240, 300, 280, 260, 250 K for D-tau-H_2O and 20, 250, 200, 220, 235, 260, 280, 300 K for D-MBP-H_2O. The total measuring time per

temperature was 10h (20-220K) or 9h (230-300K) for D-tau-H_2O and 12h for D-MBP-H_2O. Detector efficiencies were corrected with a measurement of the scattering intensities from the samples at 20 K. Data points were binned in 0.04 μeV channels. Spectra at some selected temperatures were also collected in the full ±30.9 μeV energy range accessible on SPHERES, in order to check possible effects of the accessible energy window on the results of data analysis (see Supplementary Figure 3).

Neutron data analysis

QENS data were analysed using a rotation-translation model[53] for hydration water dynamics with four components: a translational term, taking into account the center-of-mass motion of water molecules; a rotational term, accounting for motions around the center-of-mass; a term for the "immobile" water molecules (*i.e.* the waters characterized by motions too slow to be detected in the dynamic range probed by the neutron spectrometer); and a flat energy-independent term for the instrumental background and for motions faster than the time window probed by the spectrometer. All the energy-dependent terms were convoluted with the instrumental resolution function, which was obtained by measuring the spectra of the same samples at 20 K, where the incoherent scattering signal is almost exclusively of elastic nature. Both translational and rotational terms have a Lorentzian-like dependence on energy transfer, thus making them difficult to distinguish when analyzing a single spectrum at a given q value. The main difference between the translational and rotational terms lies in the q-dependence of their shape parameters: the first one has a q-dependent width and a q-independent intensity, while the opposite holds for the second one. These characteristics allow us to distinguish and assign the two different dynamical contributions to the QENS spectrum when using a global fitting approach in the q-E space, like the one adopted here. The model for the scattering function $S(q,E)$ has the following analytical form:

$$S(q,E) = DW(q) \cdot \left[\left[A_0 \cdot \delta(E) + A_t \cdot S_t(q,E) + A_r \cdot S_r(q,E) \right] \otimes R(q,E) + k(q) \right] \text{ (eq.1)}$$

$DW(q) = \exp(-q^2 \langle u^2 \rangle)$ is the Debye-Waller factor due to the vibrational motions. The intensity of the elastic term ($A_0 = A_r \cdot J_0^2(qb) + A_i$, where J_0 is the 0th order spherical Bessel function) contains a q-dependent term from the rotational diffusion model and a q-

independent term A_i that takes into account hydrogen atoms not moving on the time scale set

by the resolution function. $S_t(q,E) \propto \dfrac{Dq^\lambda}{E^2+(Dq^\lambda)^2}$ is the contribution of translational motions

(for $\lambda=2$ the standard Brownian diffusion is recovered, while $\lambda<2$ indicates a sub-diffusive

character, often found in polymers, glass-forming and supercooled liquids)[54].

$S_r(q,E) \propto \sum_{l=1} (2l+1) J_l^2(qb) \dfrac{1}{\pi} \dfrac{l(l+1)\Gamma_r}{E^2+(l(l+1)\Gamma_r)^2}$ is the quasi-elastic broadening originating from

rotational motions[55]. $J_l(qb)$ are the spherical Bessel functions, while b is the radius of the

hydrogen rotation around the center-of-mass of a water molecule, which was set to 0.98 Å,

i.e. the H-O distance (the center-of-mass position essentially coincides with the oxygen atom

position). The first five terms of the series, accounting for more than 99% of the rotational

term in the explored q-range, were taken into account. $k(q)$ is an energy-independent

background that accounts for all the contributions faster than the timescale detected in the

SPHERES energy range, as well as the instrumental background. The free parameters in this

model are then: $<u^2>$, A_i, A_t, D_t, λ, A_r, Γ_r, $k(q)$ for a complete dataset at a given temperature.

The fitting procedure consists in minimizing the function $\chi^2 = \sum_{q,E} \dfrac{\left[S_{q,E}^{\text{exp}} - S(q,E)\right]^2}{\sigma_{q,E}^2}$, where the

summation is over all the (q,E) values, S^{exp} are the experimental points and σ the experimental

errors associated with the experimental points. To carry out the fitting, a Fortran code was

written, based on the Minuit minimization routine (lcgapp.cern.ch/project/cls/work-

packages/mathlibs/minuit/index.html), released by the CERN computing group.

Molecular dynamics simulations

The MD simulations were performed using the NAMD program[56] with the

CHARMM27 force field[57,58] for the protein and the SPC/E[59] model of water. The SPC/E

model has been shown to more accurately reproduce neutron scattering data on protein

hydration water than the TIP3P water model that is typically used in conjunction with the

CHARMM protein force field[60]. The simulations were carried out at constant temperature

using Langevin dynamics, and a constant pressure of 1 atm using the Nosé-Hoover-Langevin

piston algorithm with anisotropic cell fluctuations[61,62]. The equations of motion were

integrated using the Verlet-I/r-RESPA multiple-time step algorithm[63,64] with time steps of 4 fs

for the long-range non-bonded forces, 2 fs for the short-range non-bonded forces, and 2 fs for

the bonded intra-molecular forces. The SHAKE algorithm[65] was used to constrain the lengths of all bonds to H atoms. Electrostatic interactions were computed using the smooth particle-mesh Ewald sum[66], and the van der Waals interactions and the real-space part of the Ewald sum were smoothly switched to zero over the range 10-11 Å for the MBP powder and 10-12 Å for the tau powder.

The MBP powder model was prepared as described earlier[15]. Briefly, the powder consists of four molecules of MBP (PDB entry 1JW4) and 3,460 water molecules, corresponding closely to the hydration level of the samples used in the neutron experiments. The tau powder is composed of ten tau molecules and was prepared as follows. Out of an ensemble of 200 tau conformations determined by NMR and SAXS experiments[67], we randomly selected a sub-ensemble of ten conformations, for which the average radius of gyration matched the experimental value measured of tau in solution (62 Å)[5]. Those ten conformations were placed in the same simulation box before adding 10,105 water molecules (corresponding to the experimental hydration level of 0.4 g H_2O/g protein) spread over the first hydration shell of each protein. Snapshots from the simulations of both the tau and MBP powder models are shown in Supplementary Figure 10. We then started the simulation at constant temperature and pressure, so that the simulation box collapsed. After the box dimensions had stabilized we set the temperature to 500 K for 1 ns. The temperature was then reduced to 300 K, the simulation was prolonged for an additional 2 ns, and the final configuration was used as a starting point for the simulations at a series of temperatures. The two powder models were equilibrated for about 30 ns at 20, 150, 200, 230, 240, 260, 280, 300 K for MBP and 20, 150, 200, 220, 240, 260, 280, and 300 K for tau, followed by 14 ns of simulations at each temperature during which configurations were saved at 1 ps intervals for subsequent analysis. For the calculations of the protein-water hydrogen bond lifetimes (τ_{HBC}, defined below), the simulations were continued for another 400 ps, during which configurations were saved at 8 fs intervals.

The water dynamics in our powder models were validated by comparing the computed and experimental water mean-squared displacements (MSD) as a function of temperature, as shown in Supplementary Figure 7. The temperature dependence of the MSD from the simulations agrees well with the experimental data, although the absolute values are different for reasons discussed in the Supplementary Discussion.

Two measures of protein-water HB relaxation rates were considered to discriminate

between fast (~ps) HB formation/breakage (characterized by the "continuous" HB relaxation time τ_{HBC}) due to water rotational/librational motion, and the slower (> 10 ps) relaxation of the protein-water HB network (characterized by the "intermittent" HB relaxation time τ_{HBI}) due to diffusion of water on the protein surface or exchange of water molecules between hydration shells. τ_{HBC} is defined by the decay of the time correlation function $S_{HB}(t) = <h(0)H(t)>/<h>$[68] and τ_{HBI} by the decay of $C_{HB}(t) = <h(0)h(t)>/<h>$[69], where h(t) is equal to one if a given donor-acceptor (D-A) pair is hydrogen bonded at time t and zero otherwise, $h(t)$ is equal to one if the D-A hydrogen bond remains intact continuously from time 0 to time t and zero otherwise, and the angular brackets denote an average over all D-A pairs and time origins. The relaxation times τ_{HBC} and τ_{HBI} were defined as the times at which $S_{HB}(t)$ and $C_{HB}(t)$ decay, respectively, to 1/e. When the $C_{HB}(t)$ did not decay to 1/e within 4 ns, they were fitted with a stretched exponential and extrapolated to 1/e.

For the calculation of the correlation functions $C_{HB}(t)$ and $S_{HB}(t)$, a D-A pair was considered to be hydrogen bonded when the distance D-A was less than 3.5 Å and the D-H-A angle was greater than 150°. While the values of the hydrogen bond lifetimes (τ_{HBC}) at a given temperature depend on the details of the criterion used to define hydrogen bonds, previous work on bulk water suggests that the definition is not expected to affect the qualitative temperature dependence of the lifetimes[46].

Author contributions

MW proposed experiments; FXG, JPC, MM and MiH prepared samples; FXG, MW and JW performed neutron experiments; YF, MaH, and DJT carried out and analyzed the molecular dynamics simulations; GS analyzed neutron data with input from JW, FXG, AO, AP, KW, FG and MW; YF, GS, DJT and MW wrote the manuscript with input from all authors.
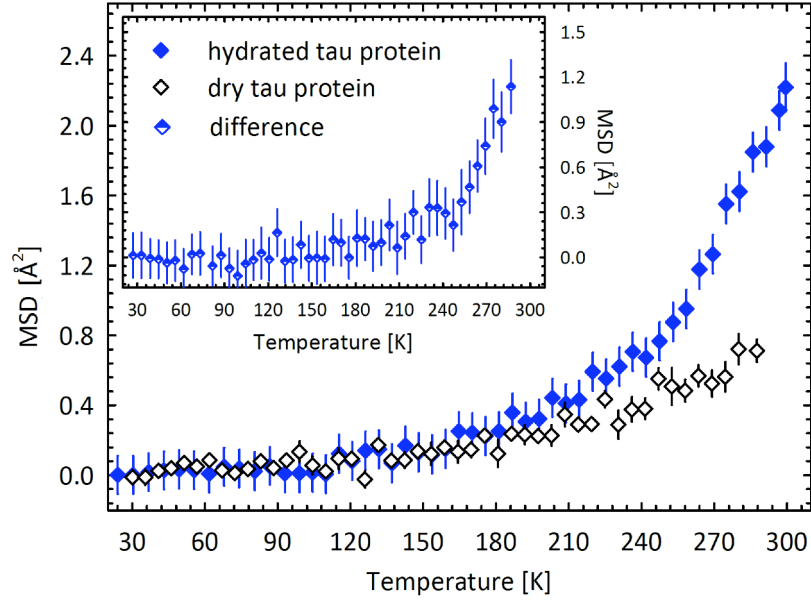
Figure 1. **Dynamical transition in hydrated proteins.** Mean-squared displacements (MSD) of the intrinsically disordered protein tau[5], in a hydrated (full blue diamonds) and a dry (open black diamonds) state, measured by elastic incoherent neutron scattering on the backscattering spectrometer IN16 (0.9 µeV resolution, ILL, Grenoble)[70]. Inset: the MSD difference between the hydrated and the dry protein highlights an onset of large-amplitude protein motions at around 240 K. Protein samples were not deuterated and were hydrated in D_2O. Experimental error bars are indicated as vertical lines.
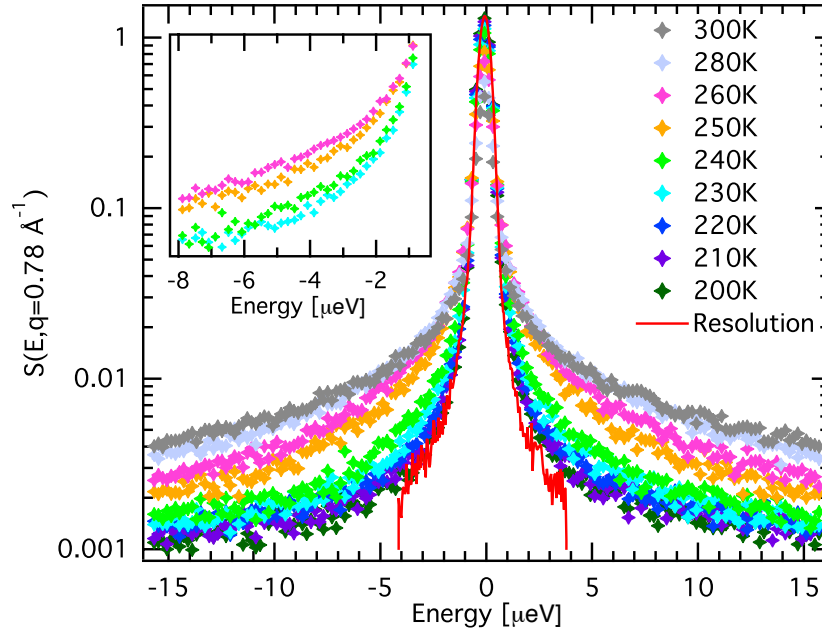
Figure 2. **Neutron spectra reveal a change in hydration water dynamics at 240 K.** Quasi-elastic neutron scattering spectra of D-tau-H_2O at different temperatures and at q=0.78 Å^{-1}. The spectrum in red corresponds to the experimental resolution function, obtained by a measurement of the same sample at 20 K, and was truncated below and above -4µeV and 4 µeV. Inset: zoom into the quasi-elastic spectra between -8 and -0.6 µeV for 230, 240, 250 and 260 K, highlighting the change at 240 K.
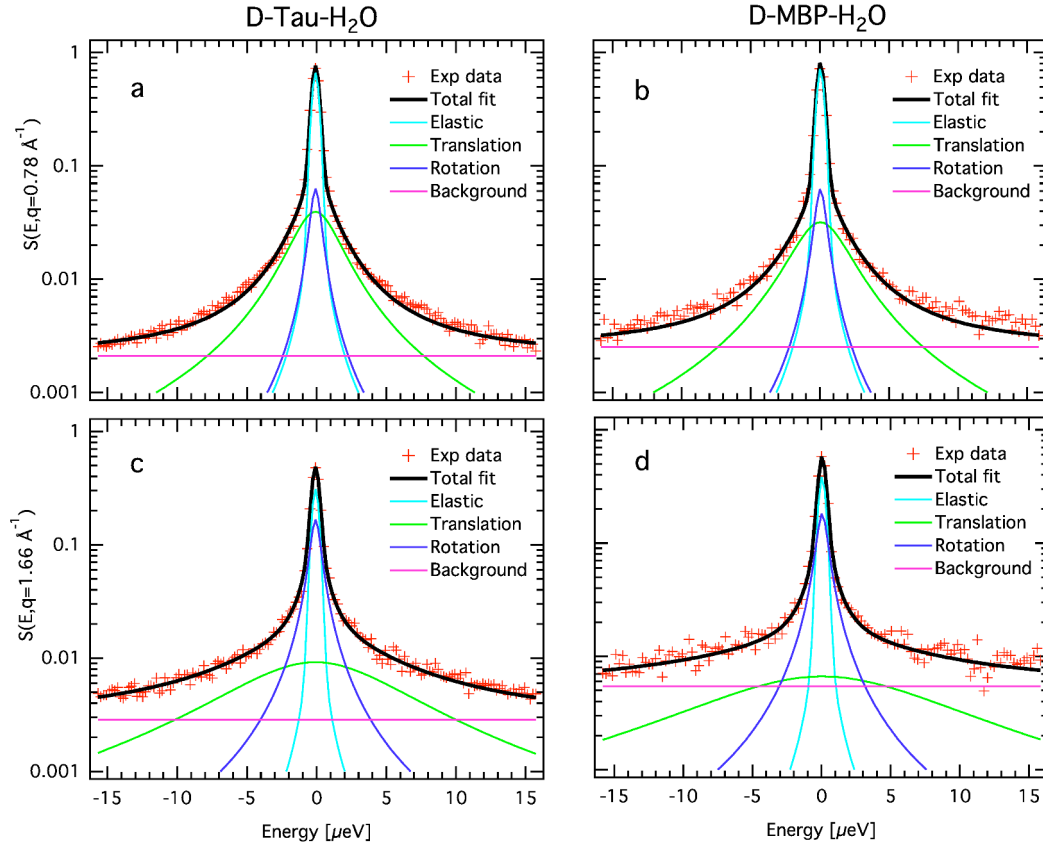
Figure 3. **Neutron spectra and fits.** Quasi-elastic neutron scattering spectra of D-tau-H_2O (a,c) and D-MBP-H_2O (b,d) at 260 K and for q=0.78 Å^{-1} (a,b) and q=1.66 Å^{-1} (c,d). The continuous lines represent the fitting curves with a model in which water molecules either translate, rotate or remain immobile (see main text).
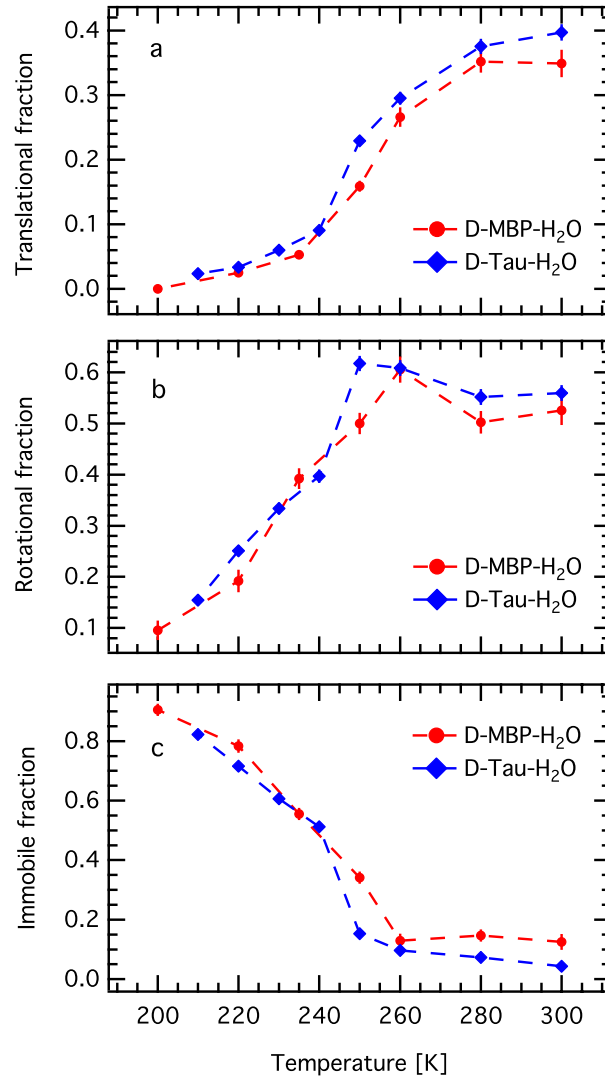
Figure 4. **Different populations of water molecules as a function of temperature.** Fractions of different dynamic contributions to the quasi-elastic spectra as a function of temperature: center-of-mass translation of water molecules (a), rotation of water molecules around their center-of-mass (b) and water molecules not moving in the dynamic window investigated (c). Dashed lines are guides to the eye. Red circles: MBP; blue diamond: tau
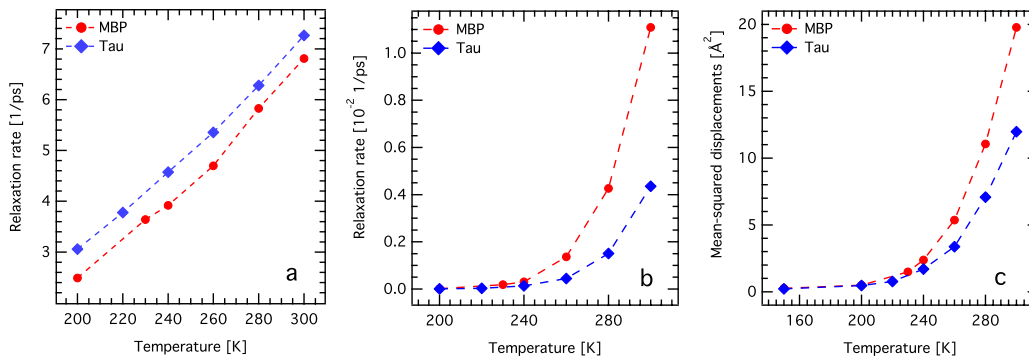
Figure 5. **Analysis of MD simulations of hydration water dynamics on tau and MBP surfaces**. (a) Continuous HB relaxation rates ($1/\tau_{HBC}$) as a function of temperature. (b) Intermittent HB relaxation rates ($1/\tau_{HBI}$) as a function of temperature. (c) Mean-squared displacements of hydration water oxygen atoms at 100 ps, as a function of temperature. Error bars on the relaxation rates and mean-squared displacements, estimated by computing the respective quantities separately over the two halves of the trajectory segments used for the analysis, are smaller than the plotting symbols.

References

1   Ball, P. Water as an active constituent in cell biology. *Chem Rev.* **108**, 74-108. (2008).

2   Rupley, J. A. & Careri, G. in *Advances in Protein Chemistry* Vol. 41 (eds Anfinsen C.B., Edsall J.T., Richards F.M., & Eisenberg D.S.) 37-172 (Academic Press, 1991).

3   Doster, W. & Settles, M. Protein-water displacement distributions. *Biochim Biophys Acta.* **1749**, 173-186. (2005).

4   Frauenfelder, H. *et al.* A unified model of protein dynamics. *Proc Natl Acad Sci U S A.* **106**, 5129-5134. (2009).

5   Gallat, F.-X. *et al.* Dynamical coupling of intrinsically disordered proteins and their hydration water: comparison with folded soluble and membrane proteins. *Biophys J* **103**, 129-136 (2012).

6   Khodadadi, S. *et al.* Dynamics of biological macromolecules: not a simple slaving by hydration water. *Biophysical journal* **98**, 1321-1326, (2010).

7   Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754-756 (1989).

8   Zaccai, G. How Soft Is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering. *Science* **288**, 1604-1607, (2000).

9   Becker, T., Hayward, J. A., Finney, J. L., Daniel, R. M. & Smith, J. C. Neutron frequency windows and the protein dynamical transition. *Biophys J* **87**, 1436-1444. (2004).

10   Khodadadi, S. *et al.* The origin of the dynamic transition in proteins. *J Chem Phys* **128**, 195106 (2008).

11   Doster, W. The dynamical transition of proteins, concepts and misconceptions. *Eur Biophys J.* **37**, 591-602. (2008).

12   Frauenfelder, H., Young, R. D. & Fenimore, P. W. Dynamics and the free-energy landscape of proteins, explored with the Mossbauer effect and quasi-elastic neutron scattering. *J Phys Chem B* **117**, 13301-13307 (2013).

13   Schiro, G., Natali, F. & Cupane, A. Physical origin of anharmonic dynamics in proteins: new insights from resolution-dependent neutron scattering on homomeric polypeptides. *Phys Rev Lett* **109**, 128102 (2012).

14   Ngai, K. L., Capaccioli, S. & Paciaroni, A. Change of caged dynamics at T(g) in hydrated proteins: trend of mean squared displacements after correcting for the methyl-group rotation contribution. *J Chem Phys* **138**, 235102, (2013).

15   Wood, K. *et al.* Coincidence of dynamical transitions in a soluble protein and its hydration water: direct measurements by neutron scattering and MD simulations. *Journal of the American Chemical Society* **130**, 4586-4587, (2008).

16   Doster, W. *et al.* Dynamical transition of protein-hydration water. *Phys Rev Lett* **104**, 098101 (2010).

17   Doster, W. The protein-solvent glass transition. *Biochim Biophys Acta* **1804**, 3-14, (2010).

18   Nickels, J. D. *et al.* Dynamics of protein and its hydration water: neutron scattering studies on fully deuterated GFP. *Biophys J* **103**, 1566-1575, (2012).

19   Chen, S. H. *et al.* Observation of fragile-to-strong dynamic crossover in protein hydration water. *Proc Natl Acad Sci U S A* **103**, 9012-9016, (2006).

20 Lehnert, U., Reat, V., Weik, M., Zaccai, G. & Pfister, C. Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light- induced conformational changes. *Biophys J* **75**, 1945-1952. (1998).

21 Kleinert, T., Doster, W., Post, F. & Settles, M. in *Proceedings of the Italian Physical Society.* (eds M.U. Palma, M.B. Palma-Vittorelli, & F. Parak) 127-130.

22 Ostermann, A., Waschipky, R., Parak, F. G. & Nienhaus, G. U. Ligand binding and conformational motions in myoglobin. *Nature* **404**, 205-208 (2000).

23 Tarek, M. & Tobias, D. J. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys Rev Lett* **88**, 138101. (2002).

24 Tournier, A. L., Xu, J. & Smith, J. C. Translational Hydration Water Dynamics Drives the Protein Glass Transition. *Biophysical journal* **85**, 1871-1875, (2003).

25 Dyson, H. J. Expanding the proteome: disordered and alternatively folded proteins. *Q Rev Biophys* **44**, 467-518, (2011).

26 Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **107**, 8183-8188 (2010).

27 Tompa, P. Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527-533 (2002).

28 Fink, A. L. Natively unfolded proteins. *Curr Opin Struct Biol* **15**, 35-41 (2005).

29 Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* **18**, 756-764, (2008).

30 Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021-1025, (2007).

31 Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome informatics. Workshop on Genome Informatics* **11**, 161-171 (2000).

32 Gaspar, A. M., Appavou, M. S., Busch, S., Unruh, T. & Doster, W. Dynamics of well-folded and natively disordered proteins in solution: a time-of-flight neutron scattering study. *Eur Biophys J* **37**, 573-582, (2008).

33 Jensen, M. R., Zweckstetter, M., Huang, J. R. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev* **114**, 6632-6660, (2014).

34 Gabel, F. *et al.* Protein dynamics studied by neutron scattering. *Q. Rev. Biophys.* **35**, 327-367 (2002).

35 Bellissent-Funel, M. C., Teixeira, J., Bradley, K. F. & Chen, S. H. Dynamics of Hydration Water in Protein *J. Phys. I* **2**, 995-1001 (1992).

36 Wood, K. *et al.* Coupling of protein and hydration-water dynamics in biological membranes. *Proc Natl Acad Sci U S A* **104**, 18049-18054, (2007).

37 Achterhold, K. *et al.* Dynamical properties of the hydration shell of fully deuterated myoglobin. *Physical Review E* **84**, 041930 (2011).

38 Abseher, R., Schreiber, H. & Steinhauser, O. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins* **25**, 366-378 (1996).

39 Bizzarri, A. R. & Cannistraro, S. Molecular dynamics simulation evidence of anomalous diffusion of protein hydration water. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **53**, R3040-R3043 (1996).

40   Tarek, M. & Tobias, D. J. Single-particle and collective dynamics of protein hydration water: a molecular dynamics study. *Phys Rev Lett* **89**, 275501. (2002).

41   Lagi, M. *et al.* The low-temperature dynamic crossover phenomenon in protein hydration water: simulations vs experiments. *J Phys Chem B* **112**, 1571-1575, (2008).

42   Heyden, M. & Tobias, D. J. Spatial dependence of protein-water collective hydrogen-bond dynamics. *Phys Rev Lett* **111**, 218101 (2013).

43   Tarek, M. & Tobias, D. J. The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-scattering Experiments. *Biophysical journal* **79**, 3244-3257, (2000).

44   Mandelkow, E. M. & Mandelkow, E. Tau in Alzheimer's disease. *Trends Cell Biol.* **8**, 425-427. (1998).

45   Wuttke, J. *et al.* SPHERES, Jülich's High-Flux Neutron Backscattering Spectrometer at FRM II. *Rev. Sci. Instr.* **83**, 075109 (2012).

46   Starr, F. W., Nielsen, J. K. & Stanley, H. E. Fast and Slow Dynamics of Hydrogen Bonds in Liquid Water. *Physical Review Letters* **82**, 2294-2297 (1999).

47   Sartor, G., Hallbrucker, A. & Mayer, E. Characterizing the secondary hydration shell on hydrated myoglobin, hemoglobin, and lysozyme powders by its vitrification behavior on cooling and its calorimetric glass-->liquid transition and crystallization behavior on reheating. *Biophys J* **69**, 2679-2694. (1995).

48   Awile, O., Krisko, A., Sbalzarini, I. F. & Zagrovic, B. Intrinsically disordered regions may lower the hydration free energy in proteins: a case study of nudix hydrolase in the bacterium Deinococcus radiodurans. *PLoS Comput Biol* **6**, e1000854. (2010).

49   Bokor, M. *et al.* NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys J* **88**, 2030-2037 (2005).

50   Qvist, J., Ortega, G., Tadeo, X., Millet, O. & Halle, B. Hydration dynamics of a halophilic protein in folded and unfolded states. *J Phys Chem B* **116**, 3436-3444, (2012).

51   Pal, S. K., Peon, J. & Zewail, A. H. Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc Natl Acad Sci U S A* **99**, 1763-1768, (2002).

52   Pieper, J. *et al.* Temperature- and hydration-dependent protein dynamics in photosystem II of green plants studied by quasielastic neutron scattering. *Biochemistry* **46**, 11398-11409 (2007).

53   Bee, M. *Quasi Elastic Neutron Scattering. Principles and Applications in Solid State Chemistry, Biology and Materials Science.* (Adam Hilger, 1988).

54   Zanotti, J. M., Bellissent-Funel, M. C. & Chen, S. H. Relaxational dynamics of supercooled water in porous glass. *Physical Review E* **59**, 3084-3093 (1999).

55   Sears, V. F. Theory of cold neutron scattering by homonuclear diatomic liquids: I. Free rotation. *Canadian Journal of Physics* **44**, 1279-1297 (1966).

56   Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802 (2005).

57   MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616, (1998).

58   Mackerell, A. D., Jr., Feig, M. & Brooks, C. L., 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum

    mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry* **25**, 1400-1415, (2004).

59   Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91**, 6269-6271, (1987).

60   Tarek, M. & Tobias, D. J. The dynamics of protein hydration water: a quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophysical journal* **79**, 3244-3257 (2000).

61   Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 4177-4189 (1994).

62   Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **103**, 4613-4621 (1995).

63   Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **97**, 1990-2001 (1992).

64   Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation* **6**, 121-142 (1991).

65   Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

66   Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995).

67   Schwalbe, M. *et al.* Predictive atomic resolution descriptions of intrinsically disordered hTau40 and alpha-synuclein in solution from NMR and small angle scattering. *Structure* **22**, 238-249 (2014).

68   Bagchi, B. Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem Rev* **105**, 3197-3219 (2005).

69   Luzar, A. & Chandler, D. Effect of environment on hydrogen bond dynamics in liquid water. *Phys Rev Lett* **76**, 928-931 (1996).

70   Frick, B. & Gonzalez, M. Five years operation of the second generation backscattering spectrometer IN16--a retrospective, recent developments and plans. *Physica B: Condensed Matter* **301**, 8-19 (2001).

3.2   Supplementary information
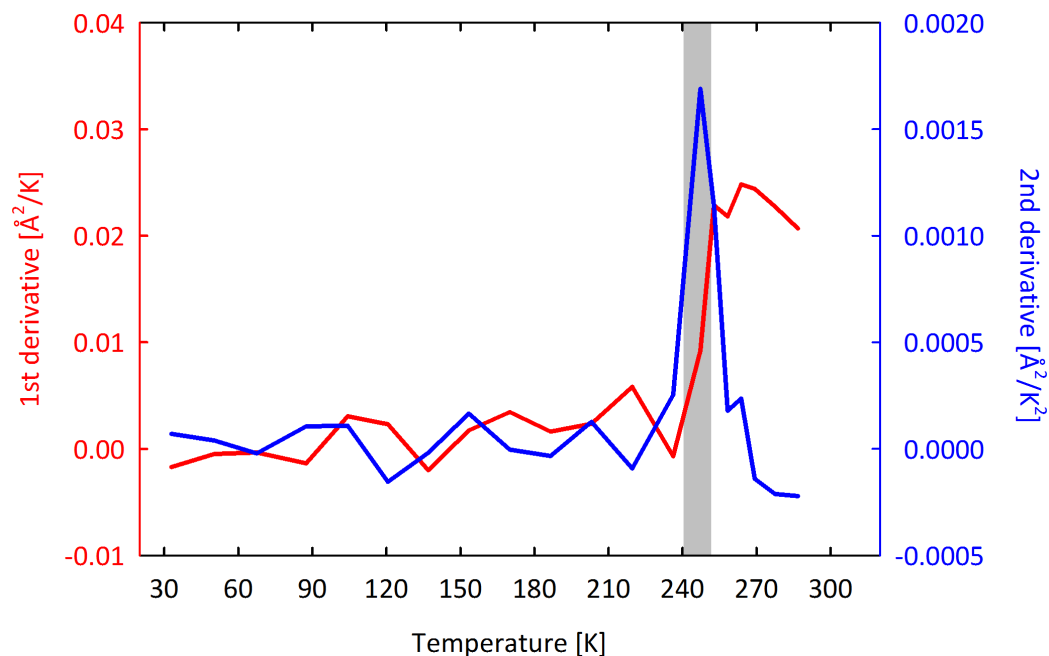
Supplementary Figure 1: **First (red curve, left axis) and second (blue, right axis) derivative of the MSD difference between the hydrated and the dry tau protein reported in the inset of Figure 1.** The presence of a step and a peak in the first and the second derivative, respectively, between 240 and 250 K (indicated by the grey area) defines the temperature region where the dynamical transition occurs.
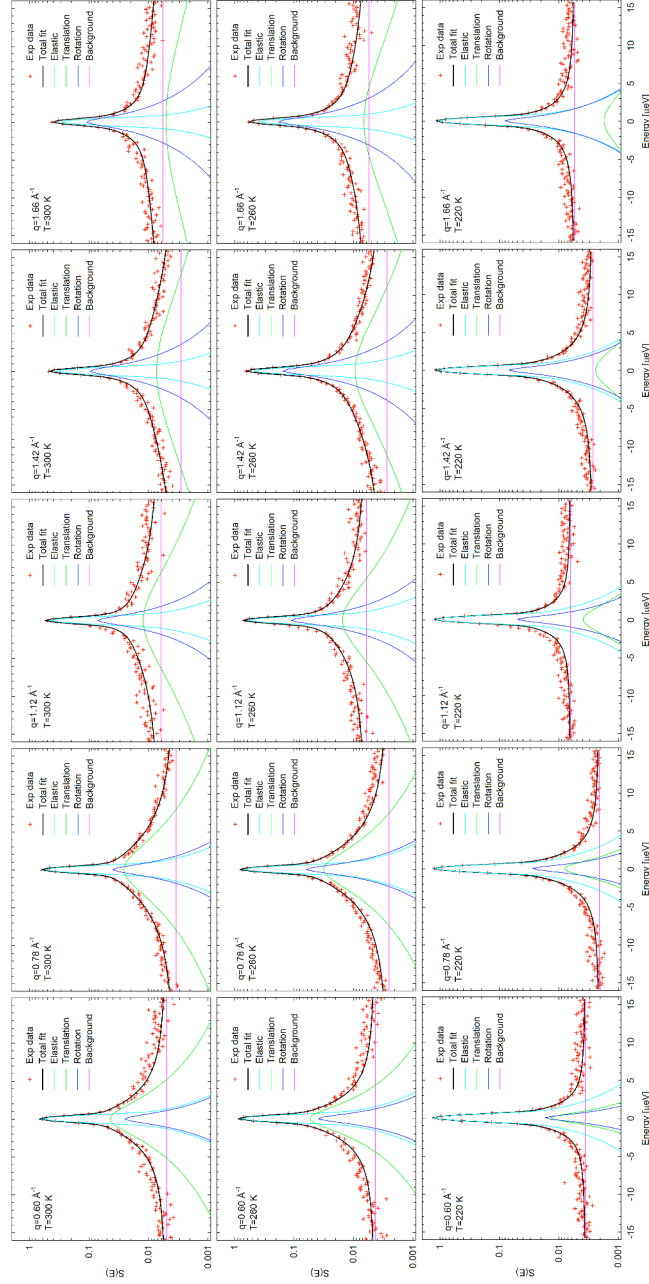
Supplementary Figure 2a: Neutron spectra and fits resulting from a model in which water molecules either translate, rotate or remain immobile. Quasi-elastic neutron scattering spectra of hydration water on the surface of D-tau-H_2O at 220, 260 and 300 K and for five different q-values. The continuous lines represent the fitting curves (see main text). Reduced χ^2 at 220, 260 and 300 K are 2.1, 1.7 and 1.6, respectively.
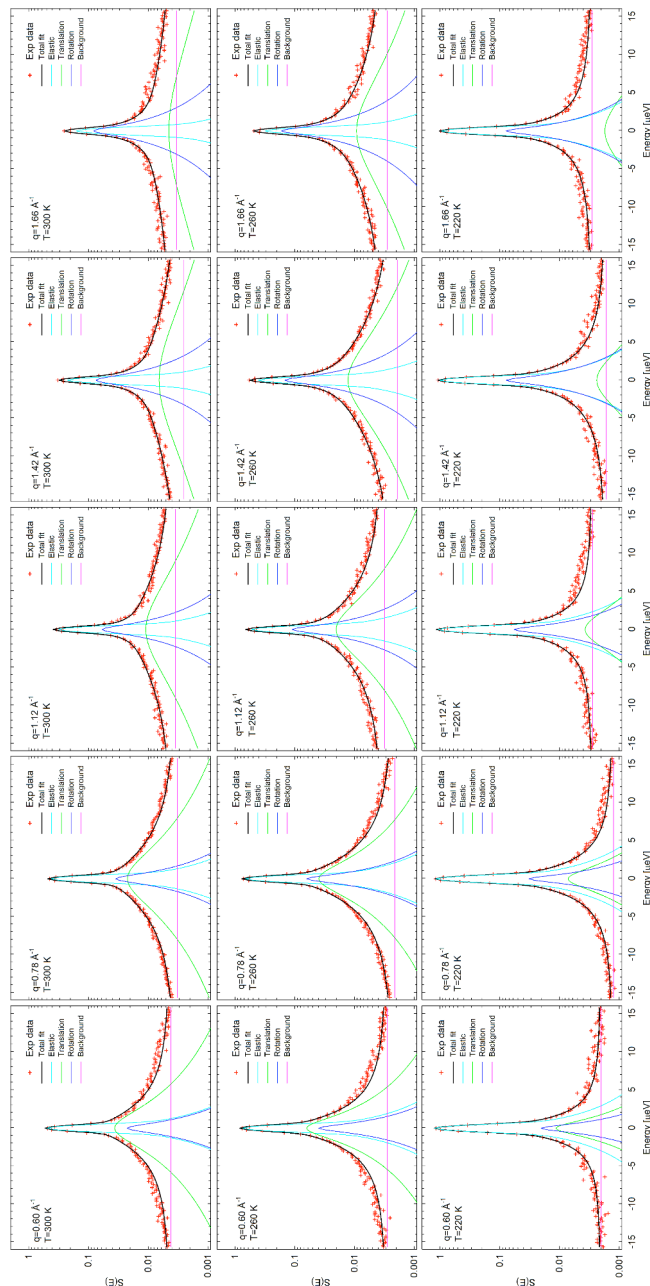
Supplementary Figure 2b: Neutron spectra and fits resulting from a model in which water molecules either translate, rotate or remain immobile. Quasi-elastic neutron scattering spectra of hydration water on the surface of D-MBP-H$_2$O at 220, 260 and 300 K and for five different q-values. The continuous lines represent the

fitting curves (see main text). Reduced χ^2 at 220, 260 and 300 K are 1.6, 1.4 and 1.4, respectively.
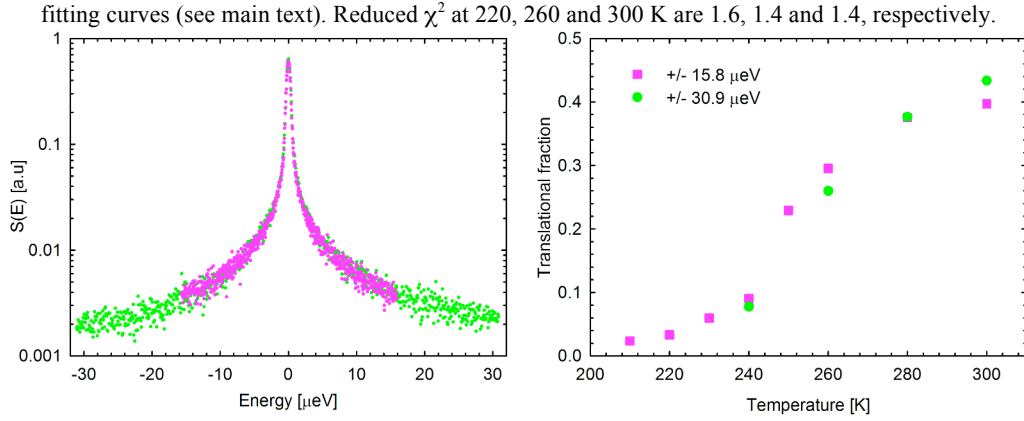
Supplementary Figure 3: **Effect of the energy range on data analysis.** Left panel: comparison of QENS spectra of D-tau-H_2O at 280 K, q=0.78 Å^{-1}, collected using two different energy ranges, +/- 30.9 μeV (green points) and +/- 15.8 μeV (magenta points). Right panel: fraction of the translational contribution as a function of temperature, relative to the analyses of D-tau-H_2O QENS data in the reduced (magenta squares) and extended (green circles) energy ranges.
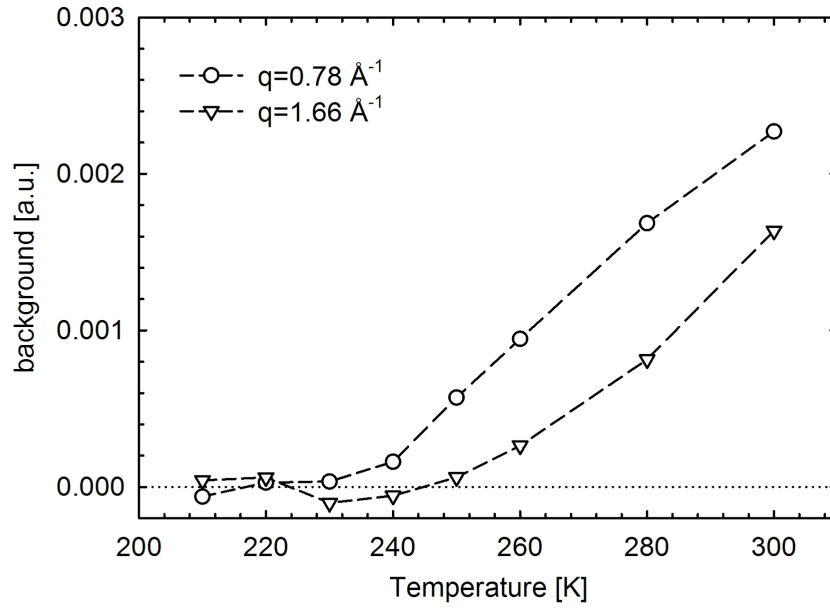
Supplementary Figure 4: Temperature dependence of the flat background obtained by the analysis of D-tau-H_2O data.
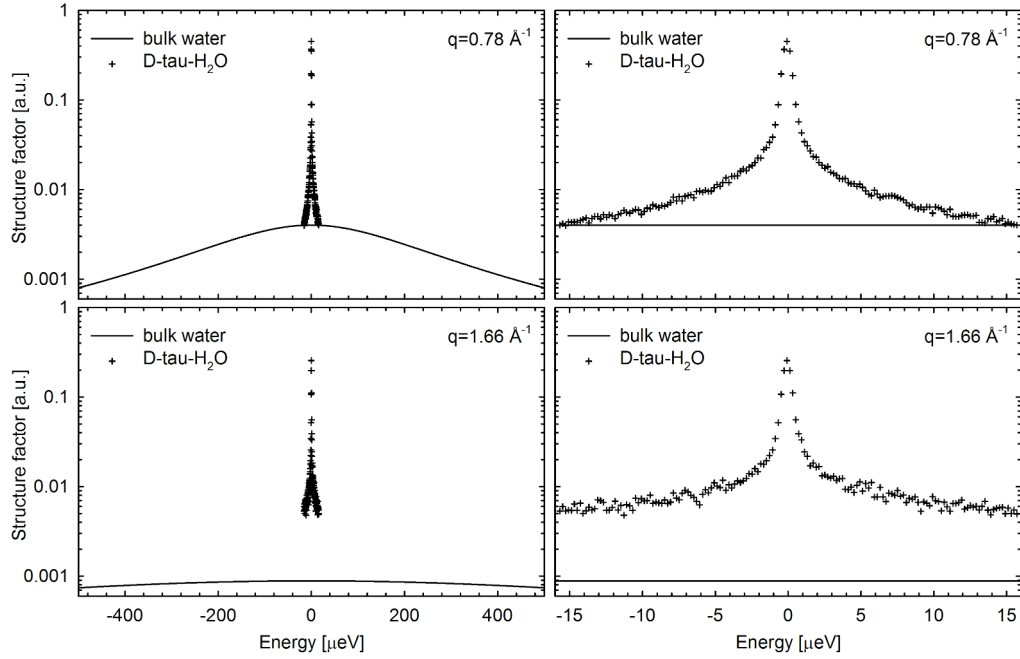
Supplementary Figure 5: A fast translational term is revealed on SPHERES as a flat background. D-tau-H_2O QENS spectra at T=300 K for two different q values: q=0.78 Å^{-1} (top panels) and q=1.66 Å^{-1} (bottom panels), superimposed on a water translational contribution calculated using a typical diffusion coefficient of bulk water (i.e. D=1×10^{-5} cm^{2}s^{-1}). The right panels are the same of the left plots zoomed in the SPHERES energy window.
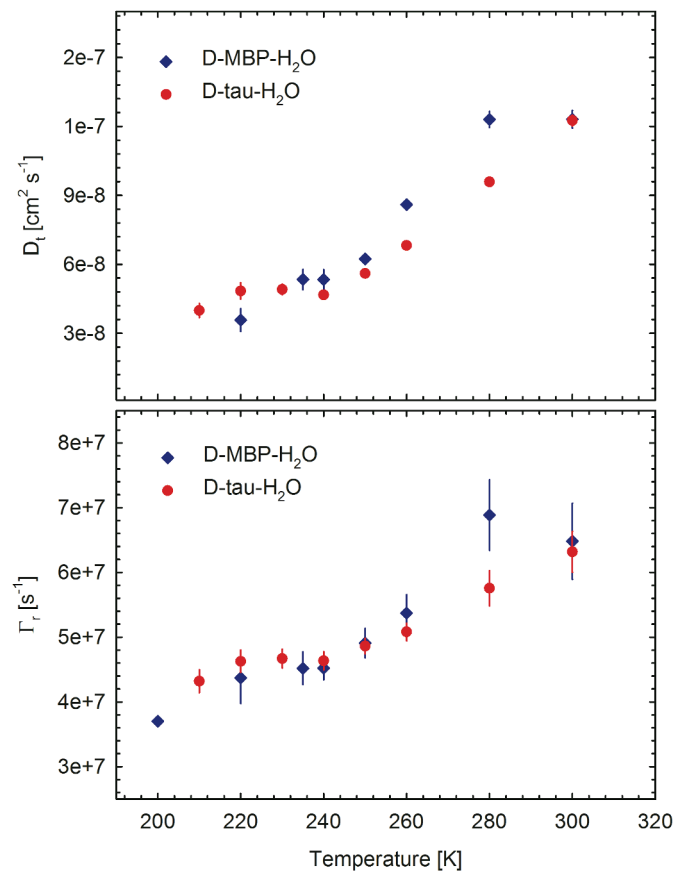
Supplementary Figure 6: Translational diffusion coefficients and rotational correlation rates of hydration water. Temperature dependence of the translational diffusion coefficients (top panel), and rotational correlation rates (bottom panel) obtained by the analysis of D-MBP-H_2O and D-tau-H_2O data.
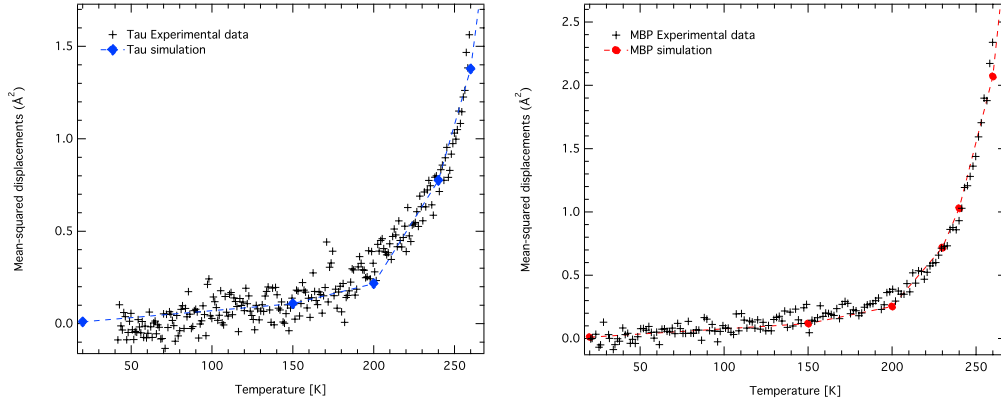
Supplementary Figure 7: Comparing MSD from neutron scattering and MD simulations. Mean-squared displacements of tau (left panel) and MBP (right panel) hydration water measured by elastic incoherent neutron scattering (black crosses; data are extracted from Gallat et al.[5] and Wood *et al.*[6]) and computed from the MD trajectories at 100 ps. The simulated MSD are scaled to the experimental values at 240 K.
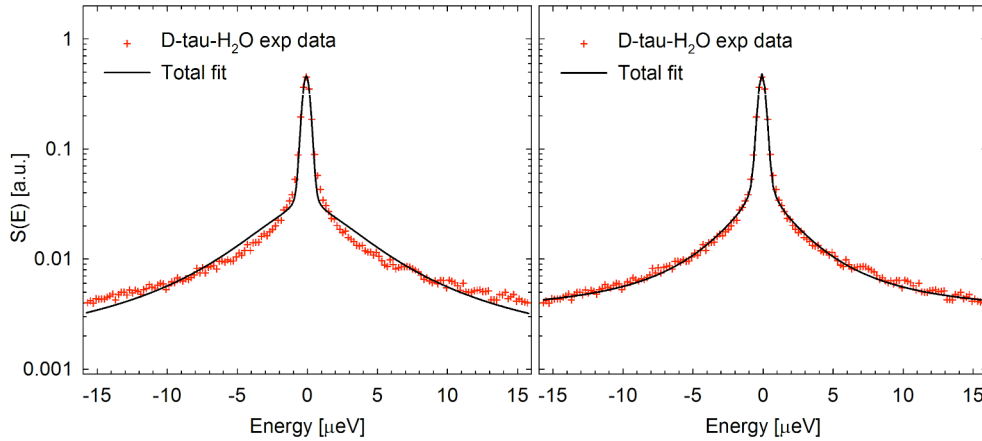
Supplementary Figure 8: Convoluting versus adding rotational and translational components. Results of fitting procedures using a convolution between translational and rotational contributions (left panel) and the model used in the present work, where the two contributions are summed (right panel). Data refers to the D-tau-H_2O sample at 300 K, q=0.78 Å^{-1}.
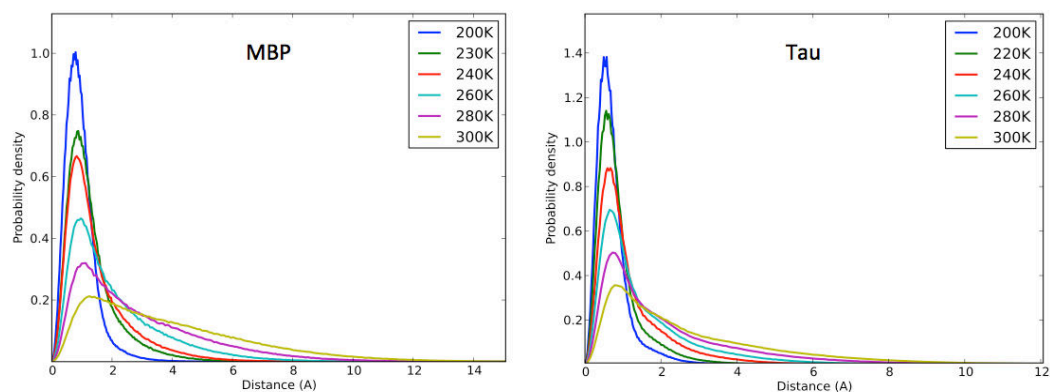
Supplementary Figure 9: Water displacement probability densities (4πr²G(r,t)). They are computed from MD simulations of MBP (left panel) and tau (right panel) powder models at t = 100 ps.
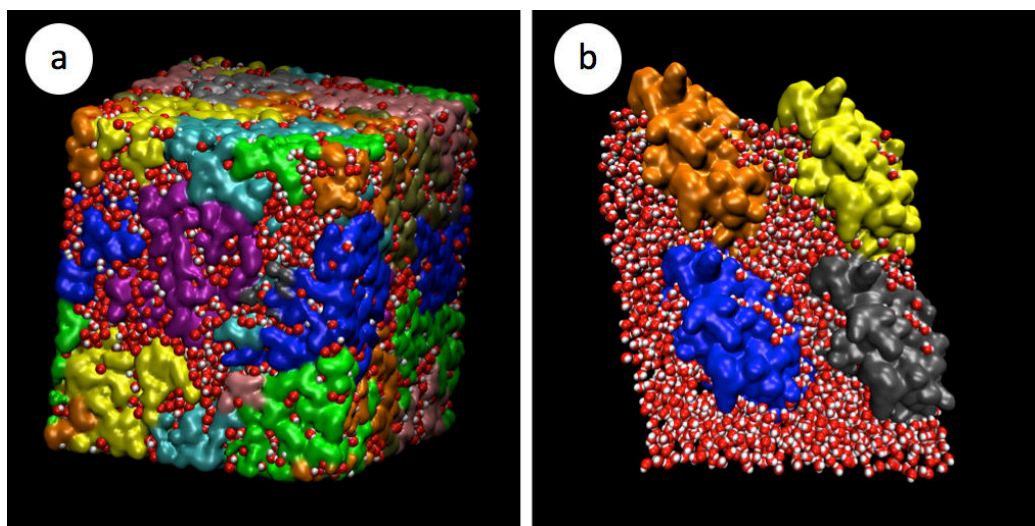
Supplementary Figure 10: Snapshots of the simulation boxes of the tau (a) and MBP (b) powder models. The protein molecules are drawn in surface representation, with different colors for each individual molecule. Water molecules are drawn with van der Waals spheres, with O atoms red and H atoms white.

Supplementary Discussion

Effect of the energy range on data analysis. In order to test a possible effect of the limited energy range on the data analysis, and in particular the influence of the background determination on the fitting parameters, we collected on SPHERES a set of QENS spectra of D-tau-H_2O at four selected temperatures (240, 260, 280, and 300 K) in the energy range +/- 30.9 μeV. As evidenced by the plots in Supplementary Figure 3 (left panel), the spectra are compatible with the data collected in the +/-15.8 μeV range, while a data analysis with the same model described above produces results perfectly in agreement with what we obtained in the reduced energy range. As a representative result, in Supplementary Figure 3 (right panel) we show a comparison between the temperature dependence of the translational fraction obtained in the two energy ranges, which confirms that the main results of our analysis are not affected by the choice of the energy range.

On the temperature dependence and physical origin of the background. In Supplementary Figure 4 we plot the constant background $k(q)$, obtained from D-tau-H_2O data analysis (eq. 1 in main text), as a function of temperature, at two different q values, *i.e.*, 0.78 and 1.66 Å^{-1}. The data points are vertically offset by subtracting the mean value of the points at temperatures below 240 K, in order to eliminate contributions originating from the instrumental q dependent background and from possible dynamic terms independent of (or barely dependent on) temperature. It is evident that a transition at about 240 K is present at both q values, yet the absolute value is larger at smaller q. The same trend (i.e. $k(q)$ decreasing with the increasing q) is observed at all q values. The observed change as a function of temperature reveals that a dynamical component is contained in the background outside of the instrumental contribution. What is the origin of this dynamical component? Since the quasi-elastic broadening in our data is essentially due to hydration water dynamics, the origin has to be due to water motions with a relaxation rate faster than the time window probed by SPHERES. As discussed in the main text, the rotational and translational components are characterized by different behaviors as a function of q. In particular, the rotational term has a width independent of q and an intensity that increases with the increasing q. A contribution with these characteristics would be detected by SPHERES as an increase of the background at high q. As a consequence, it is not compatible with the observed

q dependence of *k(q)* (Supplementary Figure 4). On the other hand, the translational term has an intensity independent of *q* and a width that increases with increasing *q*. Such a contribution would be detected by SPHERES as a reduction of the background at high *q*, which is compatible with the observed *q* dependence of *k(q)*.

The plots in Supplementary Figure 5 show how a fast translational term (compared to the SPHERES time window) is revealed by SPHERES as a flat background that decreases at high q. We calculated the line shape of a bulk water component (continuous line in Supplementary Figure 5), using a typical value for the room temperature diffusion coefficient (i.e. $D=1\times10^{-5}$ cm^2s^{-1}), and superimposed it to the experimental spectra collected at SPHERES (crosses in Supplementary Figure 5), at two different *q* values, i.e. 0.78 and 1.66 $Å^{-1}$. The intensity of the bulk-like contribution is set to an arbitrary value for clarity. The difference in the shape of the calculated translational term between $q=0.78$ $Å^{-1}$ and $q=1.66$ $Å^{-1}$ (Supplementary Figure 5 left panels) is due to the increase of its width Γ according to the relation $\Gamma=D\times q^2$ for Brownian diffusion. When zooming in the SPHERES energy window, the broadening of the calculated translational term is detected as a decrease of the background intensity (Supplementary Figure 5 right panels), in agreement with our experimental observation (Supplementary Figure 4).

Dynamical parameters and comparison with the literature. We report in Suppleme,ntary Figure 6 the temperature dependence of translational diffusion coefficient and rotational correlation rate of hydration water, obtained by the analysis of D-MBP-H_2O and D-tau-H_2O data. The translational diffusion coefficient shows that hydration water has a greatly reduced water mobility with respect to bulk water: the value at room temperature is about two orders of magnitude smaller than the corresponding bulk water value (1.2×10^{-7} cm^2s^{-1} for protein hydration water *vs.* 10^{-5} cm^2s^{-1} for bulk water). The same effect has been recently observed by M. Rosenstihl and M. Vogel using nuclear magnetic resonance on myoglobin and lysozyme powders hydrated at a similar hydration level (h=0.3). They found that in the temperature range 230 - 300 K, water diffusion is a factor of 30 - 100 slower than in bulk water[1]. Dielectric spectroscopy on hydrated protein powders is mainly sensitive to rotational relaxation of hydration water, due to the large permanent dipole moment of the water molecule. The time scale of the rotational component detected in our work is compatible with the relaxation time detected by dielectric spectroscopy on hydrated protein powders and usually attributed to hydration water dynamics (see e.g. [2-4]).

Comparison of water mean-squared displacements obtained from MD simulations with those extracted from elastic incoherent neutron scattering data. We computed the water hydrogen and exchangeable protein hydrogen mean-squared displacements (MSD) at different temperatures and extracted their values at 100 ps. The values of the MSDs at $t = 100$ ps showed a very similar temperature dependence to water MSDs determined by analysis of elastic incoherent neutron scattering data obtained previously on powders of D-MBP-H_2O and D-tau-H_2O powders.[5] However, the magnitudes of the MSDs at $t = 100$ ps were roughly an order of magnitude greater than the experimental values. The discrepancy is largely due to the differences in the way the MSDs are determined from simulations and neutron scattering experiments. The simulation values are computed directly from space-time trajectories, while the neutron values are obtained by fitting the q-dependence of the elastic intensity within a Gaussian approximation over a restricted range of q. Both quantities are measures of water mobility, but they cannot be directly compared. To enable a closer comparison of the temperature dependence of the MSDs, we scaled the values computed from the simulation trajectories at t = 100 ps to match the experimental value at 240 K. As can be seen in Supplementary Figure 7, the agreement between the scaled simulation values and the experimentally derived MSDs is very good for both proteins.

Accuracy of fitting with a sum or a convolution of rotational and translational water diffusion. The choice of the model function (see eq. 1 in the main text) used to fit the data is the result of a thorough exploration of different model functions based on the rotational-translational nature of water motions. First, we tried to fit the data with a single rotational-translational function where translational and rotational terms were convoluted (representing a single homogeneous population of water molecules). The statistical accuracy obtained was insufficient to justify using the model (Supplementary Figure 8 left panel), indicating the presence of more than one population of water molecules. Then we added a second rotational-translational term: in this case the model fitted the experimental data well but the number of independent parameters was too high to extract a meaningful trend in the fitting parameters. Finally, the choice of the sum of a translational and a rotational term (plus the other terms described in the main text) matched both the requirements of a high accuracy of fitting results (Supplementary Figure 8 right panel) and a small number of meaningful free parameters. The physical basis of such a model in terms of dynamical heterogeneity and finiteÒ instrumental

dynamical window is discussed in the main text. Briefly, the only scenario implying a single rotational-translational term is that, in the time window probed by SPHERES, a homogeneous contribution is observed, arising from one population of water molecules with the same dynamical properties. This is in disagreement with several experimental evidences about the heterogeneity of water populations and water motions on the protein surface. The finding that a model with two decoupled contributions is necessary to fit the quasi-elastic spectra of protein hydration water within the time window probed by SPHERES, reveals the presence of two contributions from distinct (or barely overlapping) water populations: one very slow and predominantly rotational (probably from hindered molecules strongly interacting with protein surface or in docking sites) and the other one predominantly translational (probably from more free water molecules) whose rotational term is larger than the SPHERES energy window.

Water displacement distributions from MD simulations reveal the existence of two populations of hydration water molecules with different dynamics. The global fitting procedure used to analyze the neutron scattering data employed a model in which it was assumed that there are distinct populations of water molecules undergoing primarily librational/rotational or translational dynamics. To assess the validity of this assumption, we used the MD simulation trajectories to calculate displacement distribution functions, $4\pi r^2 G(r,t)$, for the H atoms of the water molecules. $G(r,t)$ is the van Hove correlation function, the time and space Fourier transform of which is the dynamic structure factor probed by neutron spectroscopy, and $4\pi r^2 G(r,t) dr$ is the probability that a tagged particle at $t = 0$ has been displaced by a distance r at time t. The displacement distribution function, $4\pi r^2 G(r,t)$, is the corresponding probability density. The displacement distribution functions for protein/membrane hydration water are qualitatively different from those of bulk water[7]. As expected for Brownian particles, in bulk water under ambient conditions the displacement distribution is a unimodal function whose mean moves to larger distances r and width increases with increasing time t. In protein/membrane hydration water, the displacement distributions exhibit the existence of two classes of water molecules, the populations of which change with temperature, as exemplified by the distributions for water H atoms plotted in Supplementary Figure 9 for MBP and tau hydration water in model powders at $t = 100$ ps. At 200 K, the distributions are dominated by a sharp peak at ~1 Å, which corresponds to water

molecules librating and rotating but translationally confined to a solvation cage consisting of surrounding water and protein atoms. As the temperature is increased, the peak corresponding to translationally retarded, librating/rotating water molecules shrinks, while a shoulder, corresponding to water molecules escaping their solvation cages *via* translational diffusion, grows and extends to larger *r*.

Supplementary References

1   Rosenstihl, M. & Vogel, M. Static and pulsed field gradient nuclear magnetic resonance studies of water diffusion in protein matrices. *J Chem Phys* 135, 164503 (2011).
2   Khodadadi, S. *et al.* The origin of the dynamic transition in proteins. *J Chem Phys* 128, 195106 (2008).
3   Schiro, G., Cupane, A., Vitrano, E. & Bruni, F. Dielectric relaxations in confined hydrated myoglobin. *J Phys Chem B* 113, 9606-9613, (2009).
4   Jansson, H., Bergman, R. & Swenson, J. Relation between solvent and protein dynamics as studied by dielectric spectroscopy. *J Phys Chem B* 109, 24134-24141, (2005).
5   Gallat, F.-X. *et al.* Dynamical coupling of intrinsically disordered proteins and their hydration water: comparison with folded soluble and membrane proteins. *Biophys J* 103, 129-136 (2012).
6   Wood, K. *et al.* Coincidence of dynamical transitions in a soluble protein and its hydration water: direct measurements by neutron scattering and MD simulations. *Journal of the American Chemical Society* 130, 4586-4587, (2008).
7   Tobias, D. J., Sengupta, N. & Tarek, M. Hydration dynamics of purple membranes. *Faraday Discuss* 141, 99-116 (2009).

3.3   Summary and lead-in for the next chapter

Summary   Hydration water is the natural matrix of biological macromolecules and is essential for molecular activity in cells. The coupling between water and protein dynamics has been intensively studied for more than a decade, yet remains controversial. Here we combined protein perdeuteration, quasi-elastic neutron scattering and all-atom molecular dynamics simulations to explore the nature of hydration water motions at temperatures between 200 and 300 K, across the so called protein dynamical transition, in the intrinsically disordered human protein tau and the globular maltose binding protein. Quasi-elastic broadening is fitted with a model of translating, rotating and immobile water molecules. In both experiment and simulation, the translational component markedly increases at the protein dynamical transition (around 240 K), regardless of whether the protein is intrinsically disordered or folded. Thus, we generalize the notion that the translational diffusion of water molecules on a protein surface promotes the large-amplitude motions of proteins that are required for their biological activity.

Introduction to the next chapter   The results presented in this chapter suggest that both IDPs and globular proteins share a similar hydration coupling, *i.e.* the requirement of water translational diffusion in order to allow large-amplitude biologically-relevant protein motions to occur. Rotational and translational properties of water molecules, probed by incoherent neutron scattering and MD simulations, are single-particle dynamical features that are averaged over all atoms in the system. Another aspect of particle dynamics that can be analyzed is how atoms move with respect to each other, *i.e.* what is their collective dynamics. Both collective and single particle hydration water dynamics are suspected to be biologically relevant, although their properties are quite different (Tarek and Tobias, 2002b; Bagchi, 2005). In order to investigate whether or not IDPs and globular proteins have different influences on their hydration water, we compare in the next chapter the collective hydration water dynamics of the tau protein and the MBP, using THz spectroscopy.

Chapter 4

Comparing collective water dynamics around intrinsically disordered and globular proteins

The manuscript presented herein is to be submitted to the *Biophysical Journal* shortly.

4.1   Manuscript

Yann Fichou[1,2,3], Gudrun Niehues[4], François-Xavier Gallat[1,2,3], Valeria Conti Nibali[5], Aude Vernet[1,2,3], Pau Bernado[6], Matthias Heyden[7], Martin Blackledge[1,2,3], Giuseppe Zaccai[1,2,3,8], Douglas Tobias[9], Martin Weik[1,2,3], Martina Havenith[5]

[1]Univ. Grenoble Alpes, IBS, F-38044 Grenoble, France; [2]CNRS, IBS, F-38044 Grenoble, France; [3]CEA, IBS, F-38044 Grenoble, France; [4]Research Center for Development of Far-Infrared Region, University of Fukui, Fukui 910-8507, Japan; [5]Department of Physical Chemistry II, Ruhr-University Bochum, 44780 Bochum, Germany; [6]Centre de Biochimie Structurale, INSERM U1054, CNRS, UMR 5048, Université Montpellier 1 and 2, Montpellier, France; [7]Max Planck Institut fur Kohlenforschung, Mulheim, Germany; [8]Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France; [9]Department of Chemistry, University of California, Irvine, California, USA

Abstract

Water is the natural matrix of biological macromolecules and as such is essential for a variety of molecular processes, including protein folding, misfolding and stability. Intrinsically disordered proteins (IDPs) interconvert among multiple extended conformations rather than remaining in a well-defined folded structure and are prone to aggregation and fibrillation. The intrinsically disordered protein tau has been shown to display a different coupling to hydration water dynamics than globular proteins, which is modified when tau forms amyloid fibers involved in Alzheimer disease. Here, we focus on the collective dynamics of hydration water of tau in comparison to that of the folded globular maltose binding protein (MBP). By carrying out THz spectroscopy experiments, we found that the volume of water influenced by a tau protein molecule is two times less important than that influenced by a MBP molecule. The dynamical and structural differences between IDPs and folded proteins might thus be correlated with differences in collective hydration water dynamics.

Introduction

Intrinsically disordered proteins (IDP) lack a well-defined three-dimensional structure in their native state, are highly flexible and continuously change conformations in solution (1, 2). IDPs are ubiquitous throughout the proteome in all three kingdoms of life (3) and fulfill are large variety of biological roles that benefit from the intrinsic protein flexibility, such as e.g. the capability to interact with several partners (4). Due to their intrinsic disorder, IDPs display an inherent propensity to aggregate, which is at the origin of their involvement in a variety of diseases, mostly neurodegenerative disorders (5). The intrinsically disordered human protein tau, e.g., can form so-called neurofibrilar tangles that are among the pathological hallmarks of Alzheimer disease (6). In its non-aggregated, functional state, the tau protein regulates microtubule stability in neuronal cells. Its longest isoform, studied here, is the most common isoform (htau40) and is composed of 441 amino acids. Fibrilation of htau40 has been shown recently to be accompanied by an increase in the mobility of hydration water at the protein surface (7, Fichou *et al.* under review).

Water is an essential component of macromolecular biological activity, including protein folding and stability, ligand binding, and enzyme reactions (8). It is the unique property of water to form a dense hydrogen bond network on a macromolecular surface that

plays a key role in assisting biological function (9) and that makes hydration water a reservoir of entropy for protein fluctuations (10). Hydration water has been demonstrated to enable the large-amplitude motions that are required for protein functions (11, 12), and *vice versa* protein dynamics has been shown to influence hydration-water dynamics (13, 14). In particular, a gradient of protein-water dynamical coupling among different classes of proteins has been suggested to exist, from weak coupling for membrane proteins to tighter coupling for soluble proteins, including globular and intrinsically disordered proteins (14). A complete understanding of biological hydration-water dynamics remains elusive, however, in particular the perturbation of collective water motions in the vicinity of macromolecules only begins to be uncovered (15). Several experimental and computational techniques are sensitive to collective water motions, including MD simulations (16), neutron scattering (17) and THz spectroscopy (15, 18, 19).

THz spectroscopy is a unique tool to probe collective water motion on the sub-picosecond timescale. In particular the absorption at around 80 cm^{-1} (2.4 THz) was shown to correspond to fluctuations of the water hydrogen bond network delocalized over several water molecules (20). These collective modes are perturbed in the vicinity of biomolecules and an extended dynamical hydration shell of several layers has been suggested to exist around proteins (18). Our table-top p-Ge THz laser difference spectrometer is particularly adapted to study THz fluctuations because of its high sensitivity to subtle changes of absorption in the THz range (21). Recent studies have suggested that hydration water dynamics in the THz domain assist enzyme catalysis and that they are substrate dependent (15, 22, 23). Furthermore, Ebbinghaus and coworkers have shown that the THz perturbation of hydration water essentially disappears when a protein is unfolded, either chemically or by mutagenesis (24). The collective water dynamics on the surface of an IDP, however, has not been studied yet and might differ from the one of globular or intentionally unfolded proteins.

Here we used THz absorption spectroscopy to study the collective dynamics of hydration water around the intrinsically disordered protein tau and compared it to that of the maltose binding protein (MBP), a globular protein with a similar molecular mass than tau. In order to analyze the experimental results, we developed a Monte Carlo like simulation using the conformational ensemble of tau previously determined by NMR spectroscopy and small angle scattering. We find that a single tau protein perturbs the collective dynamics of two times less water than MBP. This result might be explained by a weaker short-range protein-

water interaction for tau than for MBP. We propose that the perturbation of a limited amount of water surrounding IDPs, leading to a limited hydration funnel effect, might compensate for the large radius of capture inherent to their geometry.

Materiel and Methods

Protein expression and purification

The htau40 isoform of the human tau protein was expressed and purified as described previously (14). Briefly, the protein was expressed as a histidine-tagged fusion protein in *Escherichia coli* BL21(DE3), which grew in two-liter flasks of LB medium. The protein was purified by nickel affinity chromatography, followed by size-exclusion chromatography. The maltose binding protein (MBP) was expressed and purified as described in (25). Briefly, the protein was expressed as a histidine-tagged fusion protein in *Escherichia coli* BL21(DE3) using high cell density cultures. The purification of MBP was performed in a unique step of cobalt affinity chromatography. Following purification, both protein samples were in 0.5 M ammonium acetate buffer (pH 7) for THz measurements. For THz measurements, the protein solutions were concentrated to the highest possible concentration at which the protein was still soluble (80 mg/ml for tau; 110 mg/ml for MBP). The absence of protein aggregation after concentration was ascertained by dynamic light scattering.

THz absorption spectroscopy measurements

For THz measurements, the protein solutions were diluted to the various desired values and passed through a filter with pores 0.22 μm in diameter. For each diluted sample, the protein concentration was verified immediately before the THz measurement measurement by monitoring the absorption at 280 nm. The molar extinction coefficients used were 64720 and 6640 $M^{-1}.cm^{-1}$ for tau and MBP, respectively.

THz absorption spectroscopy measurements were carried out using our unique, table top p-Ge THz laser difference spectrometer (21). The integrated THz absorption coefficients of the investigated samples were measured in the spectral range between 2.1 and 2.8 THz (73–93 cm^{-1}). The difference spectrometer has a double-beam configuration (two channel setup) that allows simultaneously measuring the sample and the buffer solution (0.5 M ammonium acetate, pH 7) as a reference under identical conditions (temperature and humidity). The humidity in the experimental chamber was controlled to be < 3%, and the temperature fixed

to 293 ± 0.05 K. For the measurements, we used a commercial sample cell with z-cut quartz windows and Teflon spacers for a fixed layer thickness of 50 μm.

In general, the transmitted intensity *I(n)* and hence, the absorption of a sample can be described by Beer's law:

$$I(n) = I_0 \exp(-\alpha(n)d) + c$$

where I_0, $\alpha(n)$, d, and c correspond to the intensity of the laser source, the absorption coefficient of the sample, the layer thickness of the sample, and the detector offset, respectively.

For both proteins, three THz absorption measurement series were carried out, based on three different protein solutions that were expressed, purified and concentrated independently. For the tau protein, a first series covered the concentration range from 1 to 80 mg/ml, a second from 1 to 40 mg/ml, and a third from 20 to 50 mg/ml. MBP series were carried out from 1 to 80, 1 to 30 and 20 to 80 mg/ml. Measurements from different series at similar concentrations yielded similar THz absorbances, prompting us to merge data from the three series for tau and MBP.

Three-component model analysis

The absorption curves were fitted with a three-component model, which was introduced earlier (26). In this model, the total absorbance α_{total} of the solution at a concentration *c* can be expressed as follows:

$$\alpha_{total}(c) = \frac{V_{prot}(c)}{V_{total}} \times \alpha_{prot} + \frac{V_{shell}(R,c)}{V_{total}} \times \alpha_{shell} + \frac{(V_{total} - V_{prot}(c) - V_{shell}(R,c))}{V_{total}} \times \alpha_{bulk}$$

where V_{prot} and V_{shell} are the volume of the protein fraction and the volume of the dynamical hydration shell, respectively; α_{prot}, α_{shell} and α_{bulk} are the absorption of the protein, the hydration water and the bulk water, respectively; *R* is the extension of the dynamical hydration shell.

The absorption of bulk water was set to 420 cm^{-1}. An empirical value of the partial volume of a protein of 0.73 cm^3/g (27) is commonly used. Several studies have shown that a substantial number of water molecules "tightly bound" to the protein surface share the same

dynamical properties in the THz domain than the protein itself (28, 29). They found about 0.22 g of such water molecules per gram of protein. In our model we took V_{prot} as a sum of the volumes of the protein and the "tightly bound" water. Because tau is intrinsically disordered, its solvent-accessible surface area is about three times larger than the one of MBP (14). Therefore, we included three times more water molecules in the volume of the tau protein than in the volume of the MBP. It led us to set V_{MBP} to 64400 Å³ and V_{tau} to 88700 Å³ per protein. The absorption coefficients of the proteins were extracted from a linear fit of the absorption at high concentration (see figure 1). We extrapolated the line to the theoretical high concentration limit, where only the protein component would be present, *i.e.* 1050 mg/ml for the MBP and 858 mg/ml for the tau protein, and extract the absorbance from the intersection with the y-axis. α_{MBP} and α_{Tau} were found to be 200 cm⁻¹ and 110 cm⁻¹ for the MBP and the tau protein, respectively. α_{shell} and $V_{shell}(R)$ are the two parameters that are varied in the fitting procedure. The expression of $V_{shell}(R)$ was numerically determined by a Monte Carlo like simulation outlined in the next section.

The goodness of the fit is evaluated by calculating the least square X such as

$$\chi = \sum_{C} \left[\alpha_{fit}(c) - \alpha_{\exp}(c) \right]^2$$

where $\alpha_{fit}(c)$ and $\alpha_{exp}(c)$ are the absorption predicted by the model and measured at a concentration c, respectively. X is displayed in the color map of the figures 3b and 4b as a function of the parameters R and α_{shell}.

Monte-Carlo like simulation

Ten protein molecules (tau or MBP), from which the side chains were stripped of, were randomly placed, one after the other, in a cubic box, whose dimensions were set according to the protein concentration chosen. Periodic boundary conditions were applied. Any addition of a protein molecule that led to an overlap with an already positioned protein was discarded and the molecule placed at a new random location. The proteins were considered to overlap when at least one of their backbone atoms was closer than a distance D to a backbone atom of a neighboring molecule. D was set to 7.9 Å and 8.6 Å for tau and MBP, respectively, in order to account for a steric exclusion from an average side chain of 3.95 Å (tau) and 4.3 Å (MBP) that includes the tightly bound water mentioned in the previous

section, such that the total volume per protein molecule was 64400 and 88700 Å^3, for MBP and tau, respectively. The reason why D was set to different values for the two proteins is that the backbone atoms of MBP included N-C_α-C-O (4 atoms), whereas the backbone of the tau included N-C_α-$C_\beta H_2$-C-O (7 atoms).

The conformation of the 10 MBP molecules was that of the crystal structure (PDB access code 1JW4). Concerning tau, a sub-set of ten conformations were selected for which the radius of gyration matched the experimental value of 62 Å measured in solution (14, 30) were selected from an ensemble of conformations determined by NMR and SAXS (31).

Each voxel, defined as a 5x5x5 Å^3 cube, of the protein box was then identified to be either part of the protein, of the hydration water component or of the bulk water component. The assignment to one of these groups was done as follows. The voxels that were closer than the distance *D* from any protein backbone atom were considered part of the protein. The voxels that were between *D* and *R*, were assigned to hydration water, with R being thus the extension of the dynamical hydration shell. Finally the other voxels that did not fulfill one of these two conditions were considered as belonging to bulk water. Every volume calculation was an average over 50 random box configurations, which were sufficient to converge statistically.

Molecular dynamics simulations

The molecular dynamics (MD) simulations were performed with the NAMD program (32) using the CHARMM27 force field (33, 34) for the protein and the TIP3P model (35) for water. The simulations were maintained at constant temperature using Langevin dynamics, and a constant pressure of 1 atm using the Nosé-Hoover Langevin piston algorithm with anisotropic cell fluctuations (36, 37). The equations of motion were integrated using the Verlet-I/r-RESPA multiple-time step algorithm (38, 39) with time steps of 4 fs for the long-range non-bonded forces, 2 fs for the short-range non-bonded forces, and 2 fs for the bonded intra-molecular forces. The SHAKE algorithm (40) was used to constrain the lengths of all bonds to H atoms. Electrostatic interactions were computed using the smooth particle-mesh Ewald sum (41), and the Van Der Waals interactions and the real-space part of the Ewald sum were smoothly switched to zero over the range 10-12 Å.

The MBP solution was modeled by placing one MBP molecule (PDB access code 1JW4) in a box of 27823 water molecules with periodic boundary conditions. An energy

minimization was run for 2 ps before equilibrating the system until the RMSF of the protein had stabilized. Several nanoseconds of additional trajectories saved every 0.1 ps were then produced for analysis. The tau protein was modeled by placing each of the ten tau conformations, selected as described in the section entitled *Monte-Carlo like simulation*, in different parallelepiped boxes of water. The dimensions of the simulation boxes were set so that the minimum distance between the protein and the box boundaries was 15 Å. For each of the ten boxes, an energy minimization was run for 1 ps before equilibrating the system until the number of protein-protein hydrogen bonds had stabilized. Several nanoseconds of additional trajectories saved every 0.1 ps were then produced for analysis. The dynamical parameters of the tau protein shown in figure 5 are averaged over the ten simulation boxes. The dynamical parameters of bulk water were obtained from a box of pure water containing 233 water molecules.

The protein-water hydrogen bond (HB) relaxation time, which reflects the diffusion of water on the protein surface, or exchange of water molecules between hydration shells, was defined by the time correlation function $C(t) = <h(0)h(t)>/<h>$, where $h(t)$ is equal to one if a given donor-acceptor (D-A) pair is hydrogen bonded at time t and zero otherwise, and the angular brackets denote an average over all D-A pairs and time origins. A pair was considered to be hydrogen bonded when the distance D-A was less than 3.5 Å and the D-H-A angle was greater than 150°. The function $C(t)$ is the probability that a random D-A pair that is hydrogen bonded at time zero is still bonded at time t, regardless of whether or not the bond was broken at intermediate times.

Results

THz spectroscopy was employed to investigate the extend to which hydration water is dynamically perturbed in proximity of the IDP tau in comparison to the globular MBP, both proteins ahving about the same molecular weight (45849 Da for the tau protein and 40707 Da for the MBP). Figure 2 shows the difference, *Δα*, in absorbance of the protein solutions compared to bulk buffer as a function of protein concentration. Assuming a two-component model with protein and only bulk water, an increase in protein concentration is expected to result in a linearly decreasing absorption coefficient. However, at low protein concentration, we observe for both protein solutions an initial non-linearity, the so-called *terahertz excess*, before the absorption coefficient changes linearly as a function of protein concentration. This

excess has been attributed to the presence of protein hydration water, which is dynamically perturbed and absorbs stronger than bulk water in the THz range. A linear behavior of the absorption as a function of the protein concentration is regained when the volume of bulk water is significantly reduced, thus leaving only two components contributing to the total absorption, the hydration water and the protein. In the following, we analyze data in figure 2 either qualitatively or by employing a three-component model with the aim to estimate and compare the hydration-water volumes of tau and MBP.

Qualitative estimate of tau versus MBP hydration water volumes

The protein concentration at which a linear behavior is reached in figure 2 allowed us to estimate the volume of hydration water per protein molecule. The colored zones (figure 2) highlight concentration windows in which the linear regime is reached, *i.e.* between 6 and 9 mg/ml for MBP, and between 20 and 30 mg/ml for the tau protein. In order to estimate a lower limit of the difference in hydration-water volumes of the two proteins, we considered the linear regimes are reached at 9 and 20 mg/ml for MBP and tau, respectively. Above these concentrations, the remaining amount of bulk water can not be detected anymore, so that a linear regime is observed in which only the amount of protein and hydration water changes as a function of protein concentration. Although this uncertainty is difficult to estimate, a reasonable assumption is that it is similar for the two proteins solutions that are evaluated under identical experimental conditions.

In order to estimate the volume of hydration water per protein molecule that results in the same amount of bulk water in MBP and tau solutions at 9 and 20 mg/ml, respectively, we carried out Monte Carlo like simulations. Such a simulation has been carried out previously for globular proteins, by approximating each molecular by a sphere (18). Since this approximation is obviously not appropriate for IDPs, we extracted ten tau conformations from a larger ensemble determined by NMR spectroscopy and small angle scattering (31) and used them for the Monte Carlo like simulation (see Material and Methods for details). The protein solution was modeled by randomly placing each tau conformation in a cubic box with periodic boundary conditions, for which the dimensions were set to represent the chosen concentration (see snapshot in figure 1a). Once the ten proteins were placed in the box, each voxel was assigned either to the protein, to the hydration water (within R Å away from any protein molecule) or to the bulk water (see material and methods for more details). The same

procedure was followed for the MBP, except that ten identical conformations were used (PDB entry code 1JW4) as shown in figure 1c. The hydration water volumes of hydration and bulk water as a function of protein concentration for several extensions R are shown in figures 1b and d. If we assume an arbitrary, but realistic, residual proportion of bulk water of 25%, figures 1b and 1d indicate that the extension R of the dynamical hydration shell that fulfills this criteria in a MBP solution at 9 mg/ml and in a 20 mg/ml-tau solution are 96 and 46 Å, respectively. In other words, R is 2.1 times smaller around a tau than around an MBP molecule. The corresponding hydration-water volumes per protein are about 1.8 times smaller for tau than for MBP. Up to residual bulk water of 40%, the hydration-water volume per tau molecule is always 1.8 times lower than that per MBP molecule.

THz absorption data fitted with a three-component model

The THz data were fitted using a three-component model (see Material and Methods for details) as shown in figures 3 and 4 for tau and MBP, respectively. A reasonable fit was obtained for hydration water extending to an R of 46 Å away from the tau protein surface (figure 3a) with a THz absorbance 2% higher (*i.e.* about 429.5 cm^{-1}) than bulk water. The sensitivity of the fit to the two fitting parameters can be seen as a 2D map in figure 3b. It shows that the minimum extension R of hydration water is about 30 Å, although it can be up to 70 Å for a slightly smaller absorption coefficient. The fit of MBP data is shown in figure 4a for an extension R of 96 Å, with an absorption coefficient of hydration water of 422.3 cm^{-1}. Figure 3b shows that the fit is reasonable for R = 90 Å and above. Although our model reproduces qualitatively the trend of the experimental absorption, *i.e.* a THz excess at low concentration turning quickly into a linear decrease, it is not able to reproduce the small increase at very low concentration (about 4 mg/ml). The extension of the dynamical hydration shells around tau is about 2.1 times (46 Å compared to 96 Å) less important than around MBP, which corresponds to about 1.9 times less volume of perturbed water per protein molecule.

MD simulations

MD simulations of tau and MBP were performed in order to evaluate the single-particle dynamics of the first hydration layer (defined by water molecules being within 3 Å from the protein surface). The protein-water hydrogen bond (HB) correlation functions

(shown in figure 5a) reflect the average time scale on which the HB network reorganizes due to water diffusion. The mean-squared displacements (MSD; shown in figure 5b) provide the averaged dynamics of the considered water molecules. The shorter protein-water HB relaxation times (figure 5a) and the higher water MSD of the first hydration layer (figure 5b) of the tau protein, compared to MBP, reveal a faster water single-particle dynamics at the surface of the tau protein. In other words, the retardation of the hydration water dynamics, due to the protein-water interactions, is less important for the tau protein than for the MBP.

Discussion

The output parameters from the three-component model are in good agreement with the qualitative approach. Indeed, the latter showed that the volume of hydration water around tau is about 1.8 times less important than around MBP, in line with the fits shown in figure 3 and 4 that indicates a factor of 1.9 between the volume of hydration water of the two proteins. However, we acknowledge that the distance from the protein surface at which the water is perturbed, between 30 Å and 70 Å for the tau protein and larger than 90 Å for the MBP, are unusually high. Indeed, the dynamical hydration shell of the engineered lambda repressor λ^*_{6-85} (molecular weight of about 8.7kDa) has been shown to be 20 Å (18). One assumption underlying the Monte Carlo simulation is that the particles do not interact in solution. This assumption might not be strictly correct and might lead to an overestimation of the extension of dynamical hydration shells. Nevertheless, no THz absorption experiments have been carried out on proteins of that molecular weight so far and it would thus be highly interesting to perform further measurements, in order to reveal if large proteins have generally much large dynamical hydration shells.

Conti Nibali and Havenith have empathized two conditions that are crucial for a long-range perturbation of the hydration water: (i) the protein-water short-range interactions (protein-water hydrogen bonds), representing the direct coupling that is necessary to transmit the long range interaction and (ii) the vibrational properties of the solute in the THz range (15). The MD simulation of both proteins in solution showed that the retardation of the first hydration layer is less important for the tau protein, as evidence by higher MSD (figure 5b) and shorter protein-water relaxation times (figure 5a). These results are consistent with the less extended dynamical hydration shell of tau found in the THz measurements, by invoking the criteria (i) described above. Indeed, the first hydration layer being less "hooked" to the tau

protein (*i.e.* weaker retardation of the first hydration shell), the vibrational motions of the protein surface are weakly coupled to the surrounding water molecules, resulting in a perturbation of the hydration-water collective motions that is of shorter range around the tau protein than around MBP.

THz spectroscopy applied on the lambda repressor showed that when it is denatured, the extension of the dynamical hydration shell diminished drastically to almost zero (24). The fact that denatured (non-functional) proteins have a weak influence on the surrounding water, while the intrinsically disordered protein tau is shown here to perturb a large volume of water, suggests that the dynamical hydration shell is biologically relevant. It is in line with recent findings that the hydration water of an enzyme is not only substrate specific, but also contributes to enzyme activity (22).

The recently introduced concept of the hydration funnel proposes that a long-range gradient of retardation in hydration water might assist the binding of proteins by guiding the ligand into its binding site (15). Because of their large geometrical extension, IDPs have been shown to have a large radius of capture, facilitating their interaction with potential partners (42). Thus, a limited extension of the dynamical hydration shell in IDPs as compared to globular proteins, leading to a short range hydration funnel, might compensate for their large radius of capture inherent to their geometry.
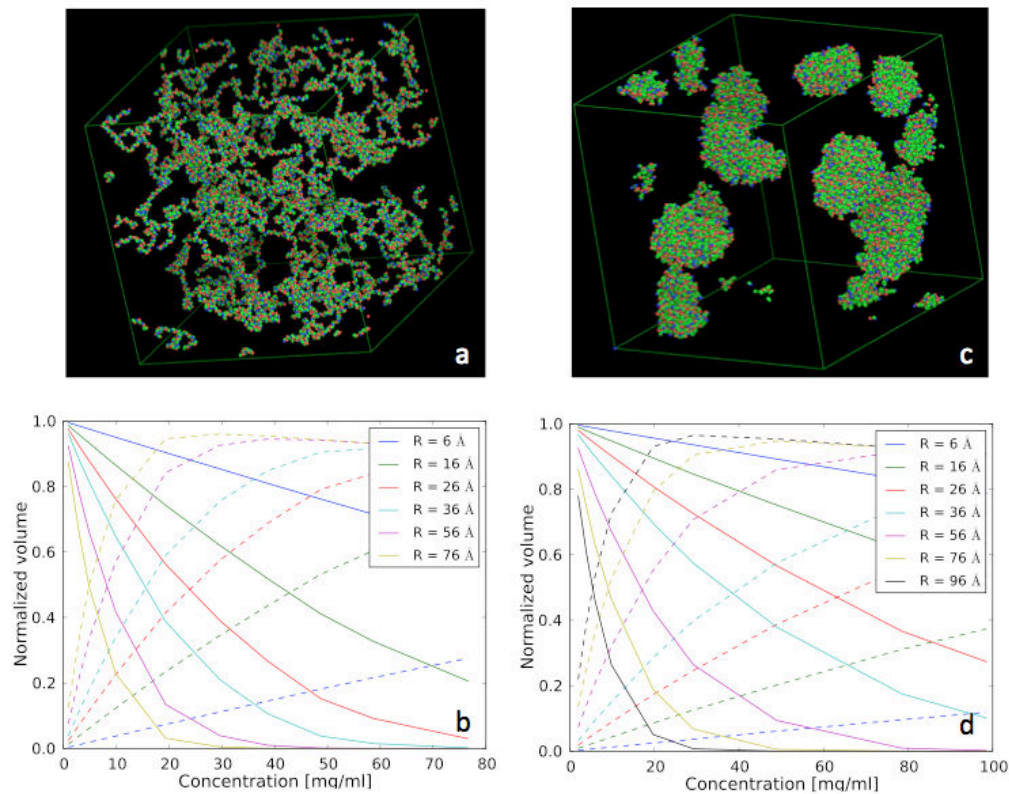
Figure 1: Monté Carlo simulations were used to evaluate the global volume of hydration water in solution, given a distance R from the protein surface within which the water is perturbed. A protein solution is modeled by randomly placing ten protein molecules in a box with periodic boundary conditions. Snapshots of a (a) tau protein box and a (c) MBP box are shown at a protein concentration of 80 mg/ml. The volumes of bulk water (continuous lines) and hydration water (dash lines) as a function of the protein concentration are shown for different extension R of the perturbation around the protein, in (b) a tau protein solution and (d) a MBP solution.
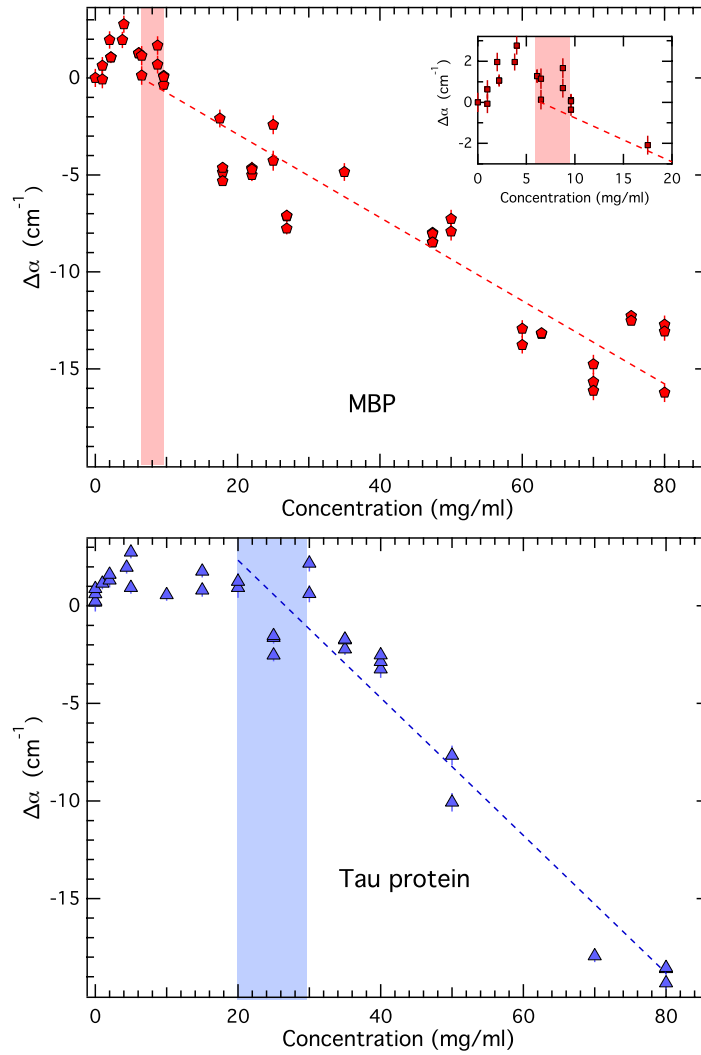
Figure 2: Concentration-dependent THz absorption (relative to pure buffer) of tau protein (bottom) and MBP (top) solution. The absorption of the tau protein enters in a linear regime at 20-30 mg/ml (blue zone) while the MBP reach this regime around 6-9 mg/ml (red zone). The dash lines show the linear fit at high concentration.
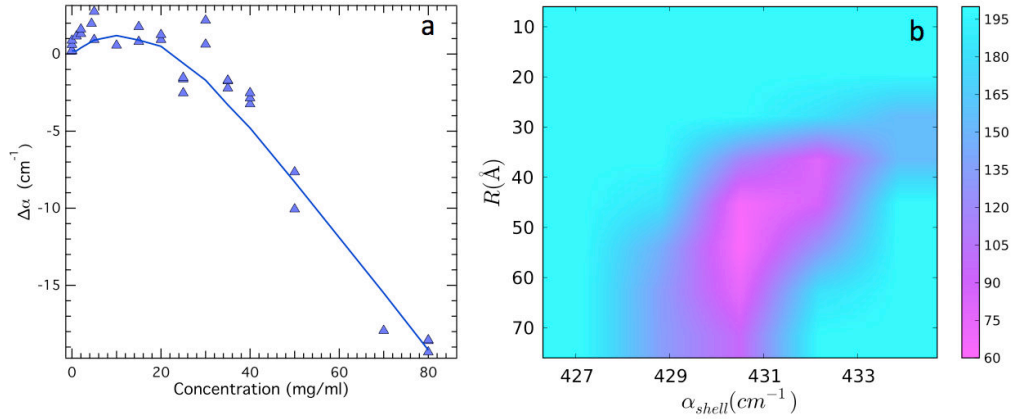
Figure 3: (a) Absorption of the tau solution as a function of the protein concentration and its associated fit based on the three-component model with R=46 Å and α_{shell} =429.5 cm^{-1} . (b) 2D map representing the least square X of the fit as a function of the free parameters (R and α_{shell}). Pink color shows the smaller least squares, *i.e.* the best fits.
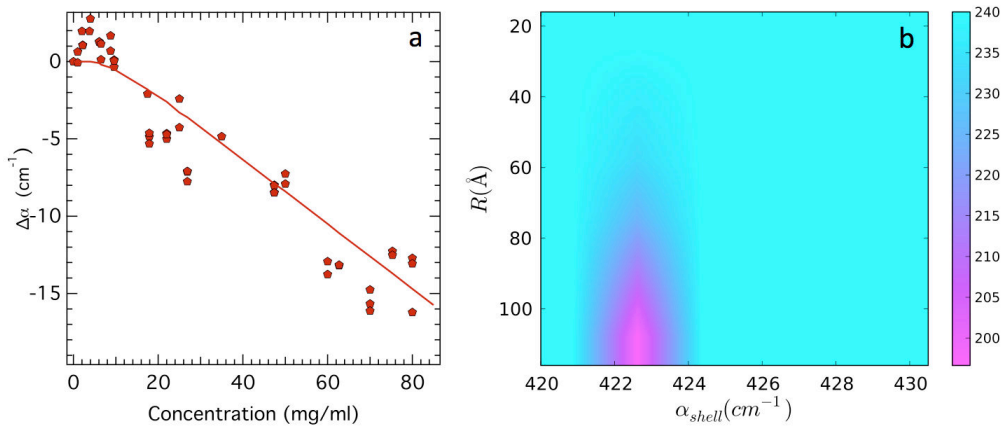
Figure 4: (a) Absorption of the MBP solution as a function of the protein concentration and its associated fit based on the three-component model with R=96 Å and α_{shell} =422.3 cm^{-1} . (b) 2D map representing the least square X of the fit as a function of the free parameters (R and α_{shell}). Pink color shows the smaller least squares, *i.e.* the best fits.
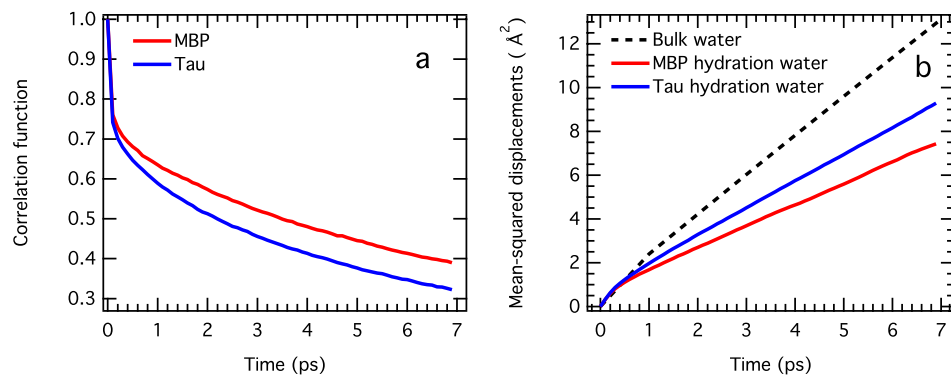
Figure 5: MD simulations of the MBP and Tau proteins solutions. (a) Protein-water HB correlation functions and (b) water mean-squared displacements of the first hydration shell (water molecules within 3 Å away from the protein surface).

References

1. Jensen, M.R., P.R.L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernadó, and M. Blackledge. 2009. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. Structure. 17: 1169–1185.
2. Stadler, A.M., L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, and D. Richter. 2014. Internal Nanosecond Dynamics in the Intrinsically Disordered Myelin Basic Protein. J. Am. Chem. Soc. 136: 6987–6994.
3. Dunker, A.K., Z. Obradovic, P. Romero, E.C. Garner, C.J. Brown, and others. 2000. Intrinsic protein disorder in complete genomes. Genome Inform. Ser. : 161–171.
4. Dyson, H.J. 2011. Expanding the proteome: disordered and alternatively folded proteins. Q. Rev. Biophys. 44: 467–518.
5. Uversky, V.N., C.J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L.M. Iakoucheva, Z. Obradovic, and A.K. Dunker. 2009. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics. 10: S7.
6. Goedert, M., A. Klug, and R.A. Crowther. 2006. Tau protein, the paired helical filament and Alzheimer's disease. J. Alzheimers Dis. 9: 195–207.
7. Thirumalai, D., G. Reddy, and J.E. Straub. 2012. Role of Water in Protein Aggregation and Amyloid Polymorphism. Acc. Chem. Res. 45: 83–92.
8. Ball, P. 2008. Water as an Active Constituent in Cell Biology. Chem. Rev. 108: 74–108.
9. Brovchenko, I., and A. Oleinikova. 2008. Which Properties of a Spanning Network of Hydration Water Enable Biological Functions? ChemPhysChem. 9: 2695–2702.
10. Frauenfelder, H., G. Chen, J. Berendzen, P.W. Fenimore, H. Jansson, B.H. McMahon, I.R. Stroe, J. Swenson, and R.D. Young. 2009. A unified model of protein dynamics. Proc. Natl. Acad. Sci. : pnas.0900336106.
11. Ferrand, M., A.J. Dianoux, W. Petry, and G. Zaccaï. 1993. Thermal motions and function of bacteriorhodopsin in purple membranes: effects of temperature and hydration studied by neutron scattering. Proc. Natl. Acad. Sci. 90: 9668–9672.
12. Ostermann, A., R. Waschipky, F.G. Parak, and G.U. Nienhaus. 2000. Ligand binding and conformational motions in myoglobin. Nature. 404: 205–208.
13. Khodadadi, S., J.H. Roh, A. Kisliuk, E. Mamontov, M. Tyagi, S.A. Woodson, R.M. Briber, and A.P. Sokolov. 2010. Dynamics of Biological Macromolecules: Not a Simple Slaving by Hydration Water. Biophys. J. 98: 1321–1326.
14. Gallat, F.-X., A. Laganowsky, K. Wood, F. Gabel, L. van Eijck, J. Wuttke, M. Moulin, M. Härtlein, D. Eisenberg, J.-P. Colletier, G. Zaccai, and M. Weik. 2012. Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. Biophys. J. 103: 129–136.
15. Conti Nibali, V., and M. Havenith. 2014. New Insights into the Role of Water in Biological Function: Studying Solvated Biomolecules Using Terahertz Absorption Spectroscopy in Conjunction with Molecular Dynamics Simulations. J. Am. Chem. Soc. 136: 12800–12807.
16. Tarek, M., and D.J. Tobias. 2002. Single-Particle and Collective Dynamics of Protein Hydration Water: A Molecular Dynamics Study. Phys. Rev. Lett. 89.
17. Orecchini, A., A. Paciaroni, A.D. Francesco, C. Petrillo, and F. Sacchetti. 2009. Collective Dynamics of Protein Hydration Water by Brillouin Neutron Spectroscopy. J. Am. Chem. Soc. 131: 4664–4669.

18. Ebbinghaus, S., S.J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D.M. Leitner, and M. Havenith. 2007. An extended dynamical hydration shell around proteins. Proc. Natl. Acad. Sci. 104: 20749–20752.

19. Meister, K., S. Ebbinghaus, Y. Xu, J.G. Duman, A. DeVries, M. Gruebele, D.M. Leitner, and M. Havenith. 2013. Long-range protein–water dynamics in hyperactive insect antifreeze proteins. Proc. Natl. Acad. Sci. 110: 1617–1622.

20. Heyden, M., J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx. 2010. Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. Proc. Natl. Acad. Sci. 107: 12068–12073.

21. Bergner, A., U. Heugen, E. Brundermann, G. Schwaab, M. Havenith, D.R. Chamberlin, and E.E. Haller. 2005. New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water. Rev. Sci. Instrum. 76: 063110.

22. Dielmann-Gessner, J., M. Grossman, V. Conti Nibali, B. Born, I. Solomonov, G.B. Fields, M. Havenith, and I. Sagi. 2014. Enzymatic turnover of macromolecules generates long-lasting protein–water-coupled motions beyond reaction steady state. Proc. Natl. Acad. Sci. : 201410144.

23. Grossman, M., B. Born, M. Heyden, D. Tworowski, G.B. Fields, I. Sagi, and M. Havenith. 2011. Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. Nat. Struct. Mol. Biol. 18: 1102–1108.

24. Ebbinghaus, S., S.J. Kim, M. Heyden, X. Yu, M. Gruebele, D.M. Leitner, and M. Havenith. 2008. Protein Sequence- and pH-Dependent Hydration Probed by Terahertz Spectroscopy. J. Am. Chem. Soc. 130: 2374–2375.

25. Wood, K., A. Frölich, A. Paciaroni, M. Moulin, M. Härtlein, G. Zaccai, D.J. Tobias, and M. Weik. 2008. Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water: Direct Measurements by Neutron Scattering and MD Simulations. J. Am. Chem. Soc. 130: 4586–4587.

26. Heugen, U., G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D.M. Leitner, and M. Havenith. 2006. Solute-induced retardation of water dynamics probed directly by terahertz spectroscopy. Proc. Natl. Acad. Sci. 103: 12301–12306.

27. Harpaz, Y., M. Gerstein, and C. Chothia. 1994. Volume changes on protein folding. Structure. 2: 641–649.

28. Vinh, N.Q., S.J. Allen, and K.W. Plaxco. 2011. Dielectric Spectroscopy of Proteins as a Quantitative Experimental Test of Computational Models of Their Low-Frequency Harmonic Motions. J. Am. Chem. Soc. 133: 8942–8947.

29. Xu, J., K.W. Plaxco, and S.J. Allen. 2006. Collective Dynamics of Lysozyme in Water: Terahertz Absorption Spectroscopy and Comparison with Theory. J. Phys. Chem. B. 110: 24255–24259.

30. Mylonas, E., A. Hascher, P. Bernadó, M. Blackledge, E. Mandelkow, and D.I. Svergun. 2008. Domain Conformation of Tau Protein Studied by Solution Small-Angle X-ray Scattering†. Biochemistry (Mosc). 47: 10345–10353.

31. Schwalbe, M., V. Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M.R. Jensen, J. Biernat, S. Becker, E. Mandelkow, M. Zweckstetter, and M. Blackledge. 2014. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution from NMR and Small Angle Scattering. Structure. 22: 238–249.

32. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J Comput Chem. 26: 1781–802.

33. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 102: 3586–616.

34. Mackerell, A.D., Jr., M. Feig, and C.L. Brooks 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem. 25: 1400–15.

35. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926–935.

36. Martyna, G.J., D.J. Tobias, and M.L. Klein. 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101: 4177–4189.

37. Feller, S.E., Y. Zhang, R.W. Pastor, and B.R. Brooks. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103: 4613–4621.

38. Grubmüller, H., H. Heller, A. Windemuth, and K. Schulten. 1991. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. Mol. Simul. 6: 121–142.

39. Tuckerman, M., B.J. Berne, and G.J. Martyna. 1992. Reversible multiple time scale molecular dynamics. J. Chem. Phys. 97: 1990–2001.

40. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23: 327–341.

41. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen. 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103: 8577–8593.

42. Shoemaker, B.A., J.J. Portman, and P.G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. 97: 8868–8873.

4.2 Summary and lead-in for the next chapter

Summary Water is the natural matrix of biological macromolecules and as such is essential for a variety of molecular processes, including protein folding, misfolding and stability. Intrinsically disordered proteins (IDPs) interconvert among multiple extended conformations rather than remaining in a well-defined folded structure and are prone to aggregation and fibrillation. The intrinsically disordered protein tau has been shown to display a different coupling to hydration water dynamics than globular proteins, which is modified when tau forms amyloid fibers involved in Alzheimer disease. Here, we focus on the collective dynamics of hydration water of tau in comparison to that of the folded globular maltose binding protein (MBP). By carrying out THz spectroscopy experiments, we found that the volume of water influenced by a tau protein molecule is two times less important than that influenced by a MBP molecule. The dynamical and structural differences between IDPs and folded proteins might thus be correlated with differences in collective hydration water dynamics.

Introduction to the next chapter This chapter showed that the collective hydration water dynamics of the IDP tau differs from that of the globular MBP. The tau protein has attracted a lot of attention because it can form amyloid fibers that are one of the hallmarks of Alzheimer disease. In addition, some studies have suggested that the hydration water plays a key role in protein aggregation (Thirumalai et al., 2012). Thus, it is of high interest to investigate the hydration water properties of the tau protein in the context of fibrillation. In particular, the next chapter reports a neutron scattering study that aimed at comparing the hydration water dynamics of the tau protein in its native and fibrillated forms.

Hydration water mobility is enhanced around tau amyloid fibers

The manuscript presented herein has been published in the *Proceedings of the National Academy of Sciences* (Fichou et al., 2015). Its formatted version is available online.

5.1 Manuscript

Y. Fichou[a,b,c], G. Schiro[a,b,c], F-X Gallat[a,b,c], C. Laguri[a,b,c], M. Moulin[d], J. Combet[e], M. Zamponi[f], M. Härtlein[d], C. Picart[g,h], E. Mossou[d], H. Lortat-Jacob[a,b,c], J.P. Colletier[a,b,c], D. J. Tobias[i], M. Weik[a,b,c]

[a]Univ. Grenoble Alpes, IBS, F-38044 Grenoble, France; [b]CNRS, IBS, F-38044 Grenoble, France; [c]CEA, IBS, F-38044 Grenoble, France, [d]Life Sciences Group, Institut Laue-Langevin, 38000 Grenoble, France; [e]Institut Laue Langevin, 71 avenue des martyrs, 38000 Grenoble, France; [f]Jülich Centre for Neutron Science, outstation at Heinz Maier-Leibnitz Zentrum, Forschungszentrum Jülich GmbH, 85747 Garching, Germany; [g]CNRS, UMR 5628, LMGP, 38016 Grenoble, France; [h]Université de Grenoble Alpes, Grenoble Institute of Technology, 38016 Grenoble, France; [i]Department of Chemistry, University of California, Irvine, California, USA 92697-2025.

Significance statement

Protein aggregation into amyloid fibers and oligomers is observed in a variety of neurodegenerative diseases. The fibers formed by the intrinsically disordered human protein tau, for instance, are one of the hallmarks of Alzheimer disease. In this work, we report on the dynamic behavior of tau hydration water, which we found to be more mobile in tau fibers than in non-aggregated tau. This increase in mobility could promote fiber formation through an increase in hydration-water entropy. That hydration water is more mobile around the pathological form of tau corroborates that methodologies sensitive to the diffusion of water, such as diffusion magnetic resonance imaging, could be used to diagnose Alzheimer patients in an early stage of the disease.

Abstract

The paired helical filaments (PHF) formed by the intrinsically disordered human protein tau are one of the pathological hallmarks of Alzheimer disease. PHF are fibers of amyloid nature that are composed of a rigid core and an unstructured fuzzy coat. The mechanisms of fiber formation, in particular the role that hydration water might play, remains poorly understood. We combined protein deuteration, neutron scattering and all-atom molecular dynamics simulations to study the dynamics of hydration water at the surface of fibers formed by the full-length human protein htau40. In comparison to monomeric tau, hydration water on the surface of tau fibers is more mobile, as evidenced by an increased fraction of translationally diffusing water molecules, a higher diffusion coefficient and increased mean-squared displacements in neutron scattering experiments. Fibers formed by the hexapeptide ^{306}VQIVYK^{311} were taken as a model for the tau fiber core and studied by molecular dynamics simulations, revealing that hydration water dynamics around the core domain is significantly reduced after fiber formation. Thus, an increase in water dynamics around the fuzzy coat is proposed to be at the origin of the experimentally observed increase in hydration water dynamics around the entire tau fiber. The observed increase in hydration water dynamics is suggested to promote fiber formation through entropic effects. Detection of the enhanced hydration-water mobility around tau fibers is conjectured to potentially contribute to the early diagnosis of Alzheimer patients by diffusion MRI.

Introduction

Amyloid fibers are the most stable forms of ordered protein aggregates. They have attracted much attention because of their implication in so-called conformational diseases, which include a variety of neurodegenerative disorders (1). Consequently, means of hindering or reversing fiber formation are actively researched (2). Pathological fibers are often formed by intrinsically disordered proteins (IDP) that lack a well-defined three-dimensional structure in their native state and are best described by an ensemble of different conformations (3). The human protein tau is an IDP that normally regulates microtubule stability in neurons. When tau aggregates, it forms paired helical filaments (PHF) that are one of the two histological hallmarks of Alzheimer disease (AD) (4, 5). As yet, and despite considerable effort over the past 30 years, the understanding of tau fibrillation in AD and other taupathies remains largely incomplete (6). The longest human tau isoform, htau40, is composed of 441 amino acid residues and is organized into several domains (see figure 1), including the repeat domains R1-R4 (residues 244-369) that constitute together with the P1 and P2 domains the microtubule binding regions (7). Essential for the nucleation of tau fibers is the presence of hexapeptides (^{275}VQIINK^{280} and ^{306}VQIVYK^{311}) in R2 and R3 (8) that have a high propensity to form β-structures. Although precise structures of tau PHF remain unknown (6), they can be divided into two structurally different regions (see figure 1): (i) a rigid β-rich core (denoted as the fiber core domain), which is essentially composed of the four repeat domains, and (ii) the remainder, the so-called fuzzy coat, which is highly flexible (9–11).

Water is known to play key roles in protein folding, stability and activity (12). It mediates protein-protein and protein-DNA recognition, is involved in allostery, partakes in enzymatic reactions and proton and electron transfer, and more generally plasticizes biological macromolecules by providing their surface with an extensive and highly dynamic network of hydrogen bonds. Compared to folded proteins, tau has been shown to have a stronger coupling with its hydration water (13). However, very little is known about the role water plays in protein aggregation in general and in tau fibrillation in particular. A recent study on two different amyloid systems concluded that water plays a key role in fiber growth and polymorphism, *inter alia* through entropic effects (14). A study by Chong and coworkers (15) highlighted the role of water in protein aggregation propensity by revealing a tight relation between the hydration free energy of a protein and its propensity to aggregate.

Among the experimental methodologies available to study protein hydration water,

neutron scattering (NS) stands out owing to its pronounced sensitivity to motions of hydrogen atoms. Indeed, hydrogen atoms incoherently scatter neutrons about two orders of magnitude more strongly than all other atoms present in a biological sample, including deuterium atoms. Consequently, NS has been widely used to study bulk and confined water at room temperature (16), hydration water of peptides (17), proteins (18–21) and water inside cells (22). More specifically, NS probes atomic motions on the nano to picosecond time scales and on the Å length scales (23), thus ensuring the time and space resolution necessary for investigating water dynamics with atomistic detail. Elastic incoherent NS (EINS) reflects the global dynamics averaged over all atoms but does not provide any information on the nature of the observed motions. Quasi-elastic neutron scattering (QENS), however, allows the quantification of energy exchanges between the sample and the neutron beam and provides quantitative information about the nature of motions observed. Because of a pronounced isotope effect, the replacement of hydrogen by deuterium atoms effectively masks the labeled part of a sample in incoherent NS experiments. Perdeuteration of proteins (*i.e.* deuteration of the entire protein) hydrated in H_2O thus puts the focus on hydration water dynamics by minimizing the protein contribution to the neutron scattering signal. All-atom molecular dynamics (MD) simulation is a useful complement to neutron scattering because both methods probe atomic motions on the same time and length scales. Whereas incoherent NS provides an accurate measure of the average dynamics of hydrogen atoms throughout the sample, MD simulations provide atomic-scale insight into motions occurring within particular space and time windows of interest (24).

Here we experimentally and computationally address the effect of tau fiber formation on the dynamics of its surrounding hydration water. We produced perdeuterated htau40 as well as a perdeuterated heparin analogue, and measured by NS the dynamical properties of hydration water on the surface of tau monomers and of tau fibers whose formation was triggered by the heparin analogue. Both elastic and quasi-elastic NS indicate an increased mobility of hydration water on tau fibers compared to tau monomers. MD simulations provide circumstantial evidence suggesting that it is the increase in water dynamics around the disordered fuzzy coat and not around the fiber core that is at the origin of the experimentally observed increase in tau hydration water dynamics after fibrillation. We conjecture that the observed gain in water dynamics reflects an increase in water entropy that is favorable to the fiber formation.

Results

Tau fiber characterization

Fibers formed by adding heparin to monomeric tau have been reported to closely resemble those formed *in vivo* by tau hyperphosphorylation in AD brains (25). Here we produced and used deuterated versions of tau and heparin in order to largely mask their incoherent contribution in NS experiments and thus focus on the hydration-water dynamics around tau fibers. In order to exclude an isotope effect on fiber formation and morphology, we characterized deuterated fibers by complementary biophysical methods. A negative staining electron micrograph of the deuterated tau fibers is presented in figure 2a. Control micrographs of the monomeric tau did not reveal the presence of fibers. X-ray fiber diffraction patterns of deuterated tau fibers (figure 2b) showed the typical signature of amyloid structures. The diffraction ring at 4.7 Å corresponds to the cross-β sheet distance along the fiber direction and the ring at 9 Å reflects the gauge of the steric zipper, perpendicular to the fiber direction (26, 27). Similar observations were made on hydrogenated tau fibers and monomers as shown in figure S1. We conclude that adding deuterated heparin to monomeric htau40 leads to the formation of fibers resembling those formed from hydrogenated constituents.

Water mean-squared displacements are increased at the surface of tau fibers

We performed EINS on powders of fibers (denoted as D-fiber-H_2O) and monomers (D-tau-H_2O) of deuterated tau hydrated at 0.4 g H_2O / g protein. This experiment aimed at measuring the dynamics of the first hydration shell of the protein in both states. Hydrogen atoms in the hydration water contribute 71% to the incoherent neutron scattering signal in both samples, while the remainder originates from the deuterated proteins, including exchanged hydrogen atoms (see SI for detailed calculations). The contribution of deuterated heparin potentially included in the fibers is negligible (see SI). Incoherent NS experiments from D-fiber-H_2O and D-tau-H_2O thus mainly monitor hydration-water dynamics. Mean-squared displacements (MSD) of hydration water in both samples were extracted by applying a Gaussian approximation (see SI for details), and are presented as a function of temperature in figure 3a. The MSD of D-fiber-H_2O become markedly larger than those of D-tau-H_2O above around 220 K, where large amplitude water motions set in (28). At 300 K, they are about 30% higher for the fiber compared to the monomer sample. Hydration water is thus

more mobile around the tau fibers than around the monomeric protein. The total elastically scattered intensity, which provides a model-free estimation of the dynamics, confirms the difference seen in the water MSD between D-fiber-H_2O and D-tau-H_2O (figure S2a).

Water translational diffusion is increased around tau fibers

In order to provide quantitative information about the nature of water motions observed, we recorded QENS spectra on the D-fiber-H_2O and D-tau-H_2O samples at 280 K. The raw spectra are shown in figure 3b. The D-fiber-H_2O spectrum exhibits a larger quasi-elastic broadening, confirming qualitatively the enhanced dynamics of fiber hydration-water as evidenced by the MSD (figure 3a). QENS data were fitted with a model (29) that describes water diffusional dynamics as a superposition of translational and rotational motions, with a supplementary term for immobile water molecules (see SI for details and figure S3 for an example of fits). The fitting procedure allows extraction of the fraction of the scattering signal coming from translating, rotating or immobile water molecules, as well as their respective diffusion coefficients. Table S1 presents the output parameters of the fitting procedure. The fraction of water molecules undergoing translational diffusion around the fibers is 25% higher than around the monomers. Furthermore, the translational diffusion coefficient and the rotational rate are 11% and 17% higher, respectively, for the fiber hydration water.

Protein dynamics of tau monomers and fibers are identical

In order to determine if the enhancement of hydration water dynamics originates from a change in protein dynamics upon fibrillation, we measured by EINS the dynamics of the fibrillated and the monomeric tau protein. To this end, we prepared hydrogenated tau amyloid fibers, hydrated with 0.4 g D_2O /g protein (H-fiber-D_2O), strictly following the protocol established for formation of the D-fibers. Incoherent neutron scattering from such a sample amounts to 97% from the protein and only to 3% from the hydrating D_2O (detailed calculations are presented in the SI) and, hence, almost solely reflects protein dynamics, and not water dynamics. We extracted the MSD (figure 3c) of tau fibers (H-fiber-D_2O) and compared them to those of monomeric tau (H-tau-D_2O; re-processed from Gallat *et al*. (13)). We applied a q^4 correction to the Gaussian approximation (see figure S4 and SI for details) that allows fitting a large q-range and provides not only the MSD, but also the width of the MSD distribution (30, 31). This width reflects the heterogeneity of the MSD among the

various hydrogen atoms in the protein. Note that this model was not suitable to fit scattering data from the deuterated samples. The H-fiber-D_2O and H-tau-D_2O samples present identical MSD (figure 3c) with similar distributions (figure S5). Thus, both tau monomers and fibers exhibit the same protein dynamics on the nano- to picosecond time scales, and the differences in water dynamics are inherent to the hydration shell.

MD simulation shows decreased water dynamics around the fibrillated hexapeptide $^{306}VQIVYK^{311}$

The structure of tau fibers is heterogeneous in the sense that about 30% of the monomer is included in a cross-β structure (referred as the fiber core), while the rest remains disordered (fuzzy coat; see figure 1). Because neutron scattering provides information on the *average* dynamics of all hydrogen atoms in the sample, we employed MD simulations to assist the interpretation of our experimental results (see discussion). We carried out all-atom MD simulations on the amyloid-prone hexapeptide $^{306}VQIVYK^{311}$, which belong to the tau fiber core (8) and was therefore used as a model of the fiber core (32, 33). The simulation was performed on the monomeric peptide in solution (figure S6a) and the peptide fiber based on the crystal structure (34) (PDB entry 2ON9; figure S6b), following the procedure described by Zhao *et al.* in (35). We compared the dynamics of the water molecules (figure 4) around the monomeric peptide (highlighted in figure S6a) and around the peptide fiber (highlighted in figure S6b). The MSD of hydration water (figure 4a) are higher for the monomeric compared to the fibrillated peptide. In order to provide further insight into water behavior at the surface of the peptide, two types of peptide-water hydrogen bond (HB) correlation functions were analyzed: the continuous HB correlation function, the decay of which defines the timescale, on average, on which protein-water hydrogen bonds break, and the intermittent HB correlation function, the decay of which defines the timescale of the rearrangement of the protein-water HB network. The decay of the continuous HB correlation function is due primarily to the rotational/librational motions of water molecules, while the decay of the intermittent HB correlation function results from the reorganization of the protein HB network due to water translational diffusion (36). The relaxation times of both correlation functions are longer (figure 4b and 4c) when the peptide is in the amyloid state, suggesting, consistently with the water MSD (figure 4a), that the hydration water dynamics around the core domain is slowed down after fiber formation.

Discussion

The mobility of hydration water on the surface of tau fibers is increased compared to tau monomers as evidenced by elastic and quasi-elastic neutron scattering. Quantitatively, 25% more water molecules undergo translational diffusion on the fiber surface and they display a 11% higher diffusion coefficient. The protein dynamics of monomeric and fibrillar tau appear to be identical, implying that the observed differences in water dynamics are inherent to the hydration shell. MD simulation, carried out on a model of the fiber core, shows a reduced water mobility around the amyloid form of the fiber core model.

Is the experimentally observed average increase in hydration water dynamics around fibrillar tau due to enhanced dynamics of water around the core or the fuzzy coat domains (see figure 1 for a visual representation of these domains)? In order to address this question, we divide the average hydration water dynamics in two parts (see figure 1): the dynamics of the water molecules around residues 244-369 (forming the core domain in the fibers) and around the residues 1-243 and 370-441 (forming the fuzzy coat in the fibers). The subscripts *core* and *fuzz* refer to these two groups, respectively, and the subscript *tot* to the water around the entire protein. Note that the definition of these groups applies to both the monomeric and the fibrillated protein. Then, one can write : $\langle u \rangle_{tot} = p_{fuzz} \times \langle u \rangle_{fuzz} + p_{core} \times \langle u \rangle_{core}$ *(eq.1)*, where $\langle u \rangle$ are the mean values of a generic dynamic parameter, and p_{fuzz} and p_{core} are fractions of water molecules being at the surface of the fuzzy coat and the core domain, respectively, with $p_{fuzz} + p_{core} = 1$. Note that u represents any dynamical quantity, such as MSD for instance. Following eq. 1, our experimental finding can be written as $\langle u \rangle_{tot}^{fib} > \langle u \rangle_{tot}^{mon}$, where the superscripts *fib* and *mon* refer to the fibrillated and monomeric protein, respectively.

During tau fibrillation, a fraction of the core domain becomes dehydrated when it forms dry steric zippers (34) and cross-β sheets *i.e.* $p_{core}^{fib} < p_{core}^{mon}$ and thus $p_{fuzz}^{fib} > p_{fuzz}^{mon}$. If the average amino acid composition of core and fuzzy coat were markedly different, in particular resulting in a different hydrophobicity, a redistribution of water from the core to the fuzzy coat after fibrillation could affect hydration water dynamics (17, 37). However, according to the hydropathy scale of Kyle and Doolittle (38), the average hydropathy of the core and fuzzy coat are similar (-0.6±1.3 and -0.9±1.4, respectively). This similarity thus suggests that $\langle u \rangle_{fuzz}^{mon} \approx \langle u \rangle_{core}^{mon}$ and that a redistribution of water between core and fuzzy coat domains is not

at the origin of the experimentally observed increase in the average tau hydration dynamics after fibrillation.

Another possible origin of the enhanced hydration water dynamics observed on the surface of tau fibers is an increase in water dynamics on the core moiety after fiber formation. The amyloid core is composed of β-sheets that are stabilized by inter-molecular hydrogen bonds, thus reducing the number of hydrogen bond donors and acceptors available for interaction with hydration water. As a consequence, one could conjecture that water moves more freely on the fiber core, thus leading to $\langle u \rangle_{core}^{fib} > \langle u \rangle_{core}^{mon}$. In order to evaluate this possibility *in silico*, we chose the hexapeptide [306]VQIVYK[311] as a model for the fiber core and carried out MD simulations of both monomeric and amyloid fiber states. The analysis of water dynamics in the first hydration layer shows an overall lower mobility around the amyloid form, as revealed by water total MSD (see figure 4a). In particular, both the continuous (figure 4b) and the intermittent (figure 4c) protein-water hydrogen bond relaxation times are longer around the fibrillated peptides, suggesting that both rotational and translational diffusions are reduced around the amyloid assembly. A reduction in the local dynamics of water molecules around the core domain in tau fibers (in particular around the residue Cys322) has also been observed in Overhauser dynamic nuclear polarization NMR experiments (39). Therefore, we propose that it is not the water dynamics around the core domain, but around the fuzzy coat, that is increased when tau has formed amyloid fibers, resulting in the average increase in water dynamics observed in our neutron experiments. Spatially resolved water dynamics on the core and on the fuzzy domains could be monitored with time-resolved fluorescence lifetime measurements (40).

The local protein topology, determined by the protein conformation, has been proposed to be a key, if not the main, feature determining the hydration water dynamics (41– 44). Significant conformational changes in the fuzzy coat have been shown by NMR spectroscopy to accompany tau-fiber formation (11), thus a change in water mobility upon fiber formation is expected. A rough model of a tau fiber has been proposed based on EM and AFM experiments, in which the fuzzy coat resembles a two-layered polyelectrolyte brush with the protein termini sticking out of the fiber core (45). In view of this model, one might conjecture that the alignment of the protein termini within the fuzzy coat after fiber formation modifies the confinement geometry of hydration water. In particular, the confinement dimensionality might be reduced from 3D before, to 1D or 2D after fiber formation, thus

increasing water mobility, as shown in model systems by e.g. MD simulations and experiments (46, 47). Thus, we propose that the conformational change of the fuzzy coat imposed by formation of the fiber core is responsible for the experimentally observed increase in water dynamics.

A macromolecular surface perturbes hydration water with respect to bulk water, resulting in a decrease in both water dynamics and water entropy. The notion of water being a reservoir of entropy for biomolecules has been discussed in the context of protein folding, binding and aggregation (14, 48–50). Energetically unfavourable protein conformational changes can feed upon the water entropy. For instance, Breinten and colleagues showed experimentally that water entropy compensates for the poor binding enthalpy of a protein-ligand complex (49). Several studies (14, 51) have shown that the nucleation of amyloid structures requires surmounting a free energy barrier. That this thermodynamic cost can be compensated by an increase in hydration entropy through the release of water into the bulk has been shown *in silico* (14). Our observed increase in the fraction of water molecules undergoing translational diffusion on the fiber surface reflects a decrease in water perturbation and thus a gain in water entropy. A tentative estimation of the entropy gain associated with the addition of water translational degrees of freedom could be based on the increase in the number of translating water molecules around tau fibers (see Table S1). However, it has been shown (see e.g. (52)) that the translational entropy of a water molecule interacting with a protein depends not only on the surface hydrophobicity, but also largely on the water-protein distance. Consequently, without knowing the surface topology of tau fibers at high resolution, any quantitative estimation of the entropy change would not be reliable. Nevertheless, from a qualitative point of view, we propose that the entropy gain associated with the increase in the number of translating water molecules promotes fiber formation.

Many studies have examined fibers formed only by the core domains (so-called K18 and K19 fragments) as models for the biologically relevant tau PHF. Our work provides an example of a property that differs significantly between the fiber core domain (reduced hydration dynamics) and the full length tau fibers (enhanced hydration water dynamics). Care should thus be taken when extrapolating experimental results obtained on a protein-fragment fiber to a full-length protein fiber.

An increase in water diffusivity has been observed with diffusion magnetic resonance imaging (MRI) in the hippocampus of patients suffering from Alzheimer's disease (53, 54),

and proposed as a potential early biomarker of the disease. This increase has been hypothesized to originate from the decrease in neuronal cell density accompanying the progression of AD, yet experimental validation remains elusive. Our study raises the possibility that the enhanced water mobility around tau PHF could provide an additional explanation for the increased water diffusivity revealed by diffusion MRI. Indeed, while the local density of PHF within volumes of brains corresponding to the spatial resolution of MRI (approximately 1 mm^3 (55)) remains unknown to the best of our knowledge, PHF have been shown to be densely packed within neuronal cells (see e.g. figure 4 in (56)) and the proportion of accelerated PHF hydration water compared to bulk-like water might thus be substantial. In addition, islands of dystrophic neurons containing PHF in AD affected brains have been shown to reach 1 mm in size (57). Consequently, accelerated PHF hydration water might well be at the origin, at least partially, of the enhanced water diffusivity in AD affected brains detected by diffusion MRI.

In conclusion, we have provided experimental evidence for an increased hydration-water mobility of tau fibers compared to monomers, which we tentatively assigned to an increase on the surface of the fuzzy coat. The ensemble of results presented suggests a scenario in which the hydration water mobility plays a role in the formation of tau amyloid fibers by providing entropic compensation. It would be highly interesting to follow water dynamics during the fibrillation process. The question of whether or not the growth-rate of amyloid fibers can be modulated by acting on the solvent dynamics remains to be studied. If this were the case, efforts to mitigate the formation of tau amyloid fibers in the context of AD should be extended to include a focus on hydration-water dynamics.

Material and Methods

Expression, purification, fibrillation and sample preparation of htau40 for neutron experiments

The expression and purification of the human isoform htau40 has been published elsewhere (13) and is briefly recalled in the SI. The deuterated fibers (denoted as D-fiber) and hydrogenated fibers (H-fiber) were prepared identically. The purified protein was mixed with the deuterated heparin analogue (the production of which is described in SI) at a molar ratio tau:heparin 4:1 to trigger fibrillation. The fibrillation was monitored by fluorescence of thioflavin S on a BioTek Synergy H4 Hybrid microplate reader with excitation and emission

wavelengths of 430 and 490 nm, respectively. A typical fluorescence curve as a function of time of the hydrogenated protein solution is shown in figure S7.

After adding the deuterated heparin analogue, the solutions were incubated at room temperature for the hydrogenated protein and at 30°C for the perdeuterated protein for about 2 weeks. The fibers were then extracted by centrifugation at 125,000 g for 90 min and used for neutron sample preparation, as well as for biophysical characterization (see SI). After lyophilization, the deuterated fibers were rehydrated to 0.40 g H_2O / g protein (sample denoted D-fiber-H_2O) and the hydrogenated fibers were rehydrated to 0.44 g D_2O /g protein (H-fiber-D_2O). Both samples were then sealed in a 4x3 cm^2 flat aluminum sample holder. A detailed protocol of the sample preparation is given in the SI. The deuterated and hydrogenated monomeric tau samples (i.e. D-tau-H_2O and H-tau-D_2O, respectively) were prepared previously (13).

Elastic incoherent neutron scattering (EINS) experiments

EINS of D-fiber-H_2O and D-tau-H_2O was carried out on the backscattering spectrometer SPHERES (58) (Jülich Centre for Neutron Science at the Heinz Maier-Leibnitz-Zentrum Garching (MLZ), Garching, Germany). The instrumental energy resolution of 0.65 μeV (full width at half-maximum) allowed motions faster than about 1 ns to be probed. The instrument uses neutrons with a wavelength of 6.27 Å and scattering vectors q in the range of 0.21–1.84 Å^{-1}. EINS of H-fiber-D_2O was carried out on the backscattering spectrometer IN16 (59) at the ILL (Grenoble, France). The instrumental energy resolution of 0.9 μeV (full width at half-maximum) allowed motions faster than about 1 ns to be probed. The instrument uses neutrons with a wavelength of 6.27 Å and scattering vectors q in the range of 0.19–1.95 Å^{-1}. Data on H-tau-D_2O, also measured on IN16 were taken from previous work (13) and re-processed. The MSD were calculated using the Gaussian approximation (SI).

Quasi-elastic neutron scattering (QENS)

QENS experiments on the D-fiber-H_2O sample was carried out on SPHERES (58). The data were collected over an energy range of -15.8/+15.6 μeV at 20 and 280 K, for 10 and 12 h, respectively. The spectra at 280 K (figure 3b and S3) were normalized to the 20 K spectra. QENS data on D-tau-H_2O, measured on SPHERES according to the same protocol,

were obtained by Schiro and co-workers (Schiro *et al.*, in press (DOI: 10.1038/ncomms7490)). Details on data processing and modeling are available as SI.

MD simulation

The monomeric peptide model was built by placing the hexapeptide ^{306}VQIVYK^{311} in a box containing 9253 water molecules and one chloride ion for electroneutrality (see a snapshot of the simulation box in figure S6a). The fiber model was based on the work of Zhao *et al.* (35). The fiber was made of the antiparallel two-sheet crystal structure published by Sawaya and coworkers (34) (PDB entry 2ON9). Five strands were stacked on top of each other with a distance of 4.7 Å (figure S6b) and placed in a box of 9258 water molecules. After 18 ns of equilibration, the dynamical parameters shown in figure 4 were computed over 2 ns for water molecules within 3 Å of the protein (see figure S6). Details on the simulation protocol and the calculations of the dynamical parameters are given as SI.

Author Contributions

YF, JPC, MW designed research; YF proposed the experiment; YF, FXG, CL, MM, MH, HLJ, EM prepared samples; YF, FXG, MZ, JC, CP performed experiments; YF, GS analyzed data; YF, DT performed MD simulations; YF, MW wrote the manuscript with input from all authors.
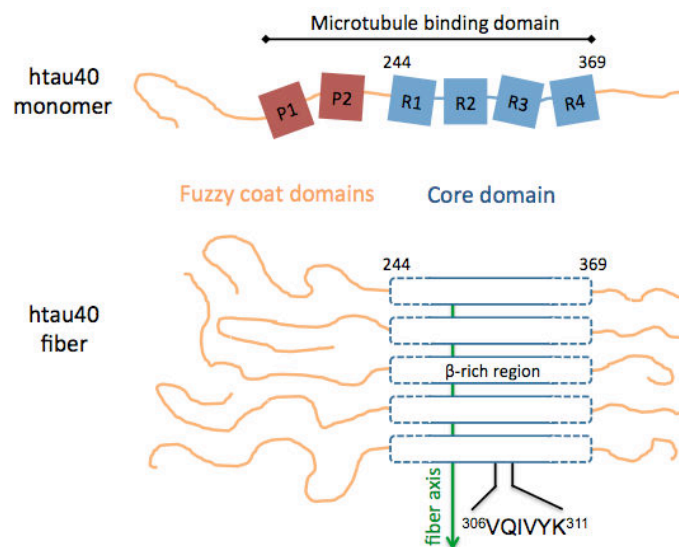
Figure 1: Schematic representation of the tau isoform htau40 in its monomeric (top) and fibrillated (bottom) form. The microtubule binding domain is roughly composed of the four repeat domains R1-R4 (residues 244-369) and the proline-rich domains P1-P2. R1-R4 constitute the core domain, which forms cross-β structures as well as steric zippers in the fiber, while the rest of the protein is referred as the fuzzy coat domains, which remains disordered in the fiber form. The amyloidogenic hexapeptide ^{306}VQIVYK^{311} can be used as a model for the fiber core.
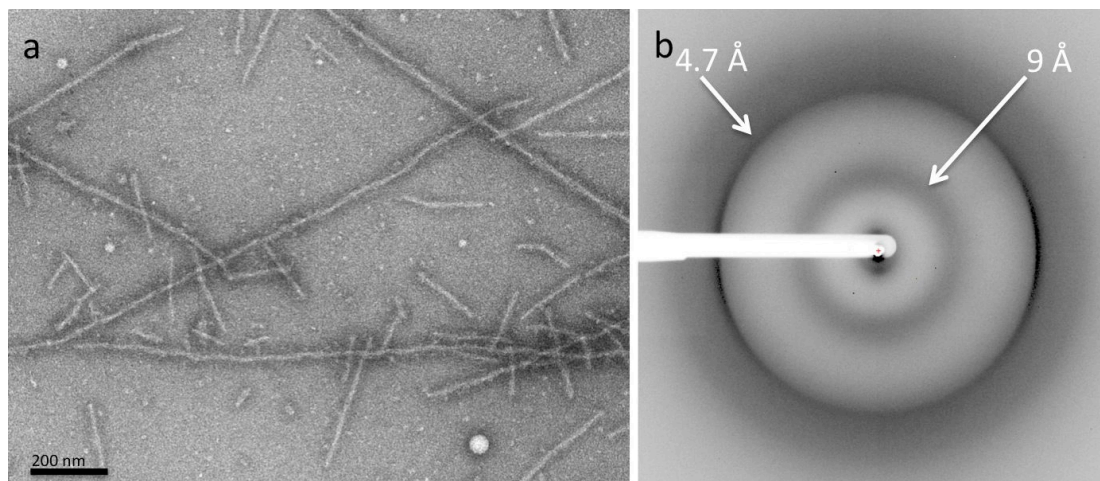
Figure 2: (a) Electron micrograph and (b) X-ray fiber diffraction pattern of deuterated tau amyloid fibers. The white arrows highlight the diffraction rings observed at 4.7 and 9.0 Å, which are characteristic of amyloid structures.
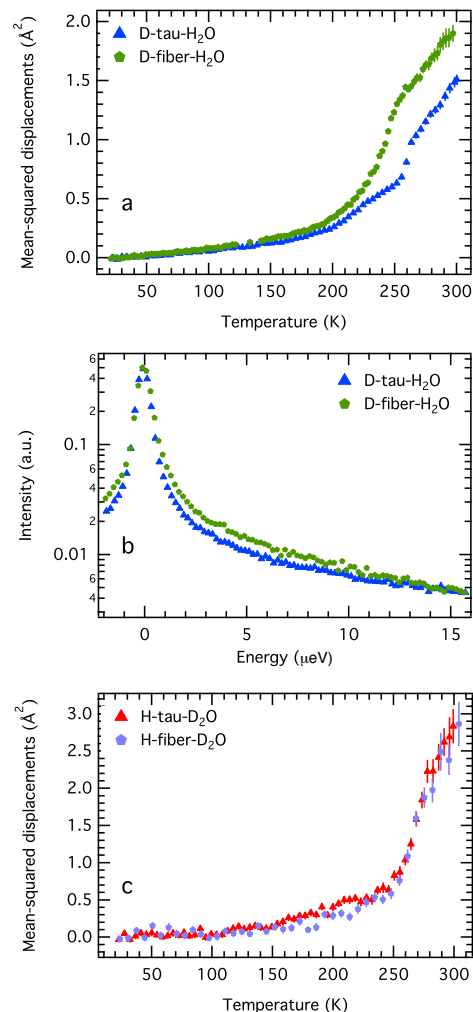
Figure 3: (a) MSD of the hydration water around monomers (D-tau-H_2O) and fibers (D-fiber-H_2O) of the tau protein. The MSD were extracted from the Gaussian approximation fitted between q-values of 0.78 and 1.76 $Å^{-1}$. (b) Comparison of the raw QENS spectra of D-fiber-H_2O and D-tau-H_2O at 280 K, binned over q-values from 0.45 to 1.66 $Å^{-1}$. The spectra of D-tau-H_2O were extracted from Schiro *et al.* (Schiro *et al.* under review). For a visual comparison, the spectra were scaled with a multiplicative factor and corrected for a linear background. Because QENS spectra are symmetrical, the focus was put on positive energy exchange. (c) MSD of monomers (H-tau-D_2O) and fibers (H-fiber-D_2O) of the tau protein. They were extracted from the Gaussian approximation with a q^4 correction between q-values of 0.43 and 1.93 $Å^{-1}$. The error bars represent the SD output from the fitting procedure. Data on H-tau-D_2O were taken from (13) and re-analyzed.
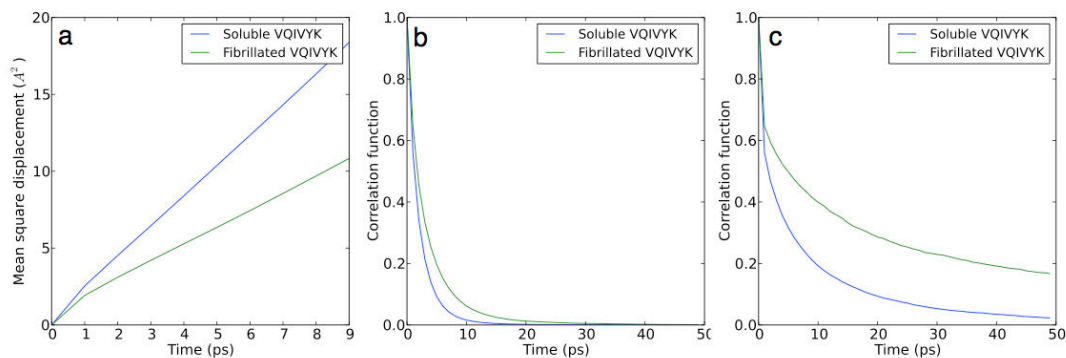
Figure 4: Dynamical properties of the hydration water around the fibrillated and the monomeric peptide [306]VQIVYK[311], obtained by MD simulations. (a) MSD of the first hydration shell (defined as water molecules within 3 Å of the peptide), (b) protein-water continuous hydrogen bond correlation function (providing information on the timescale of water rotational/librational dynamics) and (c) protein-water intermittent hydrogen bond correlation function (providing information on the timescale of the rearrangement of the protein-water hydrogen bond network due to water translational diffusion).

References

1. Chiti F, Dobson CM (2006) Protein Misfolding, Functional Amyloid, and Human Disease. *Annu Rev Biochem* 75(1):333–366.
2. Lednev IK (2014) Amyloid Fibrils: the Eighth Wonder of the World in Protein Folding and Aggregation. *Biophys J* 106(7):1433–1435.
3. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu Rev Biophys* 37(1):215–246.
4. Brion JP, Couck AM, Passareiro E, Flament-Durand J (1985) Neurofibrillary tangles of Alzheimer's disease: an immunohistochemical study. *J Submicrosc Cytol* 17(1):89—96.
5. Kosik KS, Joachim CL, Selkoe DJ (1986) Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease. *Proc Natl Acad Sci* 83(11):4044–4048.
6. Mandelkow EM, Mandelkow E (2012) Biochemistry and cell biology of tau protein in neurofibrillary degeneration. *Cold Spring Harb Perspect Med* 2(7):a006247.
7. Amos LA (2004) Microtubule structure and its stabilisation. *Org Biomol Chem* 2(15):2153.
8. Von Bergen M, et al. (2000) Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure. *Proc Natl Acad Sci* 97(10):5129–5134.
9. Wischik CM, et al. (1988) Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. *Proc Natl Acad Sci U A* 85(12):4506–10.
10. Crowther T, Goedert M, Wischik CM (1989) The Repeat Region of Microtubule-Associated Protein Tau Forms Part of the Core of the Paired Helical Filament of Alzheimer's Disease. *Ann Med* 21(2):127–132.
11. Bibow S, et al. (2011) The dynamic structure of filamentous tau. *Angew Chem Int Ed Engl* 50(48):11520–4.
12. Ball P (2008) Water as an Active Constituent in Cell Biology. *Chem Rev* 108(1):74–108.
13. Gallat F-X, et al. (2012) Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys J* 103(1):129–136.
14. Thirumalai D, Reddy G, Straub JE (2012) Role of Water in Protein Aggregation and Amyloid Polymorphism. *Acc Chem Res* 45(1):83–92.
15. Chong S-H, Ham S (2014) Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angew Chem Int Ed* 53(15):3961–3964.
16. Bellissent-Funel M, Chen SH, Zanotti J (1995) Single-particle dynamics of water molecules in confined space. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top* 51(5):4558–4569.
17. Russo D, Hura G, Head-Gordon T (2004) Hydration dynamics near a model protein surface. *Biophys J* 86(3):1852–62.
18. Bellissent-Funel M-C, Teixeira J, Bradley K-F, Chen SH (1992) Dynamics of hydration water in protein. *J Phys I* 2(6):995–1001.
19. Doster W, et al. (2010) Dynamical transition of protein-hydration water. *Phys Rev Lett* 104(9):098101.
20. Achterhold K, et al. (2011) Dynamical properties of the hydration shell of fully deuterated myoglobin. *Phys Rev E* 84(4).

21. Nickels JD, et al. (2012) Dynamics of protein and its hydration water: neutron scattering studies on fully deuterated GFP. *Biophys J* 103(7):1566–75.
22. Frolich A, et al. (2009) From shell to cell: neutron scattering studies of biological water dynamics and coupling to activity. *Faraday Discuss* 141:117–30; dsicussion 175–207.
23. Gabel F, et al. (2002) Protein dynamics studied by neutron scattering. *Q Rev Biophys* 35(04):327–367.
24. Bizzarri AR (2004) Neutron scattering and molecular dynamics simulation: a conjugate approach to investigate the dynamics of electron transfer proteins. *J Phys Condens Matter* 16(6):R83–R110.
25. Goedert M, et al. (1996) Assembly of microtubule-associated protein tau into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature* 383(6600):550–553.
26. Astbury WT, Dickinson S, Bailey K (1935) The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem J* 29(10):2351.
27. Sunde M, et al. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* 273(3):729–739.
28. Wood K, et al. (2008) Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water: Direct Measurements by Neutron Scattering and MD Simulations. *J Am Chem Soc* 130(14):4586–4587.
29. Schirò G, et al. (2015) Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins. *Nat Commun* 6:6490.
30. Becker T, Smith JC (2003) Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular systems. *Phys Rev E Stat Nonlin Soft Matter Phys* 67(2 Pt 1):021904.
31. Yi Z, Miao Y, Baudry J, Jain N, Smith JC (2012) Derivation of Mean-Square Displacements for Protein Dynamics from Elastic Incoherent Neutron Scattering. *J Phys Chem B*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22471396.
32. Sievers SA, et al. (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* 475(7354):96–100.
33. Berhanu WM, Masunov AE (2011) Can molecular dynamics simulations assist in design of specific inhibitors and imaging agents of amyloid aggregation? Structure, stability and free energy predictions for amyloid oligomers of VQIVYK, MVGGVV and LYQLEN. *J Mol Model* 17(10):2423–2442.
34. Sawaya MR, et al. (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* 447(7143):453–457.
35. Zhao J-H, et al. (2010) Molecular dynamics simulations to investigate the stability and aggregation behaviour of the amyloid-forming peptide VQIVYK from tau protein. *Mol Simul* 36(13):1013–1024.
36. Tarek M, Tobias D (2002) Role of Protein-Water Hydrogen Bond Dynamics in the Protein Dynamical Transition. *Phys Rev Lett* 88(13). Available at: http://link.aps.org/doi/10.1103/PhysRevLett.88.138101 [Accessed September 21, 2014].
37. Jana B, Pal S, Bagchi B (2008) Hydrogen Bond Breaking Mechanism and Water Reorientational Dynamics in the Hydration Layer of Lysozyme. *J Phys Chem B* 112(30):9112–9117.
38. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105–132.
39. Pavlova A, et al. (2009) Site-specific dynamic nuclear polarization of hydration water as a generally applicable approach to monitor protein aggregation. *Phys Chem Chem Phys PCCP* 11(31):6833–6839.

40. Zhang L, et al. (2007) Mapping hydration dynamics around a protein surface. *Proc Natl Acad Sci* 104(47):18461–18466.
41. Luise A, Falconi M, Desideri A (2000) Molecular dynamics simulation of solvated azurin: Correlation between surface solvent accessibility and water residence times. *Proteins Struct Funct Bioinforma* 39(1):56–67.
42. Makarov VA, Andrews BK, Smith PE, Pettitt BM (2000) Residence times of water molecules in the hydration sites of myoglobin. *Biophys J* 79(6):2966–2974.
43. Sterpone F, Stirnemann G, Laage D (2012) Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *J Am Chem Soc* 134(9):4116–4119.
44. Bagchi K, Roy S (2014) Sensitivity of Water Dynamics to Biologically Significant Surfaces of Monomeric Insulin: Role of Topology and Electrostatic Interactions. *J Phys Chem B* 118(14):3805–3813.
45. Wegmann S, Medalsy ID, Mandelkow E, Muller DJ (2013) The fuzzy coat of pathological human Tau fibrils is a two-layered polyelectrolyte brush. *Proc Natl Acad Sci U A* 110(4):E313–21.
46. Swenson J, Bergman R, Longeville S, Howells WS (2001) Dynamics of 2D-water as studied by quasi-elastic neutron scattering and neutron resonance spin-echo. *Phys B Condens Matter* 301(1–2):28–34.
47. Hummer G, Rasaiah JC, Noworyta JP (2001) Water conduction through the hydrophobic channel of a carbon nanotube. *Nature* 414(6860):188–190.
48. Harano Y, Kinoshita M (2004) Large gain in translational entropy of water is a major driving force in protein folding. *Chem Phys Lett* 399(4–6):342–348.
49. Breiten B, et al. (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* 135(41):15579–84.
50. Brovchenko I, Oleinikova A (2008) Which Properties of a Spanning Network of Hydration Water Enable Biological Functions? *ChemPhysChem* 9(18):2695–2702.
51. Nelson R, et al. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435(7043):773–8.
52. Sasikala WD, Mukherjee A (2014) Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding. *J Phys Chem B* 118(36):10553–10564.
53. Kantarci K, et al. (2005) DWI predicts future progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology* 64(5):902–904.
54. Nir TM, et al. (2013) Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *NeuroImage Clin* 3:180–195.
55. Le Bihan D, Johansen-Berg H (2012) Diffusion MRI at 25: Exploring brain tissue structure and function. *NeuroImage* 61(2):324–341.
56. Itoh Y, Amano N, Inoue M, Yagishita S (1997) Scanning electron microscopical study of the neurofibrillary tangles of Alzheimer's disease. *Acta Neuropathol (Berl)* 94(1):78–86.
57. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol (Berl)* 82(4):239–259.
58. Wuttke J, et al. (2012) SPHERES, Julich's high-flux neutron backscattering spectrometer at FRM II. *Rev Sci Instrum* 83(7):075109.
59. Frick B, Gonzalez M (2001) Five years operation of the second generation backscattering spectrometer IN16—a retrospective, recent developments and plans. *Phys B Condens Matter* 301(1–2):8–19.

5.2 Supplementary information

Supplementary Materials & Methods

Expression and purification of unlabeled and perdeuterated htau40 protein

The protocol for expression and purification of the human isoform htau40 has been published elsewhere (1). Briefly, the protein was expressed as a histidine-tagged fusion protein in *Escherichia coli* BL21(DE3). In order to produce hydrogenated protein, the bacteria grew in LB medium in two-liter flasks. The induction of protein expression was done by adding isopropyl-β-D-thio-galactoside (IPTG) to a final concentration of 0.5 mM, once the optical density at 600nm (OD_{600}) reached 0.8. The bacteria were harvested after 3 hours of incubation at 30°C.

In order to obtain perdeuterated protein, a high-cell-density fermentation process with deuterated Enfors minimal medium was used to grow bacteria to an OD_{600} of 12–14, followed by induction of protein expression by IPTG to a final concentration of 0.5 mM. D8-glycerol (fully deuterated glycerol) was used as the carbon source and D_2O as solvent. Bacteria were harvested when an OD_{600} of 20 was reached.

The expressed protein was purified by immobilized-metal-ion-affinity chromatography on a nickel affinity column, followed by size-exclusion chromatography. All purification steps were conducted in H_2O-based buffer systems. Protein purity was assessed by 14% Tris-Tricine SDS-PAGE.

Production of deuterated heparin analogue

E. coli K5 bacteria were grown in Enfors deuterated medium and perdeuterated heparosan (*N*-acetyl-glucosamine - Glucuronic acid repeats) polysaccharides were isolated as described in (2). Chemical modifications were adapted from Casu *et al.* (3) to produce poly *N*- and *O*- sulfated heparosans. Heparosan at 0.5 mg/ml was deacetylated in 2 M NaOH for 16 h at 56°C. pH was adjusted to 7.5 and polysaccharides were extensively dialyzed against water and freeze-dried. *N*-sulfation was achieved by addition to *N*-deacetylated heparosan of 30 mg/ml pyridine-sulfotrioxyde and Na_2CO_3 at 1.5 mg/ml in water for 16 hours at 56°C. The same amounts of pyridine-sulfotrioxyde and Na_2CO_3 were again added and further incubated for 8 hours. For *O*-sulfation, the above described materials (*N*-sulfated heparosans) were

dialyzed against water and run through an Amberlite IR120 column at 4°C. The pH was adjusted to 5.5 with 10% tributylamine (TBA), and the heparosans were freeze-dried. *N*-sulfated heparosans (TBA form) were dissolved in diméthylsulfoxyde at 3 mg/ml with 200 mg/ml pyridine sulfotrioxyde for 2 hours at 25°C. The *O*-sulfation reaction was stopped by a two times dilution in water and pH was adjusted to 8 with NaOH. *O*-sulfated heparosans were precipitated with 80% NaCl-saturated acetone at -20°C, washed in NaCl-saturated acetone, and the final pellet was dried and resuspended in 0.2 M NaCl. A final *N*-sulfation step was then performed before extensively dialysing the polysaccharides against water.

Tau fiber formation and monitoring

See materials-and-methods section in main text for details.

Sample preparation for neutron scattering experiments

After centrifugation, the pellet was lyophilized and dried over P_2O_5 for 4 days on a 4x3 cm^2 flat aluminum sample holder. Dolman and co-workers (4) reported that after such a procedure, lysozyme retains only four tightly bound structural water molecules. The resulting hydration level was defined as 0 g water/g tau.

The deuterated fibers (106 mg of dry protein) was rehydrated to 0.40 g H_2O / g protein (sample referred to as D-fiber-H_2O). The hydrogenated fibers (116 mg of dry protein) was first kept for 17 h under 100% D_2O atmosphere and then dried over P_2O_5 until it reached a hydration level of 0.44 g D_2O/g protein (sample denoted as H-fiber-D_2O). Both samples were then sealed by an aluminum cover (0.3 mm neutron path length) and an indium seal with a diameter of 1 mm.

The deuterated and hydrogenated monomeric protein samples (D-tau-H_2O and H-tau-D_2O) had been prepared earlier (1). We recall that the hydration level was 0.38 g H_2O / g protein and 0.44 g D_2O / g protein for D-tau-H_2O and H-tau-D_2O, respectively. They were reused for the present work without opening the sample holders.

If we assume that all exchangeable hydrogen/deuterium atoms (i.e. 24% of all hydrogen/deuterons in the protein) exchange, the following contributions to the total incoherent scattering cross section of the samples are obtained by summing individual incoherent scattering cross sections of the constituting atoms:

- D-tau-H_2O and D-fiber-H_2O samples: 71% from hydration water (H_2O) and 29% from protein (8% from the deuterated protein and 92% from the deuterons exchanged against hydrogens)

- H-tau-D_2O and H-fiber-D_2O samples: 97% protein (99% from hydrogenated protein and 1% from the hydrogen atoms exchanged against deuterons) and 3% from hydration water (D_2O). Several studies on tau fibrillation triggered by heparin showed that only a small amount of heparin, if at all, is found in the mature fibers (5–8). As an additional precaution, we used deuterated heparin so that, even if the ratio tau:heparin was 1:1 in the mature fibers, the heparin contribution to the total scattering intensity in both fiber samples would be less that 0.5%. Thus, the contribution of heparin in the above calculations was neglected.

Fiber characterization by electron microscopy and X-ray diffraction

Electron microscopy. The samples were negatively stained using the mica floatation technique. 4 µl of the protein sample were adsorbed onto the clean side of a carbon film on a mica sheet (carbon/mica interface) and then, after 30 s, negatively stained with 2% (w/v) uranyl acetate (pH 4.5). Micrographs were taken under low-dose conditions with a CM12 LaB6 electron microscope working at 120 kV and with nominal magnifications of 22000 x and 45000 x using a Gatan OriusTM SC1000 CCD camera. The micrograph of the D-fiber sample is shown in figure 2a of the main text and we show in figure S1a the micrograph of the H-fiber sample. As a control, we also monitored the protein before adding heparin and no fibers were visible (not shown).

X-ray diffraction. The experiments were performed at the European Synchrotron Radiation Facility Synchrotron (Grenoble, France) on beamlines ID14-1 and ID23-2 for the hydrogenated and deuterated fibers, respectively. Hydrogenated fibers were mounted after lyophilization into a nylon loop of 0.5 mm diameter. The sample was then exposed to a monochromatic X-ray beam of 0.93 Å wavelength during 100 s at room temperature, while the sample was rotated by a total of 1 degree. For the deuterated fibers, a drop of fiber solution was dried out between two thin glass capillaries. The dry residue was then exposed to a monochromatic beam of 0.87 Å wavelength during 10 s at room temperature. The diffraction pattern of the deuterated fibers is shown in figure 2b of the main text and the one of hydrogenated fibers figure S1b.

Elastic incoherent neutron scattering (EINS) experiments

Deuterated samples (D-tau-H_2O and D-fiber-H_2O). EINS experiments were carried out on the backscattering spectrometer SPHERES (Jülich Centre for Neutron Science at the Heinz Maier-Leibnitz-Zentrum Garching (MLZ), Garching, Germany). Each sample was inserted at room temperature into a Janis cryostat at 135° with respect to the incoming neutron beam. The temperature was lowered to 20 K during about 2 h. Elastically scattered neutrons were then counted while the temperature was continuously increased from 20 to 300 K at a rate of 0.093 K/min for D-fiber-H_2O and of 0.156 K/min for D-tau-H_2O.

The elastically scattered intensity of the samples were then calculated according to the following formula:

$$I_{sample,T}(q) = \frac{I_{total,T}(q) - tr \times I_{empty}(q)}{\left\langle I_{total,T}(q) - tr \times I_{empty}(q) \right\rangle_{20-40K}} \qquad \text{eq. S1}$$

where $I_{sample,T}(q)$ is the normalized intensity scattered from the sample at the temperature T, corrected for instrument effects and empty-cell scattering; $I_{total,T}(q)$ is the measured intensity; $I_{empty}(q)$ is the scattered intensity from an empty cell measured at 280 K. The brackets represent the elastic intensity averaged in the temperature window 20-40 K, and *tr* is the measured sample transmission.

Because the more dynamic the system is, the less neutrons are elastically scattered, one can sum the elastic intensity over all q values to obtain a qualitative model-free information about the dynamics of the system, as shown in figure S2. More quantitatively, the atomic mean square displacements (MSD, $<u^2>$) were extracted from the q-dependence of the elastic intensity that can be described in the Gaussian approximation by

$$I_{sample,T}(q) = \exp(\frac{-q^2 \left\langle u^2 \right\rangle}{6}) \qquad \text{eq. S2}$$

MSD were extracted in the q-range where the logarithm of the elastic intensity remains linear with respect to q^2, *i.e.* 0.78 Å^{-1} < q < 1.76 Å^{-1} for both D-tau-H_2O and D-fiber-H_2O samples.

Hydrogenated samples (H-tau-D₂O and H-fiber-D₂O). EINS experiments were carried out on the backscattering spectrometer IN16 at the Institut Laue Langevin (ILL; Grenoble, France). H-fiber-D₂O sample was inserted at room temperature into an orange ILL cryostat at 135° with respect to the incoming neutron beam. The temperature was lowered to 20 K over 2 h. Elastically scattered neutrons were then counted while the temperature was continuously increased from 20 to 300 K at a rate of 0.14 K/min. The elastically scattered signals were then processed according to equation S1. Data on H-tau-D₂O, measured on IN16 according to the same protocol, were taken from previous work (1) and re-processed. At high q values, the deviation from the Gaussian approximation expressed in equation S2 is non negligible. Such a deviation can have two different origins: non-Gaussian scattering from single atoms and dynamic heterogeneity. A simple correction with a q^4 term has been proposed (9, 10), in which the scattering intensity an be expressed as follows:

$$I_{sample,T}(q) = \exp(\frac{-q^2 \langle u^2 \rangle}{6}) \times (1 + \frac{q^2}{72}\sigma^2)$$

<div align="right">eq. S3</div>

where σ^2 is the variance of the distribution of the MSD, if we assume that the dynamical heterogeneity is the main contribution to the non-gaussian behavior. An example of the fit from the Gaussian approximation with and without q^4 correction is shown in figure S4. This model allows extracting the MSD (see figure 3c, main text) as well as the MSD variance shown in figure S5, which reflects the MSD heterogeneity among hydrogen atoms in the sample. These two parameters were extracted by fitting equation S3 between the q-values 0.43 and 1.93 Å^{-1} for both H-tau-D₂O and H-fiber-D₂O samples.

Quasi-elastic neutron scattering (QENS)

The model used and the fitting procedure has been extensively discussed in a study that will be published elsewhere (Schiro *et al.*, in press (DOI: 10.1038/ncomms7490)). Briefly we used a rotational-translational diffusion model, where each water molecule is considered to either undergo rotational or translational diffusion, or to be immobile. We used a global fitting approach where all spectra were simultaneously fitted at all q values with the following equation:

$$S(q,\omega) = DW(q)\big[[A_0 \times \delta(\omega) + A_t \times S_t(q,\omega) + A_r \times S_r(q,\omega)] \otimes R(q,\omega) + k(q)\big] \quad \text{eq. S4}$$

where $DW(q) = \exp(-q^2 \langle u^2 \rangle)$ is the Debye-Waller factor due to the vibrational motions. The intensity of the elastic term ($A_0 = A_r \cdot J_0^2(qb) + A_i$, where J_0 is the 0th order spherical Bessel function) contains a q-dependent term from the rotational diffusion model and a q-independent term A_i that takes into account hydrogen atoms not moving in the time scale set by the resolution function. $S_t(q,\omega) \propto \dfrac{Dq^\lambda}{\omega^2 + (Dq^\lambda)^2}$ is the contribution of translational motions; for $\lambda=2$ the standard Brownian diffusion is recovered, while $\lambda<2$ indicates a sub-diffusive character, often found in polymers, glass-forming and supercooled liquids.

$S_r(q,\omega) \propto \sum_{l=1} (2l+1) J_l^2(qb) \dfrac{1}{\pi} \dfrac{l(l+1)\Gamma}{\omega^2 + (l(l+1)\Gamma)^2}$ is the contribution to the quasi-elastic broadening of rotational motions ; b is the radius of the hydrogen rotation around the center of mass of water molecule, that was set to 0.98 Å, *i.e.* the H-O distance (the center of mass position essentially coincides with the oxygen atom position); $j_l(qb)$ are the spherical Bessel functions. The free parameters are in this model are then: $\langle u^2 \rangle$, A_i, A_t, D, λ, A_r, Γ_r, $k(q)$ for a complete dataset at a given temperature.

MD simulations

The molecular dynamics (MD) simulations were performed with the NAMD program (11) using the CHARMM27 force field (12, 13) for the protein and the TIP3P model (14) for water. The simulations were maintained at constant temperature using Langevin dynamics, and a constant pressure of 1 atm using the Nosé-Hoover Langevin piston algorithm with anisotropic cell fluctuations (15, 16). The equations of motion were integrated using the Verlet-I/r-RESPA multiple-time step algorithm (17, 18) with time steps of 2 fs for the long-range non-bonded forces, 2 fs for the short-range non-bonded forces, and 2 fs for the bonded intra-molecular forces. The SHAKE algorithm (19) was used to constrain the lengths of all bonds to H atoms. Electrostatic interactions were computed using the smooth particle-mesh Ewald sum (20), and the van der Waals interactions and the real-space part of the Ewald sum were smoothly switched to zero over the range 10-12 Å.

The monomeric peptide model was built by placing the hexapeptide VQIVYK in a box containing 9253 water molecules and one chloride ion (see the snapshot in figure S6a). An energy minimization was carried out for 20 ps before equilibrating the system for 12 ns. The

fiber model was based on the work of Zhao *et al.* (21). The fiber was made of the antiparallel two-sheets crystal structure published by Sawaya and coworkers (22) (pdb entry 2ON9). Five strands were stacked on each other with a spacing distance of 4.7 Å as shown in figure S6b. The fiber backbone was first harmonically constrained and placed in a box with 10 chloride ions for 10 ns. The system was then equilibrated with 9258 water molecules for 4ns with the same harmonic constrains. Finally the system was equilibrated without constraints for 18 ns.

The time correlation functions shown in figure 4 were computed over a 2-ns-long trajectory saved every 0.1 ps for water molecules within 3 Å of the protein (See figure S6). For a rigorous comparison, only the water molecules around one central peptide of the fiber were considered (See figure S6b). The intermittent HB relaxation is described by the decay of the correlation function $c(t) = \langle h(0)h(t) \rangle / \langle h \rangle$ (23), where $h(t)$ is a hydrogen bond population operator, which is equal to one if a given donor-acceptor (D-A) pair is hydrogen bonded at time t, and zero otherwise, and the angular brackets denote an average over all D-A pairs (figure 4c). The function $c(t)$ is the probability that a random D-A pair that is hydrogen bonded at time zero is still bonded at time t, regardless of whether or not the bond was broken at intermediate times. The timescale of the decay of the continuous HB relaxation time represents the lifetime of protein-water hydrogen bonds that are formed/broken by water rotational/librational motions on the protein surface (figure 4b).

Supplementary figures
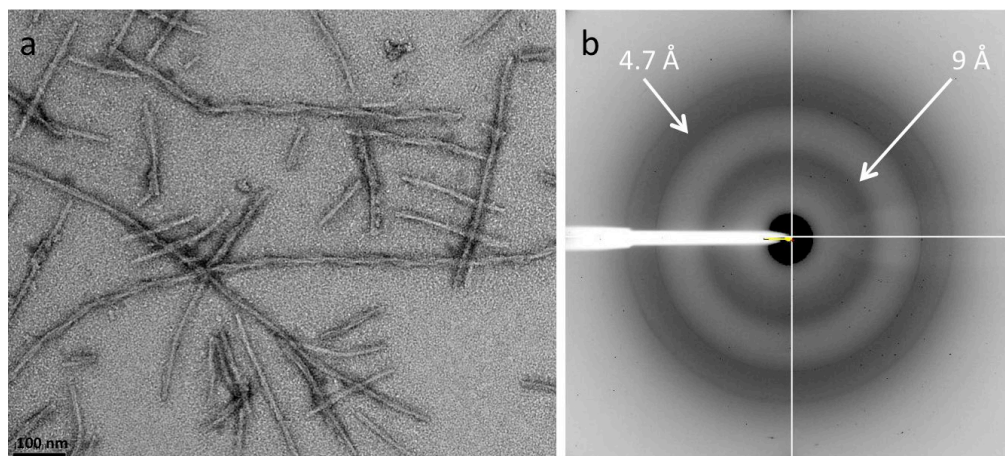
Figure S1: (a) Electron micrograph and (b) X-ray fiber diffraction pattern of hydrogenated tau amyloid fibers (H-fiber).
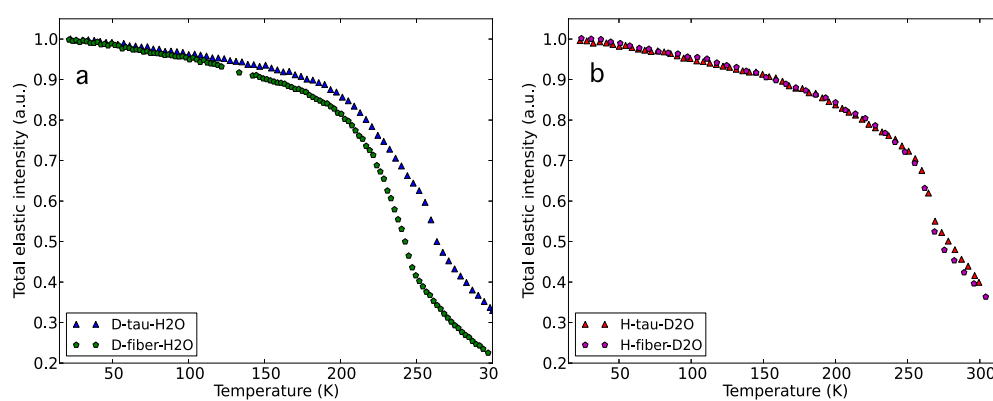
Figure S2: Total elastic intensities integrated over all detectors for (a) D-tau-H2O and D-fiber-H2O (q-range is 0.21-1.84 Å^{-1}) and (b) H-tau-D2O and H-fiber-D2O (q-range is 0.19-1.95 Å^{-1}). Data from H-tau-D2O were taken from (1).
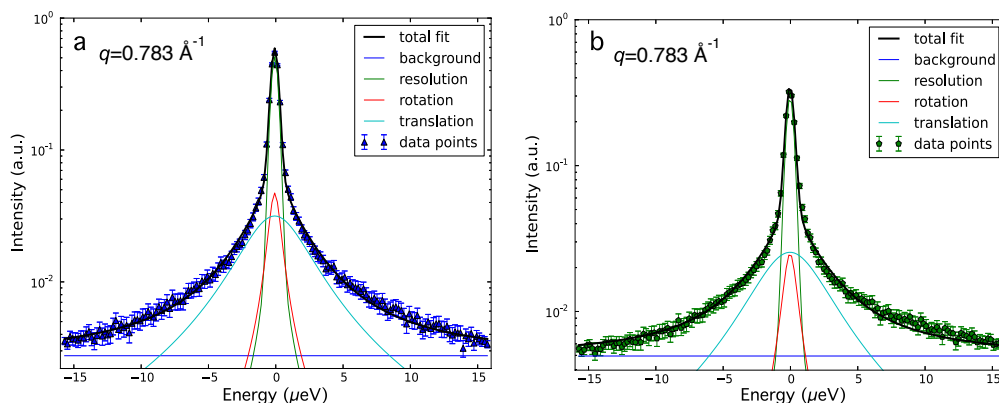
Figure S3: QENS spectra of (a) D-fiber-H_2O and (b) D-tau-H_2O at q=0.78 Å^{-1} and the corresponding fit from the translational-rotational model. Data on D-tau-H_2O were extracted from previous work (Schiro *et al.* in press (DOI: 10.1038/ncomms7490)).
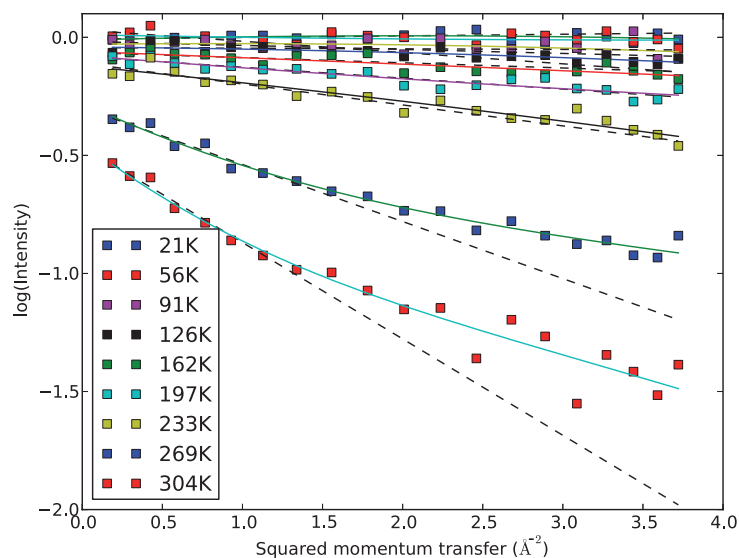
Figure S4: Example of fits from the Gaussian approximation (dashed lines) and the q^4 correction model (continuous lines) applied on the tau amyloid fibers (H-fiber-D_2O).
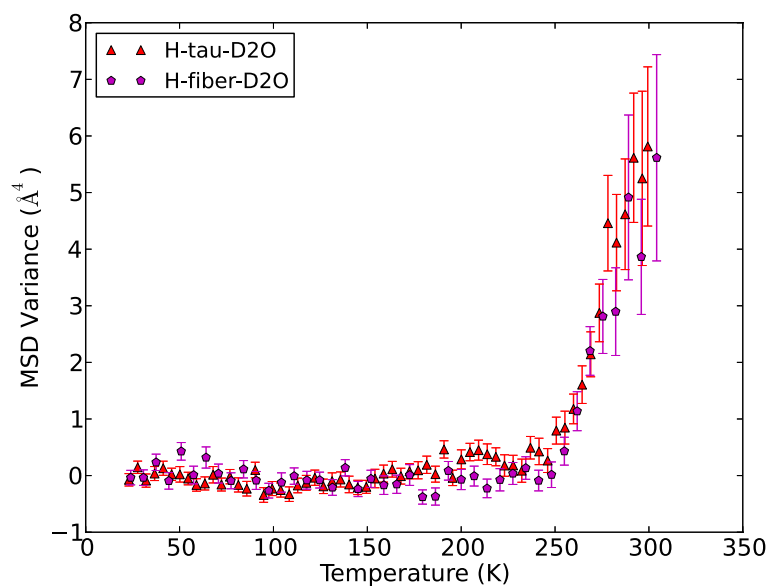
Figure S5: MSD variance obtained from q^4 correction model (eq. S3).
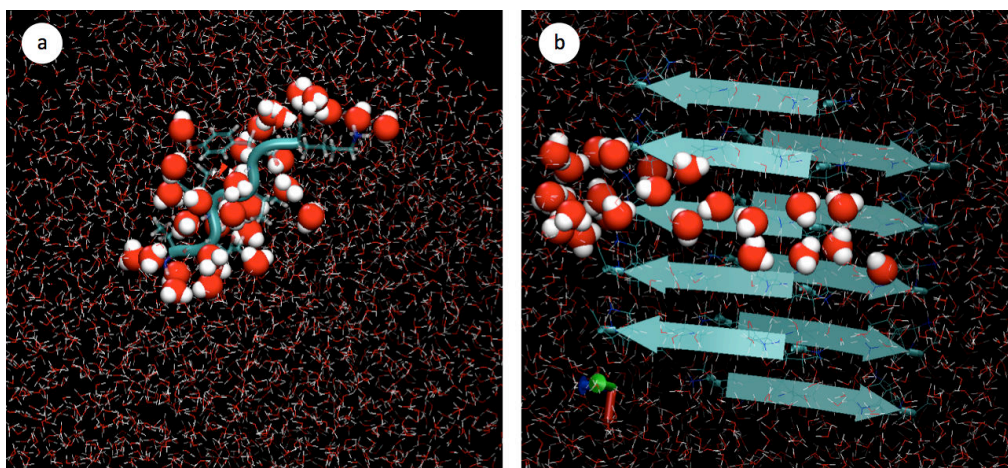
Figure S6: Snapshots of the simulation box of the amyloidogenic peptide [306]VQIVYK[311] (a) as monomer in solution and (b) in an amyloid assembly constructed from the crystal structure 2ON9. The analyzed water molecules are depicted with spheres with van der Waals radii.
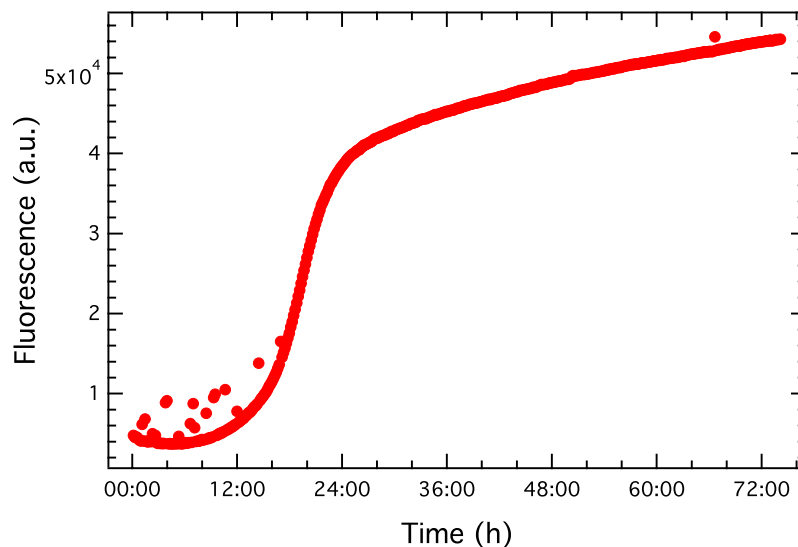
Figure S7: Fluorescence of thioflavin S that was added to the hydrogenated protein mixed with deuterated heparin. t=0 denotes the time point when heparin is added to the protein solution. The solution was excited at 430 nm and the emission measured at 490 nm.

	D-tau-H_2O	D-fiber-H_2O	Change from D-tau-H_2O to D-fiber-H_2O
Immobile fraction	0.04 ± 0.02	0.04 ± 0.03	0 %
Translational fraction	0.381 ± 0.005	0.475 ± 0.008	+ 25 %
Trans. diff. coeff. (10^{-7} cm^2/s)	0.99 ± 0.03	1.10 ± 0.04	+ 11 %
Rotational fraction	0.58 ± 0.03	0.49 ± 0.03	- 16 %
Rotational rate (ps^{-1})	58 ± 9	68 ± 5	+ 17 %
λ	1.80 ± 0.07	1.80 ± 0.08	0 %

Table S1 : Fitting parameters of the QENS spectra of D-fiber-H_2O and D-tau-H_2O at 280 K. The errors represent SD output from the fitting procedure. An example of the fit is shown in figure S3. The third column shows the relative change of the parameters between the two samples.

Supplementary references

1. Gallat F-X, et al. (2012) Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys J* 103(1):129–136.

2. Laguri C, et al. (2011) 13C-Labeled Heparan Sulfate Analogue as a Tool To Study Protein/Heparan Sulfate Interactions by NMR Spectroscopy: Application to the CXCL12α Chemokine. *J Am Chem Soc* 133(25):9642–9645.

3. Casu B, et al. (1994) Heparin-like compounds prepared by chemical modification of capsular polysaccharide from E. coli K5. *Carbohydr Res* 263(2):271–284.

4. Dolman M, Halling PJ, Moore BD, Waldron S (1997) How dry are anhydrous enzymes? Measurement of residual and buried 18O-labeled water molecules using mass spectrometry. *Biopolymers* 41(3):313–321.

5. Von Bergen M, et al. (2006) The Core of Tau-Paired Helical Filaments Studied by Scanning Transmission Electron Microscopy and Limited Proteolysis†. *Biochemistry (Mosc)* 45(20):6446–6457.

6. Sibille N, et al. (2006) Structural Impact of Heparin Binding to Full-Length Tau As Studied by NMR Spectroscopy†. *Biochemistry (Mosc)* 45(41):12560–12572.

7. Carlson SW, et al. (2007) A Complex Mechanism for Inducer Mediated Tau Polymerization†. *Biochemistry (Mosc)* 46(30):8838–8849.

8. Ramachandran G, Udgaonkar JB (2011) Understanding the Kinetic Roles of the Inducer Heparin and of Rod-like Protofibrils during Amyloid Fibril Formation by Tau Protein. *J Biol Chem* 286(45):38948–38959.

9. Becker T, Smith JC (2003) Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular systems. *Phys Rev E Stat Nonlin Soft Matter Phys* 67(2 Pt 1):021904.

10. Yi Z, Miao Y, Baudry J, Jain N, Smith JC (2012) Derivation of Mean-Square Displacements for Protein Dynamics from Elastic Incoherent Neutron Scattering. *J Phys Chem B*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22471396.

11. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–802.

12. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–616.

13. Mackerell AD Jr, Feig M, Brooks CL 3rd (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–15.

14. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.

15. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101(5):4177–4189.

16. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 103(11):4613–4621.

17. Grubmüller H, Heller H, Windemuth A, Schulten K (1991) Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Mol Simul* 6(1-3):121–142.

18. Tuckerman M, Berne BJ, Martyna GJ (1992) Reversible multiple time scale molecular dynamics. *J Chem Phys* 97(3):1990–2001.

19. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23(3):327–341.

20. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593.

21. Zhao J-H, et al. (2010) Molecular dynamics simulations to investigate the stability and aggregation behaviour of the amyloid-forming peptide VQIVYK from tau protein. *Mol Simul* 36(13):1013–1024.

22. Sawaya MR, et al. (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* 447(7143):453–457.

23. Luzar A, Chandler D (1996) Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys Rev Lett* 76(6):928–931.

5.3 Summary and lead-in for the next chapter

Summary The paired helical filaments (PHF) formed by the intrinsically disordered human protein tau are one of the pathological hallmarks of Alzheimer disease. PHF are fibers of amyloid nature that are composed of a rigid core and an unstructured fuzzy coat. The mechanisms of fiber formation, in particular the role that hydration water might play, remains poorly understood. We combined protein deuteration, neutron scattering and all-atom molecular dynamics simulations to study the dynamics of hydration water at the surface of fibers formed by the full-length human protein htau40. In comparison to monomeric tau, hydration water on the surface of tau fibers is more mobile, as evidenced by an increased fraction of translationally diffusing water molecules, a higher diffusion coefficient and increased mean-squared displacements in neutron scattering experiments. Fibers formed by the hexapeptide 306VQIVYK311 were taken as a model for the tau fiber core and studied by molecular dynamics simulations, revealing that hydration water dynamics around the core domain is significantly reduced after fiber formation. Thus, it is an increase in water dynamics around the fuzzy coat that is at the origin of the experimentally observed increase in hydration water dynamics around the entire tau fiber. The observed increase in hydration water dynamics is suggested to promote fiber formation through entropic effects. Detection of the enhanced hydration-water mobility around tau fibers might contribute to the early diagnosis of Alzheimer patients by diffusion MRI.

Introduction to the next chapter We showed in this chapter that the hydration water dynamics of the native and fibrillated tau protein are different, although the dynamics of tau in both states are similar. The latter finding is rather unexpected, given the important structural differences between the native disordered protein and the amyloid fibers. Although a previous study has investigated the dynamics of aggregated concanavalin A powders (Schiro et al., 2012), the lack of data on different proteins does not allow to draw comprehensive conclusions. We present in the next chapter a neutron scattering study of various proteins in different aggregation states, which aim at further understanding the dynamics of protein aggregates on the ns time scale.

Chapter 6

Dynamical diversity of protein aggregates

6.1 Introduction

In certain physicochemical conditions, proteins can change conformations and assemble into aggregates of different morphologies. Protein aggregation is involved in many diseases, so-called conformational diseases, and has been intensively studied. However, the relation between aggregation and cytoxicity remains incompletely understood, and the particular aggregate features (size, structure, dynamics etc.) that are responsible for their toxicity are unknown (Ross and Poirier, 2005). Aggregates can have different shapes and sizes, from small soluble oligomers to macroscopic aggregates. The so-called amyloid aggregates, or amyloid fibers, are highly structured and characterized by inter-protein β-sheet structures (see chapter 1), whereas no common structural patterns have been found for other aggregates, so-called amorphous aggregates.

While a considerable research effort has been devoted to study the structural properties of aggregates, much less is known about their dynamical behavior. A recent neutron scattering study on concanavilin A showed that amyloid aggregates, amorphous aggregates and monomers of the same protein have different dynamical properties on the nanosecond (ns) time scale (Schiro et al., 2012). Yet, the lack of data on several proteins in different aggregation states does not allow to conclude on the existence of general dynamical patterns among protein aggregates.

In this chapter, we report an elastic incoherent neutron scattering (EINS) study aiming at characterizing the dynamical properties of different aggregates. Using protein powders hydrated in heavy water, EINS allows to probe the protein internal dynamics on the ns time scale (see chapter 2.1), which correspond to side-chain motions. We investigated three different systems in different aggregation states, the lysozyme protein, the maltose binding protein (MBP) and the hexapeptides VQIVYK and VQIINK, which are two amylogenic

peptides present in the tau protein (see section 1.4). The measured samples as well as some experimental details are presented in table 6.1.

	Dry protein weight (mg)	Hydration level (g D_2O/g prot.)	Instrument	Temperature increase rate (K/min)	Fit model*	Fitted Q-range (Å^{-2})
Lysozyme monomers	149	0.39	SPHERES	0.23	Q2	0.61-3.7
Lysozyme amorphous aggregates	119	0.40	SPHERES	0.19	Q2	0.61-3.7
Lysozyme amyloid aggregates	113	0.40	SPHERES	0.19	Q2	0.61-3.7
MBP monomers	186	0.40	IN16	0.19	Q2 Q4	0.19-1.34 0.19-3.7
MBP amorphous aggregates	156	0.40	IN16	0.13	Q2 Q4	0.19-1.34 0.19-3.7
VQIVYK monomers	134	0.40	IN16	0.15	Q2 Q4	0.19-1.34 0.19-3.7
VQIVYK amyloid aggregates	165	0.40	IN16	0.15	Q2 Q4	0.19-1.34 0.19-3.7
VQIINK amyloid aggregates	160	0.40	IN16	0.19	Q2 Q4	0.19-1.34 0.19-3.7

Table 6.1: Experimental details of samples for which data are shown in the present chapter.
* Q2 and Q4 refers to equation 6.2 and 6.3, respectively.

6.2 Materials and methods

6.2.1 Proteins aggregation and samples preparation

Lysozyme The lysozyme samples were prepared by Robert Smock in the group of Dan Tawfik at the Weizmann Institute of Science, Rehovot, Israel. Hen egg white lysozyme was used to prepare amyloid and amorphous aggregates following a similar procedure as previously described (Arnaudov and de Vries, 2005; Malisauskas et al., 2003). The protein was incubated in buffered deuterium oxide at 57 °C. The buffers were composed of 20 mM glycine and 0.02 % sodium azide at pH 2.0 for fibrils, and 20 mM Tris and 0.02 % sodium azide at pH 7.0 to favor amorphous aggregates (pH was adjusted by HCl). Reasonable yields of aggregates were found in a variable time frame within weeks or up to two months. The aggregates were washed three times in D_2O by centrifugation at 17,000 g and resuspension, to remove salts and monomers. The aggregates solutions were finally centrifuged at 50,000 g for 30 min. The native lysozyme was incubated in deuterium oxide for several hours, to ensure the exchange of all protein exchangeable hydrogen atoms. Both amorphous and amyloid aggregates pellets, as well as the native protein, were lyophilized

and dried over P_2O_5 for 4 days on 4x3 cm^2 flat aluminum sample holders. Dolman and co-workers (Dolman et al., 1997) reported that after such a procedure, a lysozyme molecule retains only four tightly bound structural water molecules. The resulting hydration level was defined as 0 g water/g protein. The quantity of dry native protein, amyloid fibers and amorphous aggregates were 149, 113 and 119 mg, respectively. These three samples were rehydrated at 0.39, 0.40 and 0.40 g D_2O per gram protein (0.4 D_2O g/g), respectively.

Electon and optical microscopy were used to verify the quality of the samples. In particular, the fiber twists visible by electron microscopy (EM) in the upper picture of figure 6.1 (alternance of thin and thick parts along a fiber) are typical from amyloid fibers, while the amorphous aggregates are only observable by optical microscopy (lower picture of figure 6.1), because of their large size. The absence of fibers among the amorphous aggregates was checked by EM.
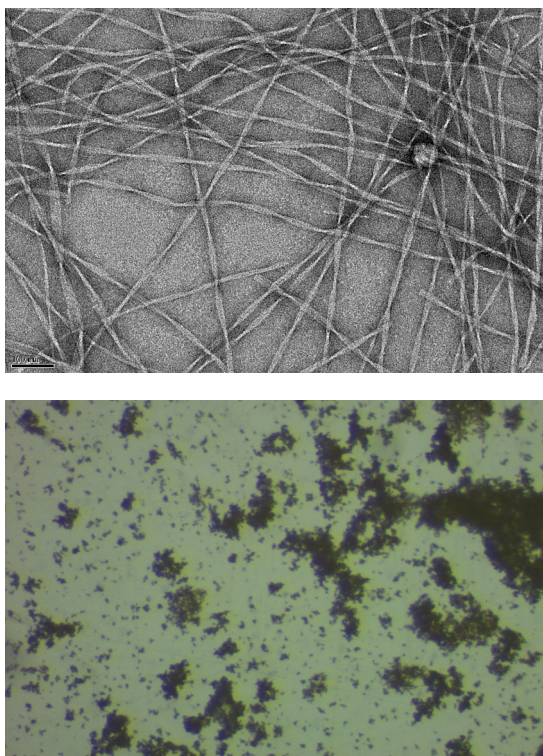
Figure 6.1: Electron micrograph of the amyloid fibers (upper picture) and optical micrograph of the amorphous aggregates (lower picture).

Maltose binding protein (MBP) The MBP was expressed and purified as outlined in chapters 3 and 4. MBP amorphous aggregates were prepared from a purified MBP solution by letting the solvent evaporate in a low-vacuum chamber at 37 °C until it was completely dry. The solvent evaporation led to high protein concentrations, eventually

inducing aggregation/precipitation. The precipitate powder was then dried over P_2O_5 and rehydrated at 0.40 D_2O g/g.

Hexapeptides VQIVYK and VQIINK   Ac-VQIVYK-amide (VQIVYK with an acetylatation and amidation of its N- and C-terminus, respectively, *i.e.* $CH_3 - CO_2-$VQIVYK $-NH_2$) and Ac-VQIINK-amide peptides were purchased from CS Bio Co. as dry powders. The peptide fibrillation was triggered by solubilization in a 0.4 M amonium acetate buffer. Quijano and coworkers (Rojas Quijano et al., 2006) showed that these peptides fibrillate spontaneously in the presence of salt. The fibrillation of both peptides was monitored by fluorescence of thioflavin T (Th T) on a BioTek Synergy H4 Hybrid microplate reader with excitation and emission wavelengths of 450 and 480 nm, respectively. As a control, the peptides were solubilized in pure H_2O and no increase of Th T fluorescence was found. After 8 days of incubation at room temperature in the 0.4 M ammonium acetate buffer, the solutions were centrifuged at 35,000 RPM (rotor Ti 45). The pellets of both peptide fibers, as well as the monomeric VQIVYK, were lyophilized and dried over P_2O_5 for 4 days on 4x3 cm^2 flat aluminum sample holders. The quantity of dry monomeric VQIVYK, amyloid VQIVYK, and amyloid VQIINK were 134, 165 and 160 mg, respectively. All three samples were rehydrated at 0.40 g D_2O per gram protein. The presence of crystalline ice was discarded by performing diffraction experiments at sub-freezing temperatures.

6.2.2   Elastic incoherent neutron scattering (EINS) experiments

EINS on the lysozyme samples was performed on SPHERES (MLZ, Germany). They were inserted at room temperature into a Janis cryostat at 135° with respect to the incoming neutron beam. The MBP and hexapeptides samples were measured on IN16 (ILL, France). They were inserted at room temperature into an Orange cryostat at 135° with respect to the incoming neutron beam. Each sample underwent a temperature ramp from 20K to 300K while measuring the elastic scattered intensity (see table 6.1 for the temperature increase rates). For each experiment, an empty cell was measured at 280 K for several hours (typically two hours).

6.2.3   Data treatment

All experimental data shown here are normalized at low temperature (20-40 K) and corrected by an empty cell measurement, such as:

$$S(Q,\omega) = \frac{S_{meas}(\omega, Q) - S_{empty}(\omega, Q)}{S_{20-40K}(\omega, Q) - S_{empty}(\omega, Q)} \quad , \tag{6.1}$$

where $S(Q,\omega)$ is the structure factor of the sample, $S_{meas}(\omega, Q)$ is the measured elastic intensity from the sample, $S_{empty}(\omega, Q)$ is the measured elastic intensity from an empty cell at 280 K, and $S_{20-40K}(\omega, Q)$ is the elastic intensity from the sample averaged between 20 and 40 K.

The Q-dependance of the elastic intensity was fit with a Gaussian approximation with (equation 6.3) or without (equation 6.2) a Q^4 correction :

$$S(Q, \omega = 0) \simeq exp(-\frac{1}{6}Q^2 < \Delta r^2 >) \tag{6.2}$$

$$S(Q, \omega = 0) \simeq exp(-\frac{1}{6}Q^2 < \Delta r^2 >) \times (1 + \frac{Q^4}{72}\sigma^2) \tag{6.3}$$

$< \Delta r^2 >$ represents the mean square displacements (MSD) averaged over all atoms, and σ^2 is a variance $\sigma^2 = \frac{1}{N}\sum_{k=0}^{N}(< \Delta r_k^2 > - < \Delta r^2 >)^2$. Equation 6.3 was proposed by (Yi et al., 2012) to take into account the deviation from the first order Gaussian approximation. The authors showed that this deviation originates mainly from the dynamical heterogeneity among hydrogen atoms in the sample, which is quantified by the parameter σ^2. Equation 6.2 could fit with acceptable error bars the data measured on both IN16 and SHPERES. However, because the model with a Q^4 correction has an addition free parameter (σ^2) compared the simple Gaussian approximation, it could not fit the data acquired on SPHERES that possesses only 16 detectors (against 22 on IN16), among which some systematic deviations are observed (for instance the elastic intensity at Q=0.61 Å^{-2} is always low, independently of the temperature or the sample, as visible in the left panel of figure 6.6).

Above the dynamical transition, MSD are often found to be linear with respect to the temperature. This linearity reflects a situation where the average potential of the system can be considered as harmonic, although the individual potentials seen by the different atoms are not harmonic (Zaccai, 2000). The effective force constant, corresponding to this average harmonic oscillator, is inversely proportional to the slope of the MSD as a function of temperature. This force constant, expressed in N/m, allows to quantify the resilience of the protein in the high temperature range (typically 260-300 K).

6.3 Results

6.3.1 EINS experiments on MBP

The dynamics of amorphous aggregates and monomers of the MBP were measured by EINS in the temperature range 20-300 K, on the backspectrometer IN16. The absence of crystalline ice was verified by measuring the diffraction of the samples at sub-freezing temperatures on IN16. No Bragg peaks were found, especially at 1.61, 1.72 and 1.82 Å^{-1}, where diffraction from ice (100), (002), (101) are expected ($\lambda = 6.27$Å) , respectively. The logarithm of the elastic intensity as a function of Q^2, as well as the fit from the Gaussian approximation are shown in figure 6.2. The linear fit (equation 6.2) can fit the low Q values, while the Q^4 correction (equation 6.3) allows to fit a larger Q range. The MSD extracted from the Gaussian approxiamtion with and without a Q^4 correction are shown in figures 6.4 and 6.3, respectively. Although the MSD extracted with the Q^4-correction model are slightly higher than those obtained from the simple Gaussian approximation, both models

show that the aggregated MBP is more dynamic above 260 K than the monomeric MBP. Moreover, the steep increase of aggregates MSD with respect to temperature reveals a lower resilience of the aggregates compared to the monomers. The MSD variances, shown in figure 6.4, exhibit no significant differences between the MSD distributions of the two samples.
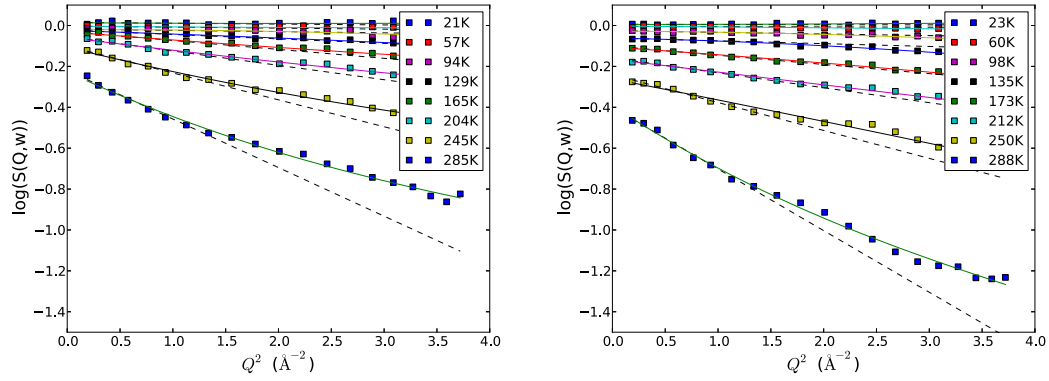
Figure 6.2: Logarithm of the elastic intensity as a function of Q^2 for the monomeric (right panel) and aggregated (left panel) MBP, and the associated fit from equation 6.2 (solid lines) and equation 6.3 (dashed lines).
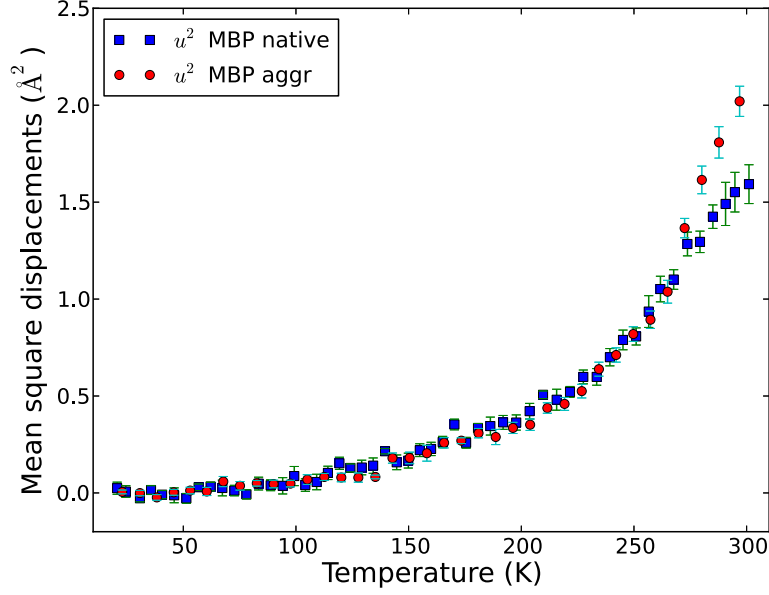
Figure 6.3: MSD of aggregated and native MBP, extracted from equation 6.2 fitted in the Q range 0.19-1.34 Å^{-2}.
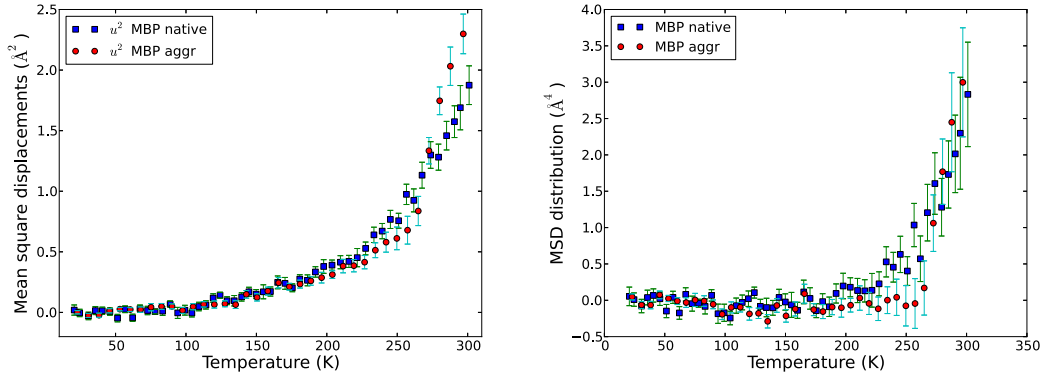
Figure 6.4: MSD of aggregated and native MBP (left panel) , extracted from equation 6.3 in the Q range 0.19-3.7 Å^{-2}, and their associated variance σ^2 (right panel) .

6.3.2 EINS experiments on lysozyme

The dynamics of the lysozyme protein in its native, aggregated (amorphous) and amyloid forms have been measured in the temperature range 20-300 K, by EINS on the backspec-

143

trometer SPHERES. The scattered elastic intensities summed over all detectors, which provide a model-free qualitative information on the system dynamics, are shown in figure 6.5. The logarithm of the elastic intensity scattered from the amyloid sample as a function of Q^2, as well as the associated fit from the Gaussian approximation (equation 6.2), are shown in the left panel of figure 6.6. The extracted MSD are shown in the right panel of figure 6.6.
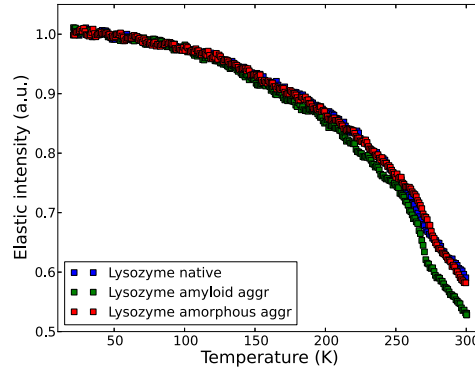
Figure 6.5: Total elastic intensity summed over all detectors, from the native protein, amorphous aggregates and amyloid aggregates of MBP.
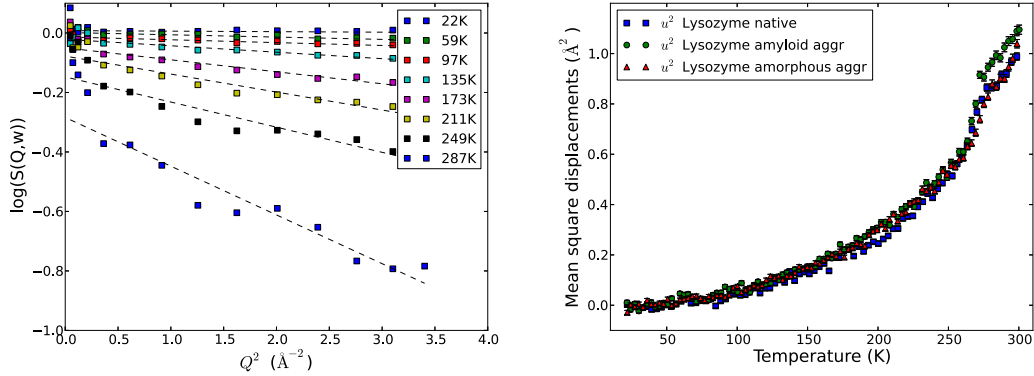
Figure 6.6: Logarithm of the elastic intensity as a function of Q^2 for the amyloid aggregates of the lysozyme, and the associated linear fits from equation 6.2 (left panel). MSD of the native protein, amorphous aggregates and amyloid aggregates, extracted from the Gaussian approximation applied in the Q range 0.61-3.7 Å^{-2} (right panel).

Monomers and unstructured aggregates exhibit a very similar dynamical behavior. However, The MSD of the amyloid fibers reveal a slightly enhanced dynamics above 266 K, which is confirmed by the total elastic intensities shown in figure 6.5. The slopes of the

MSD as a function of temperature decrease around 274K, which likely originates from atomic motions leaving the instrumental spatial window.

6.3.3 EINS experiments on the hexapeptides VQIVYK and VQIINK

Comparison of the different hexapeptide samples

We compare the dynamics of the hexapetide VQIVYK in the temperature range 20-300 K, in its monomeric and fibrillated forms, and the fibrillated VQIINK peptide. The absence of crystalline ice was verify by measuring the diffraction of the samples at sub-freezing temperatures on IN16. No Bragg peaks were found, especially at 1.61, 1.72 and 1.82 Å^{-1}, where diffraction from ice (100), (002), (101) are expected ($\lambda = 6.27\text{Å}$), respectively. The logarithm of the elastic intensity as a function of Q^2 is shown in figure 6.7 for the monomeric and fibrillated VQIVYK peptide. Both fits from the simple Gaussian approximation and the Q^4-correction model fit correctly the data. Both forms of the peptide VQIVYK exhibit a similar dynamics, which is slightly lower than that of the fibrillated peptide VQIINK above 240 K, as presented in figure 6.8.
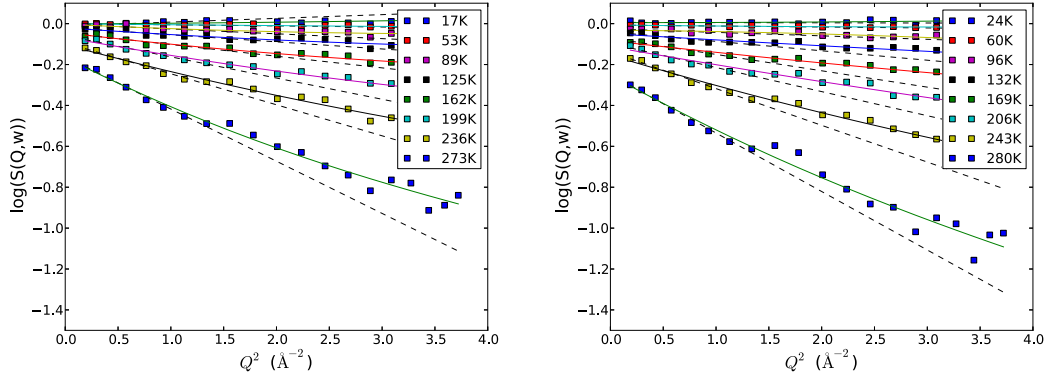
Figure 6.7: Logarithm of the elastic intensity as a function of Q^2 for the (left panel) monomeric and (right panel) fibrillated hexapeptide VQIVYK, and the associated fits from equation 6.2 (solid lines) and 6.3 (broken lines).
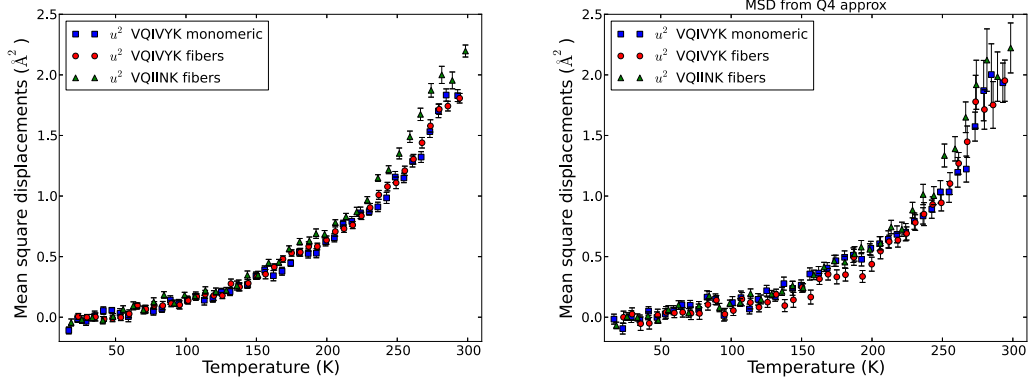
Figure 6.8: MSD of the hexapeptides VQIVYK and VQIINK, extracted from equation 6.2 fitted in the Q range 0.19-1.34 Å^{-2} (left panel) and from equation 6.3 fitted in the Q range 0.19-3.7 Å^{-2} (right panel).

Comparison of hexapeptides and tau full length dynamics

Both $^{275}VQIINK^{280}$ and $^{306}VQIVYK^{311}$ hexapeptides have been identified as key sequences for the tau protein fiber formation (see section 1.4), and, thus, are sometimes used as models of the tau fibers (see chapter 5). Here we compare the dynamics of VQIVYK with the dynamics of the full-length htau40, for which data have been presented in chapter 5. We recall that the dynamics of the native and fibrillated htau40 was found to be identical by EINS experiments performed on IN16.

Figure 6.9 shows the MSD of the monomeric peptide VQIVYK and htau40. The MSD of the peptide increase at lower temperature than those of the full-length protein, whereas the dynamics of the latter increases more drastically above around 240 K. The first onset of dynamics occurring at around 100-150 K is associated with methyl groups rotations entering the dynamical window of the spectrometer (Roh et al., 2005; Wood et al., 2010; Krishnan et al., 2008; Schiro et al., 2010a,b). The contribution of methyl group appears to be more important in the peptide sample compared to htau40. It is consistent with the larger proportion of hydrogen atoms belonging to methyl groups in VQIVYK (about 30 % of all hydrogen atoms belongs to methyl groups) than in htau40 (about 17 %). In contrast, the large-amplitude motions observed at temperatures higher than about 240 K are more important in the full-length tau sample. This is highlighted by the linear fit of MSD above 250 K, whose slope is larger for the full-length protein, reflecting a lower resilience. The variances, shown in figure 6.9, reveal a lower MSD distribution above 260 K for the peptide than for htau40, as one might expect from the larger chemical diversity in the full-length protein.
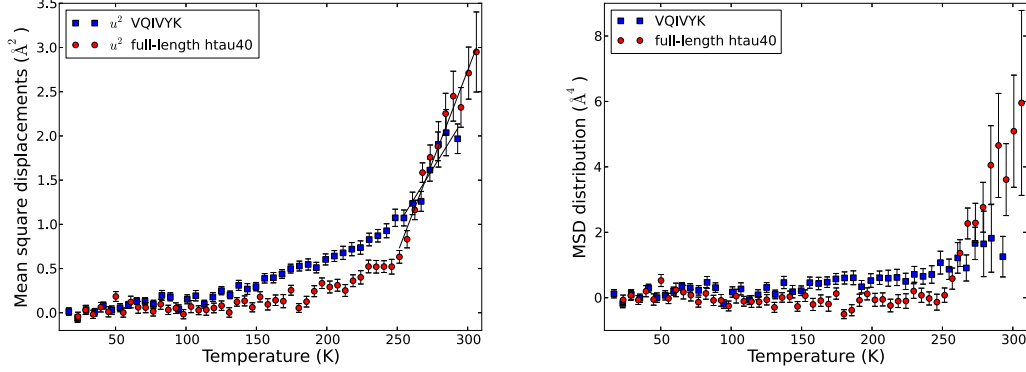
Figure 6.9: MSD of the monomeric hexapeptide VQIVYK and the full-length tau protein, extracted from the Gaussian approximation with a Q^4 correction (equation 6.3) applied in the Q range 0.19-3.7 Å^{-2} (left panel) . MSD variances σ^2 (right panel) .

6.4 Discussion and conclusion

We measured the dynamics of several proteins in different aggregation states by applying EINS. The experiments were carried out on the backscattering spectrometers SPHERES (MLZ, Germany) and IN16 (ILL, France), which probe motions on the ps to ns time scale. The data were analyzed using two models: (i) the commonly-used Gaussian approximation, which allows to extract average MSD from the low-Q dependence of the elastic intensity, and (ii) the Gaussian approximation with a Q^4 correction, proposed by (Yi et al., 2012), which allows to extract the MSDs, as well as the MSD variance from the Q dependence of the elastic intensity in a large Q range. Both models are suitable to fit elastic data measured on IN16 and they output similar values of MSD. The Gaussian approximation with the correction presents the advantage of providing an indicator of the heterogeneity of MSD among the sample (without presupposing the shape of this distribution), which is an additional dynamical property of the samples.

In order to present an overall comparison of the variation of protein dynamics according to the aggregation state, we report in table 6.2, for each sample, the change of dynamics compared to the native form (the MSD values in the temperature range 260-300 K are used to represent the dynamics of the samples). The measurements of concanavalin A carried out on IN16 by Giorgio Schiro and coworkers (Schiro et al., 2012) are also reported. Overall, there is not an obvious correlation between the aggregation state of a protein and its dynamics on the ns time scale. Indeed, while the dynamics of the amyloid form of htau40 and VQIVYK are found to be unchanged compared to their monomers, the lysozyme and the concanavalin A exhibit an enhanced dynamics in their amyloid state. Similarly, the dynamics of amorphous aggregates from lysozyme, MBP, and concanavalin A are found to be similar, higher and lower than the monomers, respectively.

147

	Hydration level (g D_2O/g prot.)	Amyloid aggregates	Amorphous aggregates
Lysozyme	0.4	+	=
MBP	0.4	ND	+
VQIVYK	0.4	=	ND
htau40	0.4	=	ND
Concanavalin A (Schiro et al., 2012)	0.2	+	-

Table 6.2: Comparison of the dynamics, above the dynamical transition (*i.e.* in the range 260-300 K), of protein aggregates with respect to the dynamics of their monomer. + , - , and = symbols mean that the considered aggregate is more, less or equally dynamic than the monomer, respectively. ND marks the forms of aggregates for which no data are available.

When comparing the dynamics of amyloid aggregates with that of monomers, it appears unchanged for tau and for VQIVYK, whereas it is higher for lysozyme and concanavalin A. The amorphous aggregates exhibit an enhanced, similar, and reduced dynamics compared to the monomers for MBP, lysozyme and concanavalin A, respectively. These results highlight a diversity of protein aggregate dynamics on the ns time scale, which is apparently not related to their structure. It is interesting to note that the aggregates present predominantly a similar or enhanced dynamics compared to monomers (it is found for six out of seven samples), which contrasts with the intuitive picture of protein aggregates as stiff and tightly packed assemblies. Rather, it suggests that the protein side chains are more or equally free to move in the aggregated than in the native forms of proteins.

When proteins assemble into aggregates, their secondary and tertiary native structures change. However, not much is known about these structural changes, because very few techniques can characterize protein aggregate structures. Some techniques, such as Fourier transform infrared spectroscopy or circular dichroism, allow to quantify the proportions of secondary structure elements, but they do not provide any spatially resolved information. Even in amyloid aggregates, which by definition are cross-β rich aggregates, it is hard to evaluate how many residues are involved in the inter-protein β sheets and what changes occur in the other parts of the protein. Therefore, one possible origin of the dynamical heterogeneity observed among different aggregates (table 6.2) is that it reflects diverse structural changes that occurred when the different proteins aggregated. Nevertheless, this hypothesis likely does not account for all the presented results, because, although the structures of the amyloid tau and VQIVYK aggregates differ from those of the monomers, their dynamics is unchanged.

One way to further investigate the origins of the measured average protein dynamics is to use MD simulations, which probe motions on the same time scale than neutron scattering, to scrutinize protein dynamics in hydrated powders. In particular, by comparing the average computed dynamics with experimental data, one can use MD simulations to further interpret neutron scattering data, as it has been done in chapters 3 and 5, and elsewhere

(see references 16,24,25,40-47 in chapter 3 for instance). In the next section, we present a study of two hydrated powders, from the monomeric tau protein and MBP, in which we analyze the molecular origins of the protein dynamical features on the ns time scale. By extrapolating the MD analysis done on monomeric proteins to aggregated proteins, we aim to further understand the results presented in this chapter.

Chapter 7

Molecular dynamics simulations of a powder model of the intrinsically disordered human tau protein

The manuscript presented herein is to be submitted to the *Journal of Physical Chemistry B* shortly.

7.1 Manuscript

Yann Fichou[1,2,3], Matthias Heyden[4], Martin Weik[1,2,3], Douglas J. Tobias[5]

[1]Univ. Grenoble Alpes, IBS, F-38044 Grenoble, France; [2]CNRS, IBS, F-38044 Grenoble, France; [3]CEA, IBS, F-38044 Grenoble, France; [4]Max-Planck-Institut fur Kohlenforschung, D-45470 Mulheim an der Ruhr, Germany; [5]Department of Chemistry, University of California, Irvine, California, USA 92697-2025

Abstract

The tau protein, whose aggregates are involved in Alzheimer's disease, is an intrinsically disordered protein (IDP) that regulates microtubule activity in neurons. An IDP lacks a single, well-defined structure and, rather, constantly exchanges among multiple conformations. In order to study IDP dynamics, the combination of experimental techniques, such as neutron scattering, and computational techniques, such as molecular dynamics (MD) simulations, is a powerful approach. Hydrated powder samples have been very useful for studying protein internal dynamics experimentally, e.g. by using neutron scattering. Thus, there is demand for realistic *in silico* models of hydrated protein powders. Here we present an MD simulation analysis of a powder of the IDP tau in the temperature range 20-300 K. By comparing with neutron scattering data, we identify the protein-water interface as the predominant feature determining IDP dynamics. The so-called protein dynamical transition is shown to be attenuated, but not suppressed, in the parts of the protein that are not exposed to the solvent. In addition, we find similarities in the dynamical properties of the core of a globular protein and dry clusters formed by the IDP in hydrated powders. Thus, the dynamical features of proteins in hydrated powders originate mainly from those residues in contact with solvent. We propose that by measuring the dynamics of protein assemblies, such as aggregates, one might assess their state of hydration.

Introduction

Intrinsically disordered proteins (IDPs) are a class of proteins that lack a well-defined three-dimensional structure. Although only recently discovered, they are ubiquitous in all kingdoms of life[1] and fulfill a wide variety of biological functions[2]. IDPs have attracted attention because of their implication in many maladies, including neurodegenerative diseases[3]. The tau protein is an IDP that regulates microtubule activity in neuronal cells. Tau is known to aggregate into amyloid fibers that are one of the pathological hallmarks of Alzheimer's disease[4]. The longest human isoform, htau40, is the most common in human brains and is composed of 441 amino acids, which are entirely disordered in solution. Because of their lack of a stable unique structure, IDPs constantly exchange between different conformations. Thus, they have been shown to be more flexible and dynamic than globular proteins[5–9].

Conformational flexibility is essential for the biological function of folded and intrinsically disordered proteins. In particular, large-amplitude anharmonic motions, observed above the temperature of the so-called dynamical transition[10] (DT), have been proposed to enable protein function[11–14]. The origin of the DT remains debated, in particular, whether it is an effect purely originating from finite instrumental time resolution[15–18] or a resolution-independent transition[19] that originates from a transition in the solvent[20–22], or a mixture of both. Irrespective of its origin, the DT is commonly accepted to be hydration dependent and suppressed in the absence of hydration water[23–25]. Thus, the question arises as to whether or not biologically relevant anharmonic protein motions, which are enabled by the presence of hydration water, are also present in the dry core of proteins. Two computational studies on protein powders have shown that these anharmonic motions animate the entire protein, although they are attenuated in the protein core[26,27], while a recent neutron scattering study did not provide evidence for an attenuation of water-enabled motions in the core of a globular protein[28].

The study of hydrated protein powders is common in several experimental techniques, including NS, solid-state NMR, and dielectric spectroscopy, because they allow for: (i) focusing on the internal protein dynamics, (ii) studying the first hydration layer of proteins, and (iii) exploring a large temperature range without ice formation. The combination of computational and experimental techniques, such as molecular dynamics (MD) simulation and neutron scattering, has proven to be a powerful approach to reveal dynamical mechanisms at a molecular level[29–31]. Indeed, MD simulations allow studying dynamical processes with full atomistic resolution on timescales accessible to several experimental techniques, on the order of pico- to nanoseconds. Thus, after validation by comparison with experimental data, MD simulations are extremely useful for revealing details of dynamical processes that are not experimentally accessible. In order to combine experiments on hydrated protein powders with MD simulations, simulation protocols for realistic *in silico* modeling of hydrated protein powders need to be established. Although MD simulations of hydrated protein powders have been reported for several globular proteins, including RNase[32], myoglobin[33], lysozyme[34], maltose binding protein (MBP)[30], and green fluorescent proteins[35], we are not aware of any hydrated powder models of IDPs.

Here we report MD simulations of a powder model of the IDP tau (isoform htau40). We simulate its protein dynamics in the temperature range 20-300 K, compare simulated

dynamical properties to NS data, and highlight the key requirements for obtaining an IDP powder model exhibiting realistic protein dynamics. We show that, in hydrated powders of both globular proteins and IDPs , the DT is attenuated, although not suppressed, in the solvent-buried parts of the proteins. By comparing powders formed by the IDP tau and the globular protein MBP, we show that protein dynamics is essentially independent of tertiary, secondary, and primary structure, with the exception of the presence of methyl groups, which are more dynamic than the rest of the protein. Rather, the protein dynamics is determined primarily by the features of the protein-water interface in hydrated powders of both tau and MBP. Our results show that in order to model a hydrated IDP powder that accurately reproduces protein dynamics, one has to pay particular attention to the protein-water interface. In addition, we propose that by measuring the dynamics of protein aggregates in hydrated powders, one might assess their state of hydration.

Methods

Construction of the powder models

The MBP powder model was prepared as described earlier[30]. Briefly, the powder consists of four molecules of MBP (PDB entry 1JW4) and 3,460 water molecules, corresponding closely to the hydration level of the samples used in NS experiments[30], i.e. 0.4 g H_2O / g protein. The tau powder is composed of ten tau molecules and was prepared as follows. Out of an ensemble of 200 tau conformations determined by NMR and SAXS experiments[36], we randomly selected a sub-ensemble of ten conformations, for which the average radius of gyration matched the experimental value measured in solution (62 Å)[8,37]. Those ten conformations were placed in the same cubic simulation box before adding 10,105 water molecules (corresponding to the experimental hydration level of 0.4 g H_2O/g protein[8]) spread over the first hydration shell of each protein. We then started the simulation at constant temperature and pressure, so that the simulation box collapsed. After the box dimensions had stabilized we set the temperature to 500 K for 1 ns. The temperature was then reduced to 300 K, the simulation was prolonged for an additional 2 ns, and the final configuration was used as a starting point for the simulations at a series of temperatures. The two powder models were equilibrated for about 30 ns at 20, 150, 200, 230, 240, 260, 280, 300 K for MBP, and 20, 150, 200, 220, 240, 260, 280, and 300 K for tau, followed by at least 14 ns of simulations at

each temperature that served as the basis for subsequent analysis. Snapshots from simulations of the tau and MBP powder models are shown in Figure 1.

MD simulation protocols

The MD simulations were performed using the NAMD[38] program with the CHARMM27 force field[39,40] for the protein and the SPC/E[41] model for water. The simulations were carried out at constant temperature using Langevin dynamics, and a constant pressure of 1 atm using the Nosé-Hoover-Langevin piston algorithm with anisotropic cell fluctuations[42,43]. The equations of motion were integrated using the Verlet-I/r-RESPA multiple-time step algorithm[44,45] with time steps of 4 fs for the long-range non-bonded forces, 2 fs for the short-range non-bonded forces, and 2 fs for the bonded intra-molecular forces. The SHAKE algorithm[46] was used to constrain the lengths of all bonds to H atoms. Electrostatic interactions were computed using the smooth particle-mesh Ewald sum[47], and the van der Waals interactions and the real-space part of the Ewald sum were smoothly switched to zero over the range 10-11 Å for the MBP powder and 10-12 Å for the tau powder.

Trajectory analysis

For a direct comparison with neutron scattering data, we computed atomic mean-squared displacements (MSDs) from the dynamical structure factor $S(\mathbf{q},\omega)$, which can be expressed as the Fourier transform of the intermediate scattering function $I(\mathbf{q},t)$:

$$S(\mathbf{q},\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} I(\mathbf{q},t) e^{-i\omega t} \; ;$$

$$I(\mathbf{q},t) = \frac{1}{N} \sum_{k=1}^{N} \left\langle e^{i\mathbf{q}\cdot\mathbf{r}_k(t)} e^{-i\mathbf{q}\cdot\mathbf{r}_k(0)} \right\rangle,$$

where $\hbar\omega$ is the energy transfer, $\hbar\mathbf{q}$ the momentum transfer, N the number of scatterers, $\mathbf{r}_k(t)$ the position of the kth scatterer at time t, and the angular brackets denote an average over time origins. We computed powder averaged $I(q,t)$ by averaging over ten randomly oriented $\mathbf{q}$ vectors for each of the following values of $q = |\mathbf{q}|$: 0.1914, 0.2928, 0.4338, 0.5440, 0.6525, 0.7589, 0.8754, 0.9640, 1.0621, 1.1567, 1.2476, 1.3346, 1.4172, 1.4953, 1.5685, 1.6367, 1.6997, 1.7571, 1.809, 1.855, 1.895, and 1.929 Å^{-1}, chosen to match the momentum transfer values accessible by the spectrometer IN16[48] at the Institute Laue-Langevin (Grenoble, France). The computed intermediate scattering functions were then numerically Fourier

transformed to obtain the $S(q,\omega)$, which were convoluted with a Gaussian function of ω with a FWHM of $\Delta E = \hbar\Delta\omega = 0.9\ \mu\text{eV}$, meant to represent the instrumental resolution function of the IN16 spectrometer and, consequently, the accessible timescale for dynamics. As in the analysis of neutron scattering data, the MSDs $\langle u^2 \rangle$ plotted in Figure 2 were obtained by fitting the q^2-dependence of the elastic intensity using the Gaussian approximation:

$$S(q,0) = e^{-q^2\langle u^2\rangle/6}.$$

The MSDs reported in Figures 3-6 were calculated directly from the trajectories (*i.e.,* not through the calculation of the structure factor) according to:

$$\text{MSD}(t) = \frac{1}{N}\sum_{k=1}^{N}\left\langle \left| \mathbf{r}_k(t) - \mathbf{r}_k(0) \right|^2 \right\rangle,$$

where the angular brackets denote an average over time origins. For the calculations of MSD(*t*) and *I*(*q,t*), only the protein internal dynamics were considered by aligning each protein with its initial configuration so as to minimize the root mean-square deviation of the backbone Cα atoms. The values plotted in Figures 3-6 are MSD(*t* = 500 ps). The value t=500 ps was chosen because, throughout the explored temperature range, the calculated MSD(t=500ps) were close to the MSDs $\langle u^2 \rangle$ extracted from the Gaussian approximation applied to the scattering function $S(q,\omega)$ calculated using the resolution of the spectrometer IN16 (ΔE=0.9 μeV). The water MSDs around different tau molecules, reported in Figure 4b, were extracted after 100 ps, a time interval during which the exchange of water molecules between the hydration shells of the different tau molecules is very small.

Intra-protein and protein-water hydrogen bond dynamics are characterized in Figure 7 based on the time correlation function[49]:

$$c(t) = \langle h(0)h(t) \rangle / \langle h \rangle,$$

where *h*(*t*) is a hydrogen bond population operator, which is equal to one if a given donor-acceptor (D-A) pair is hydrogen bonded at time *t*, and zero otherwise, and the angular brackets denote an average over all D-A pairs and time origins. The function *c*(*t*) is the probability that a random D-A pair that is hydrogen bonded at time zero is still bonded at time *t*, regardless of whether or not the bond was broken at intermediate times. A hydrogen bond

was considered to exist when the D-A distance was less than 3.5 Å and the D-H-A angle was greater than 150°.

The protein atoms that were not exposed to the solvent in hydrated powders (denoted as solvent-buried atoms in the manuscript) were defined as being further than 4 Å from any water atom. The protein atoms exposed to the solvent (denoted as solvent-exposed) were defined as being closer than 2.5 Å from any water atoms.

Results and discussion

Comparison between MD simulations and neutron scattering experiments.

We modeled a hydrated powder of the intrinsically disordered human protein tau (isoform htau40). For comparison, we also analyzed a powder model of the globular maltose binding protein (MBP), which has a similar molecular weight as tau. We chose the hydration level of 0.4 g H_2O /g protein for both models in order to match previous experimental work on both systems[8,30]. This hydration level was shown to correspond roughly to the formation of the first hydration layer of a globular protein. A snapshot of the simulation box of both powder models is shown in Figure 1. From simulations of both protein powder models at different temperatures, we computed the incoherent dynamical structure factor $S(q,\omega)$, which reflects the individual atomic dynamics and can be obtained by incoherent neutron scattering measurements. By applying the Gaussian approximation, one can extract from the elastic intensity $S(q,0)$ the atomic mean-squared displacements (MSDs). We calculate $S(q,\omega)$ for non-exchangeable protein hydrogen atoms, which represent > 90% of the scattering signal measured by neutron scattering on hydrogenated proteins hydrated by D_2O[8]. The temperature dependence of the MSDs obtained from the measured[8,30] and simulated structure factors of the tau protein and the MBP are displayed in Figure 2. Experimental and computed MSDs from the MBP powder (Figure 2b) show good agreement over the entire temperature range. In addition, the dynamics among the four protein molecules comprising the powder model is fairly homogeneous at all temperatures. Although the agreement of the tau protein MSDs is good up to 260 K, discrepancies appearing at higher temperature reveal that the simulation underestimates the dynamics of hydrogen atoms above 260 K (Figure 2a). Interestingly, the dynamics of the ten molecules present in the powder model differs substantially among each other, particularly at high temperature.

In order to gain insight into the origin of the discrepancy of the tau protein dynamics at high temperature in experiments and simulations, we computed the MSDs directly from the trajectories (*i.e.* without using the structure factor) at different temperatures. We compared the MSDs of the solvent-exposed and solvent-buried hydrogen atoms after 500 ps (Figure 3), which is roughly the timescale of the slowest motion probed by the neutron scattering experiments to which we compare. Buried hydrogen atoms were defined as being further than 4 Å from any water atom. Protein hydrogens closer than 2.5 Å from water atoms were considered as exposed. In both the tau (Figure 3a) and MBP (Figure 3b) powder models, the dynamics of solvent-exposed atoms are enhanced at high temperature compared to the buried atoms. Thus, we hypothesize that the observed discrepancy between the computed and experimental tau protein dynamics, *i.e.* low simulated MSDs at high temperatures (Figure 2a), originates from the solvent-exposed atoms. This hypothesis is in agreement with the notion that below the DT the dynamics of a dry and a hydrated protein are similar, whereas above about 240 K the protein dynamics becomes hydration dependent[23,24].

At the chosen level of hydration (0.4 g H_2O / g protein) the water molecules essentially fill the first hydration layer of the MBP. However, due to its lack of tertiary structure, tau possesses, compared to MBP, a three-times-larger solvent accessible surface area (SASA)[8], which cannot be entirely covered by water at this hydration level. Thus, in our powder model tau forms dry clusters of hydrophilic residues that remain dry on the tens of nanoseconds timescale of our simulations. It is conceivable that, on the timescale of the sample preparation for neutron experiments, *i.e.* several days, water is able to diffuse throughout the powder, breaking up dry hydrophilic clusters. An incomplete equilibration of the water distribution in our tau powder model could result in the observed underestimation of the average protein hydrogen atom dynamics. This discrepancy presumably does not occur in the MBP powder model because it is constructed with fully hydrated crystal structures of MBP.

Correlation between the tau protein dynamics and its hydration water

The tau protein MSDs presented in Figure 2a show a marked heterogeneity of dynamics among the different tau molecules at high temperatures. To gain more insight into this dynamical heterogeneity, we evaluated the correlation between hydration water and protein dynamics for each of the ten tau molecules in our model powder at 300 K. As

expected from Figure 2a, we found a quite broad distribution of protein hydrogen MSDs extracted directly from the trajectories among the different molecules (Figure 4a). For each tau molecule, we plotted the MSDs of its hydrogen atoms after 500 ps as a function of their hydration water hydrogen MSDs after 100 ps (Figure 4b) and as a function of the percentage of solvent-exposed hydrogens (Figure 4c). The hydration water hydrogens were defined as water hydrogens that are within 3 Å from any atoms of the considered tau molecule. To quantify the correlation between these quantities, we applied a linear fit and calculated the coefficient of determination R^2 (Figure 4b and 4c). We find a poor correlation between the solvent-exposure of a protein molecule and its dynamics ($R^2 = 0.238$, Figure 4c), while a better correlation is observed between the protein dynamics and its hydration water dynamics ($R^2 = 0.641$, Figure 4b). This result suggests that the observed discrepancy between simulated and experimental tau protein dynamics at high temperature (Figure 2a) originates predominantly in an underestimation of the solvent dynamics in the simulation. The latter is likely the result of a misrepresentation of the actual water distribution in the powder model due to the finite equilibration time, as discussed in the previous paragraph. In a previous study[50], we have shown that the water dynamics in our powder models reproduces neutron scattering data well up to 260 K, but the comparison could not be made between 260 K and 300 K because of the lack of experimental data.

While the conformation of a globular protein in solution and in a hydrated powder remains similar, an IDP is expected to undergo drastic conformational changes, from more extended conformations in solution to more collapsed structures in a powder. Indeed, the average radius of gyration of the tau protein in our powder model at 300 K is 26.5 Å, while the protein in solution has an average radius of gyration of about 62 Å[8]. Compared to a folded, globular protein there is an immense number of possible IDP conformations in a powder, rendering an accurate representation of an IDP powder challenging. The conformation of a protein affects the dynamics of its hydration water through the topology of the protein surface[51–54]. Here we show that, in turn, hydration water dynamics influences the dynamics of proteins. Thus, we propose that in hydrated protein powders the topology of the protein-water interface, which is particularly challenging to accurately represent in the case of IDP powders, is a key feature determining the protein dynamics.

Temperature-dependent dynamics of solvent-exposed and solvent-buried groups

In this section we analyze and discuss the dynamical behavior of solvent-exposed and buried protein hydrogens as a function of temperature, particularly across the DT. Methyl groups have been shown to have a quite distinct dynamical behavior from other chemical groups in proteins[27,55,56]. Because they appear mostly in hydrophobic side chains, methyl groups are preferentially sequestered from the aqueous solvent. Indeed, in the groups analyzed in Figure 3 (exposed and buried protein hydrogen atoms), the proportion of hydrogen atoms belonging to methyl groups was 26-29% and 18-19% in the buried hydrogen class, but only 3-4% and 4% in the exposed hydrogen class, for the MBP and tau powders, respectively. It is thus difficult to ascertain whether the observed differences in temperature-dependent dynamics between the two classes of hydrogen atoms originates in their solvent exposure or in their methyl group content. In Figure 5a, we plot the MSDs after 500 ps of the non-methyl hydrogen atoms that are either buried or exposed to water molecules, as a function of temperature. We kept the same definition as above, *i.e.*, buried hydrogens are defined as being at least 4 Å away from water atoms and exposed hydrogens are within 2.5 Å of any water atoms. In both protein powders, the MSDs of the exposed hydrogen atoms increase significantly above 200 K as compared to the buried atoms (Figure 5a). The two proteins show qualitatively the same trend, although the difference between exposed and buried atoms is smaller in the tau powder. We carried out the same analysis, presented in figure 5b, for buried and exposed backbone atoms (N, C, Cα, O). Similar to the hydrogen atoms, the MSDs of the solvent-buried backbone atoms are smaller than the solvent-exposed backbone atoms; again, in both cases the MSDs increase more rapidly with temperature above 200 K. Thus, the protein DT is more marked in the solvent-exposed parts of the proteins, as compared to the solvent-buried parts.

A less pronounced protein DT with increasing distance from the solvent was found by Tournier and colleagues in an MD study of myoglobin powder[26]. Our study on two other proteins of different nature suggests a generalization of the notion that an attenuated but not suppressed DT in the solvent-buried interior is a general feature of proteins. This finding is compatible with previous work from Hong *et al.* showing that hydration effects propagate through the whole protein, although they are weaker in the protein core[27]. Using neutron scattering and isotope labeling, Wood *et al.* probed the dynamics of the methyl group hydrogen atoms buried inside the protein calbindin, and compared to the methylene hydrogen

atoms of lysine residues present at the surface of the protein.[28] The dynamics was found not only to propagate through the protein core but also to be enhanced in buried methyl group hydrogen atoms compared to the exposed lysine methylene group hydrogens. To shed light on the apparent paradox between the experimental work of Wood *et al.* and the attenuated dynamics in the protein interior found in the previous section and elsewhere[26,27], we analyzed the dynamics of methyl groups in tau and MBP.

Dynamical behavior of methyl groups

In the previous section, we removed the contribution from methyl group hydrogen atoms to analyze the solvent effect on protein dynamics. Here we focus on the dynamics of methyl groups, as compared to the rest of the protein, as a function of temperature. The methyl group hydrogens represent about 20% of all hydrogen atoms in proteins and are preferentially found in solvent-buried parts. Figure 6 shows the MSDs after 500 ps computed for methyl hydrogen atoms and non-methyl hydrogen atoms. In both powders, the dynamics of methyl hydrogen atoms increase substantially around 150 K, and increase almost linearly with temperature above 200 K. Furthermore, the MSDs of methyl hydrogens are higher than those of other hydrogen atoms.

The apparent onset of protein anharmonic dynamics commonly observed in protein powders at low temperature has been attributed to methyl group rotations entering the time window of the probe[19,55–59]. When observing motions on roughly the nanosecond timescale, as herein, methyl group rotations were shown to appear at around 150 K[56]. In addition to their dynamical heterogeneity that has been demonstrated earlier[60,61], methyl group hydrogens are shown here to be, on average, significantly more dynamic than other protons of the protein (Figure 6), despite the fact that most methyl groups are buried from the solvent. Previous studies have suggested that the dynamics of methyl groups reflects the extent of their confinement[62], and can therefore be used to probe the packing of the protein core[63]. The methyl group dynamics in tau and MBP are similar (compare Figure 6a and 6b), thus revealing a similar degree of packing and confinement of the methyl groups in the two powder models.

Comparison of tau and MBP.

So far we have seen that the same qualitative dynamical behavior over a wide temperature range holds for both folded and disordered proteins. Here we focus on the differences between the two protein powder models. Figure 5a and 5b shows that both buried backbone and hydrogen atoms have the same dynamics in the two powders, while the dynamics of the exposed residues vary from one protein to the other.

The solvent-buried parts of the protein are highly different in chemical composition and structure between the two powder models. Indeed, in the MBP model, the solvent-buried atoms correspond to the hydrophobic core of the protein, which is naturally hidden from the solvent. In contrast, the buried atoms in the tau powder model result from the collapse of the hydrophilic disordered chain into dry clusters, due to limited hydration of the larger solvent accessible surface. Strikingly, despite these differences, the dynamical behavior on the ns timescale of solvent-buried hydrogen atoms (Figure 5a) and solvent-buried backbone atoms (Figure 5b) is very similar in both powder models. Thus, our results suggest that the temperature evolution of the protein dynamics in hydrated powders, especially across the DT, does not depend on the chemical properties (apart from methyl group composition) or on the protein structure, but rather on the features of the protein-water interface, such as its surface area or the water confinement at the interface.

Hydrogen bond network dynamics vs. temperature.

We quantified the dynamics of intra-protein and protein-water hydrogen bond (HB) network rearrangements by computing HB autocorrelation functions (ACF) in both powders. The time-dependent decay of these functions reflects changes in the HB network connectivity via breakage and formation of new HBs. Transient HB breaking events govern the initial fast decay of the ACF. However, many transiently broken HBs, e.g. through harmonic thermal fluctuations, are quickly reformed in the absence of larger rearrangements in the local HB network environment, i.e. if the involved HB donors and acceptors are not captured by other vacant hydrogen bonding partners. The latter process results in irreversible breaking of the HBs and is described by the long-term ACF decay. The corresponding ACFs of protein-water and intra-protein HBs are shown for the tau powder in Figure 7b and 7c, respectively. For the solid/glassy systems at temperatures below the DT, the HB ACF's decay to a non-zero plateau due to the absence of HB rearrangement processes. For a perfect solid, the plateau

value of the ACF determines the history independent probability of a HB to be intact, i.e. not being transiently broken. Below the DT this probability decreases linearly with increasing temperature due to the increase of harmonic thermal fluctuations. Upon approaching the DT temperature, HB network rearrangement processes start to occur, resulting in a long-term decay of the ACF instead of a constant plateau. The full relaxation times are still much larger than the timescales of our observations, especially for intra-protein HBs (Figure 7c). Therefore we report in Figure 7a the values of the intra-protein and protein-water HB ACF after 500 ps to quantify the rearrangement of intra-protein and protein-water HB networks, respectively, on this timescale. However, our conclusions are independent of this particular choice. In both protein powders, the ACF values at 500 ps for protein-water HBs decrease with a significantly steeper slope above temperatures of 200 K than at lower temperatures, indicating the onset of HB network relaxations. This is in good agreement with previous studies showing that the protein DT in the same temperature range correlates with the reorganization of the protein-water hydrogen bond network[50,64]. In contrast, the 500 ps ACF value of intra-protein HBs decreases almost linearly from 20 K to 300 K, however, with a slight break of the slope at 200 K. The difference in dynamics between the intra-protein and protein-water HBs is consistent with the picture drawn above, in which the protein DT is driven by the solvent and attenuated in the solvent-buried parts of the protein. Although the same trends are observed in both protein powder models, the difference between protein-water and intra-protein HB network dynamics is less pronounced for tau, as seen in Figure 7a, indicating a tighter dynamical coupling of hydration water with tau vs. MBP, as suggested earlier[8].

Conclusions

In the present work we have simulated a hydrated powder of the intrinsically disordered protein tau, and compared its dynamical behavior to the globular protein MBP that has about the same molecular weight. In contrast to globular proteins that possess roughly the same conformation in solution and powder, IDP powders are particularly challenging to model because of the large number of possible protein conformations. By comparing protein dynamics obtained from neutron scattering experiments and simulation, and by analyzing different parts of the proteins as well as the individual protein molecules present in the powder model, we found that the protein-water interface in an IDP powder is a key feature

allowing an accurate representation of protein dynamics. In particular we show that the dynamics of the solvent, determined by its confinement at the protein surface[53], is the predominant effect driving the protein dynamics. Thus, we propose that efforts to obtain realistic models of IDP powders should focus on optimizing the solvent-protein interface. Furthermore, we analyzed the dynamical behavior of the solvent-buried and exposed residues as a function of temperature across the so-called dynamical transition, which only occurs beyond a threshold hydration level[24]. We found that in both globular and disordered protein powders, the dynamical transition is attenuated, although not suppressed, in the solvent-buried parts of both proteins. The dynamics of hydrogen atoms belonging to methyl groups are found to increase significantly around 150 K, as shown before[19], and to be higher than for the other hydrogen atoms over the entire simulated temperature range (20-300 K). Finally, we compared the dynamical properties of the disordered protein and the globular protein as a function of temperature. Interestingly, we found that, despite major chemical and structural differences of the solvent-buried amino acids in the IDP and globular protein powders, their dynamics are very similar. In contrast, the dynamics of the solvent-exposed protein atoms are markedly different. Thus, we propose that the main feature determining the protein dynamics in hydrated protein powders is the solvent-protein interface, such as the number of exposed residues and the confinement of water molecules at the interface, rather than the primary, secondary or tertiary structure of the proteins.

The results presented herein show that the protein dynamics in a hydrated powder reflect the protein-water interface. Thus, we suggest that, by measuring the protein dynamics in a hydrated powder, *e.g. via* neutron scattering, one could obtain information on the protein hydration state. Such an experiment would be of particular interest to characterize protein materials whose intra- and supramolecular structures are unknown, such as protein aggregates. Indeed the latter are particularly challenging to characterize structurally, and whether they form big, dry clusters or partially hydrated, supramolecular assemblies remains an open debate[65–67]. Characterizing the dynamics of aggregates in the hydrated powder state on the nanosecond timescale may provide insight into features of the protein-water interface that would, in turn, provide some structural insights.

Acknowledgements
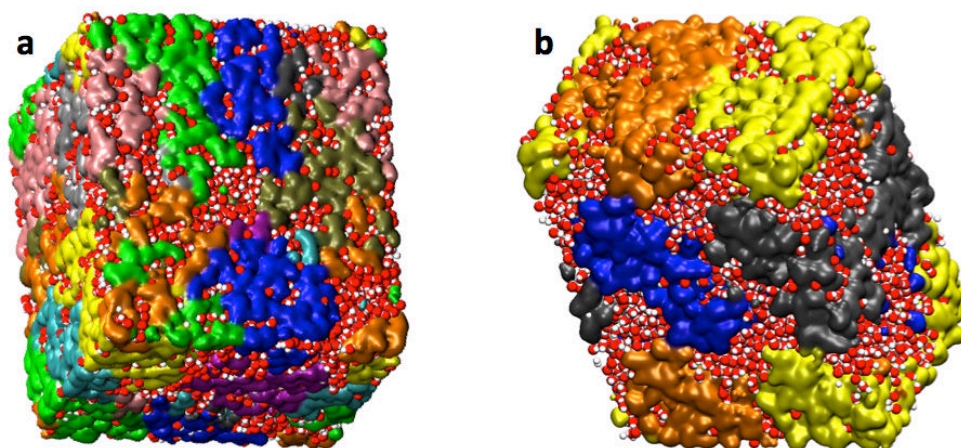
Figures

Figure 1: Snapshots of the simulation boxes of the tau (a) and MBP (b) powder models. The protein molecules are drawn in surface representation, with different colors for each individual molecule. Water molecules are drawn with van der Waals spheres, with O and H atoms represented in red and white, respectively.
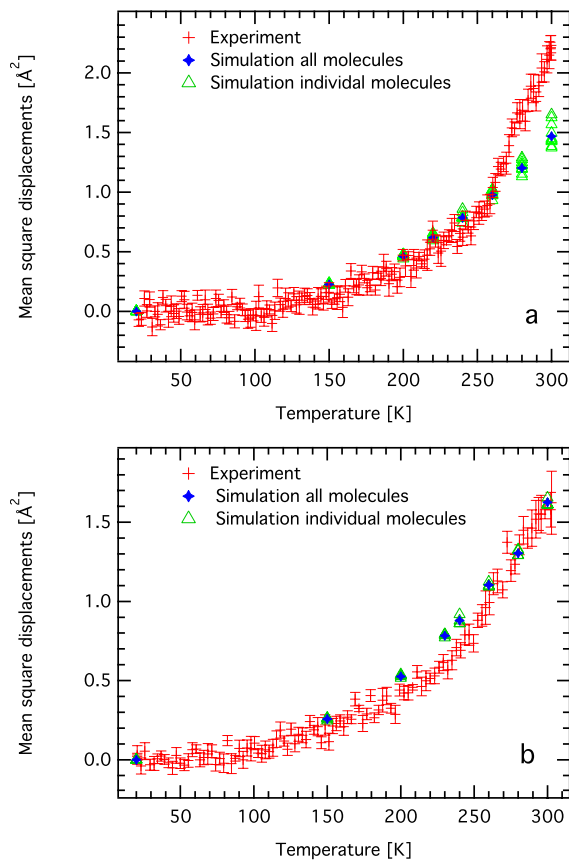
Figure 2: (a) Experimental hydrogen atom MSDs of the tau protein obtained by neutron scattering in the work of Gallat *et al.*[8] are shown as red crosses. Simulated hydrogen atom MSDs of each of the ten tau molecules composing the powder model are shown as green triangles, and their averages as blue stars. (b) Experimental hydrogen atom MSDs of MBP obtained by neutron scattering in the work of Wood *et al.*[30], and reprocessed as in Gallat *et al.*[8], are shown as red crosses. Simulated hydrogen atom MSDs of each of the four MBP molecules composing the powder model are shown as green triangles, and their averages as blue stars. The simulated and measured MSDs are extracted for both proteins from the elastic intensities $S(q,\omega=0)$ by applying the Gaussian approximation between $q^2 = 0.19$ and 1.34 Å^{-2}.
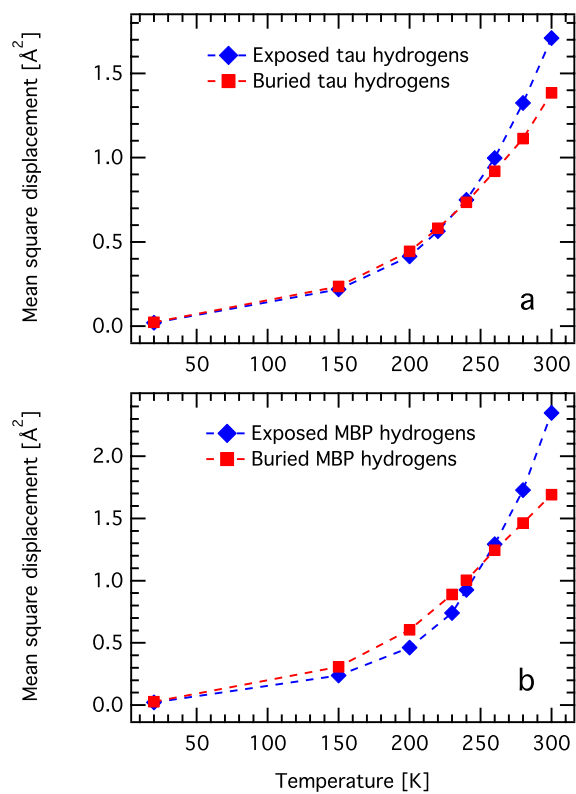
Figure 3: MSDs after 500 ps, as a function of temperature, of exposed and buried hydrogen atoms of (a) the tau protein and (b) the MBP.
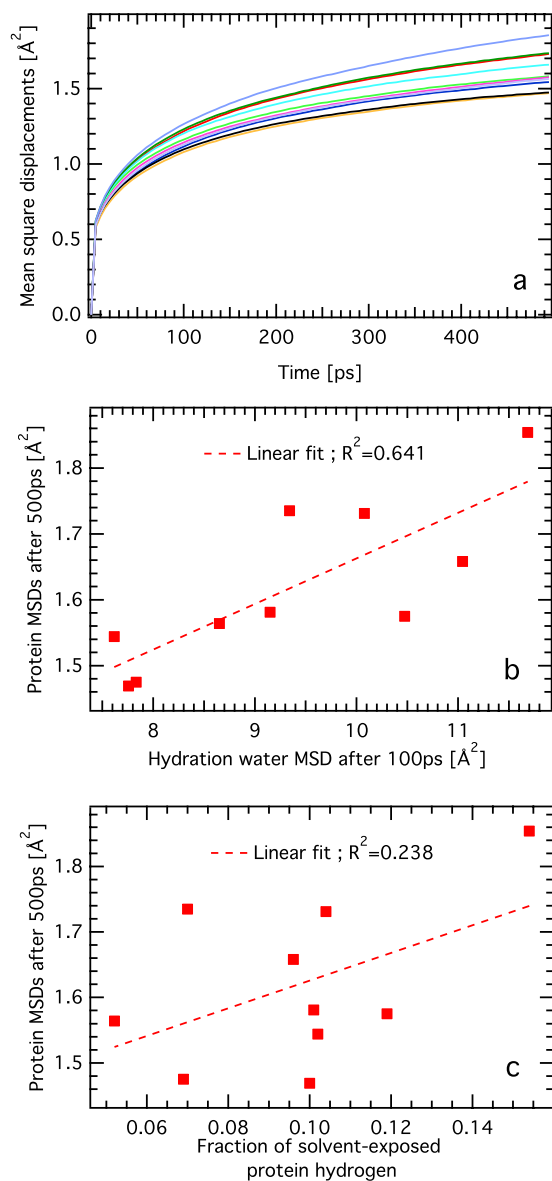
Figure 4: (a) MSD(*t*) of the hydrogen atoms of the ten individual tau molecules in the powder model at 300 K. MSDs after 500 ps of different tau molecules (b) as a function of their hydration water MSD after 100 ps and (c) as a function of the number of their exposed hydrogens. In (b) and (c) each point represents one tau molecule.

168
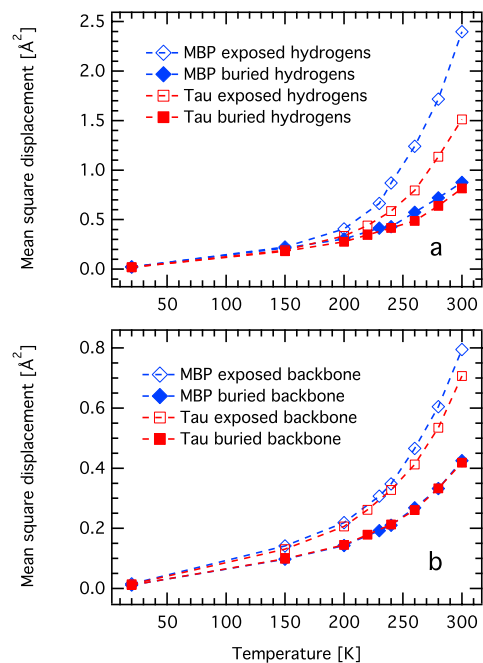
Figure 5: MSDs after 500 ps, as a function of temperature, of exposed and buried (a) non-methyl hydrogen atoms and (b) backbone atoms for both MBP and tau.
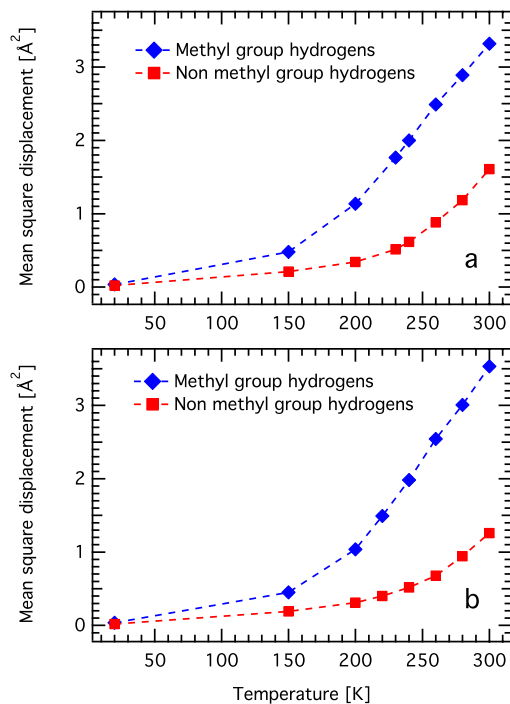
Figure 6: MSDs after 500 ps, as a function of temperature, of methyl and non-methyl hydrogen atoms of (a) the tau protein and (b) the MBP.
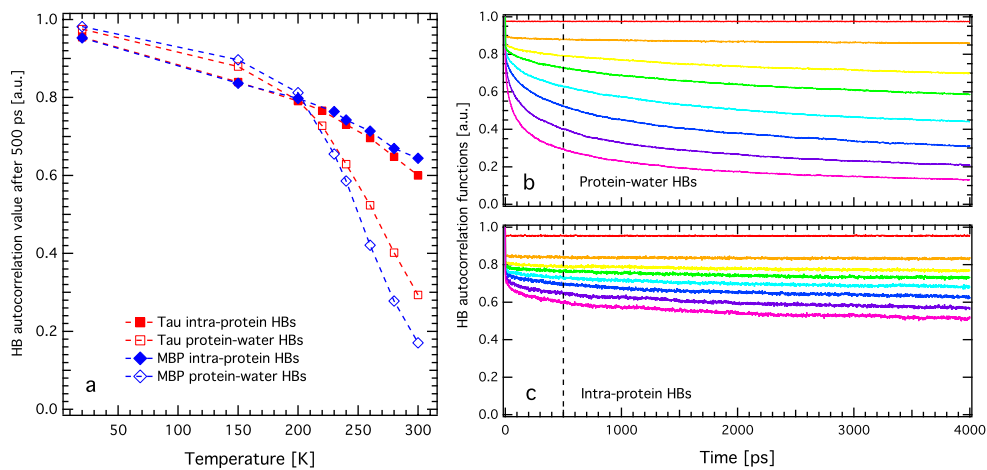
Figure 7: (a) Intra-protein and protein-water hydrogen bond autocorrelation function (ACF) values after 500 ps, as a function of temperature. ACFs of the protein-water (b) and intra-protein (c) HBs in the tau powder model at different temperatures (from top to bottom; red, 20 K; orange, 150 K; yellow, 200 K; green, 220 K; cyan, 240 K; blue, 260 K; purple, 280 K; magenta, 300 K). The vertical dashed line highlights 500 ps, the time at which the ACF values are reported in panel (a).

References

(1) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **2004**, *337* (3), 635–645.

(2) Tompa, P. Intrinsically Unstructured Proteins. *Trends Biochem. Sci.* **2002**, *27* (10), 527–533.

(3) Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. Unfoldomics of Human Diseases: Linking Protein Intrinsic Disorder with Diseases. *BMC Genomics* **2009**, *10* (Suppl 1), S7.

(4) Goedert, M.; Klug, A.; Crowther, R. A. Tau Protein, the Paired Helical Filament and Alzheimer's Disease. *J. Alzheimers Dis.* **2006**, *9* (0), 195–207.

(5) Stadler, A. M.; Stingaciu, L.; Radulescu, A.; Holderer, O.; Monkenbusch, M.; Biehl, R.; Richter, D. Internal Nanosecond Dynamics in the Intrinsically Disordered Myelin Basic Protein. *J. Am. Chem. Soc.* **2014**, *136* (19), 6987–6994.

(6) Perticaroli, S.; Nickels, J. D.; Ehlers, G.; Mamontov, E.; Sokolov, A. P. Dynamics and Rigidity in an Intrinsically Disordered Protein, B-Casein. *J. Phys. Chem. B* **2014**, *118* (26), 7317–7326.

(7) Gaspar, A. M.; Appavou, M.-S.; Busch, S.; Unruh, T.; Doster, W. Dynamics of Well-Folded and Natively Disordered Proteins in Solution: A Time-of-Flight Neutron Scattering Study. *Eur. Biophys. J.* **2008**, *37* (5), 573–582.

(8) Gallat, F.-X.; Laganowsky, A.; Wood, K.; Gabel, F.; van Eijck, L.; Wuttke, J.; Moulin, M.; Härtlein, M.; Eisenberg, D.; Colletier, J.-P.; et al. Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys. J.* **2012**, *103* (1), 129–136.

(9) Jensen, M. R.; Zweckstetter, M.; Huang, J.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014**, *114* (13), 6632–6660.

(10) Doster, W.; Cusack, S.; Petry, W. Dynamical Transition of Myoglobin Revealed by Inelastic Neutron Scattering. *Nature* **1989**, *337* (6209), 754–756.

(11) Ferrand, M.; Dianoux, A. J.; Petry, W.; Zaccaï, G. Thermal Motions and Function of Bacteriorhodopsin in Purple Membranes: Effects of Temperature and Hydration Studied by Neutron Scattering. *Proc. Natl. Acad. Sci.* **1993**, *90* (20), 9668–9672.

(12) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. Ligand Binding and Conformational Motions in Myoglobin. *Nature* **2000**, *404* (6774), 205–208.

(13) Rasmussen, B. F.; Stock, A. M.; Ringe, D.; Petsko, G. A. Crystalline Ribonuclease A Loses Function below the Dynamical Transition at 220 K. *Nature* **1992**, *357* (6377), 423–424.

(14) Pieper, J.; Hauss, T.; Buchsteiner, A.; Baczyński, K.; Adamiak, K.; Lechner, R. E.; Renger, G. Temperature- and Hydration-Dependent Protein Dynamics in Photosystem II of Green Plants Studied by Quasielastic Neutron Scattering†. *Biochemistry (Mosc.)* **2007**, *46* (40), 11398–11409.

(15) Khodadadi, S.; Pawlus, S.; Sokolov, A. P. Influence of Hydration on Protein Dynamics: Combining Dielectric and Neutron Scattering Spectroscopy Data. *J. Phys. Chem. B* **2008**, *112* (45), 14273–14280.

(16) Khodadadi, S.; Malkovskiy, A.; Kisliuk, A.; Sokolov, A. P. A Broad Glass Transition in Hydrated Proteins. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2010**, *1804* (1), 15–19.

(17)   Doster, W.; Busch, S.; Gaspar, A. M.; Appavou, M. S.; Wuttke, J.; Scheer, H. Dynamical Transition of Protein-Hydration Water. *Phys Rev Lett* **2010**, *104* (9), 098101.

(18)   Becker, T.; Hayward, J. A.; Finney, J. L.; Daniel, R. M.; Smith, J. C. Neutron Frequency Windows and the Protein Dynamical Transition. *Biophys. J.* **2004**, *87* (3), 1436–1444.

(19)   Schiro, G.; Natali, F.; Cupane, A. Physical Origin of Anharmonic Dynamics in Proteins: New Insights From Resolution-Dependent Neutron Scattering on Homomeric Polypeptides. *Phys. Rev. Lett.* **2012**, *109* (12).

(20)   Lagi, M.; Chu, X.; Kim, C.; Mallamace, F.; Baglioni, P.; Chen, S.-H. The Low-Temperature Dynamic Crossover Phenomenon in Protein Hydration Water: Simulations vs Experiments. *J. Phys. Chem. B* **2008**, *112* (6), 1571–1575.

(21)   Cupane, A.; Fomina, M.; Piazza, I.; Peters, J.; Schirò, G. Experimental Evidence for a Liquid-Liquid Crossover in Deeply Cooled Confined Water. *Phys. Rev. Lett.* **2014**, *113* (21), 215701.

(22)   Zanotti, J.-M.; Gibrat, G.; Bellissent-Funel, M.-C. Hydration Water Rotational Motion as a Source of Configurational Entropy Driving Protein Dynamics. Crossovers at 150 and 220 K. *Phys. Chem. Chem. Phys. PCCP* **2008**, *10* (32), 4865–4870.

(23)   Fomina, M.; Schirò, G.; Cupane, A. Hydration Dependence of Myoglobin Dynamics Studied with Elastic Neutron Scattering, Differential Scanning Calorimetry and Broadband Dielectric Spectroscopy. *Biophys. Chem.* **2014**, *185*, 25–31.

(24)   Roh, J. H.; Curtis, J. E.; Azzam, S.; Novikov, V. N.; Peral, I.; Chowdhuri, Z.; Gregory, R. B.; Sokolov, A. P. Influence of Hydration on the Dynamics of Lysozyme. *Biophys. J.* **2006**, *91* (7), 2573–2588.

(25)   Doster, W. The Protein-Solvent Glass Transition. *Incl. Spec. Sect. Protein-Water Interact.* **2010**, *1804* (1), 3–14.

(26)   Tournier, A. L.; Xu, J.; Smith, J. C. Translational Hydration Water Dynamics Drives the Protein Glass Transition. *Biophys. J.* **2003**, *85* (3), 1871–1875.

(27)   Hong, L.; Cheng, X.; Glass, D. C.; Smith, J. C. Surface Hydration Amplifies Single-Well Protein Atom Diffusion Propagating into the Macromolecular Core. *Phys. Rev. Lett.* **2012**, *108* (23), 238102.

(28)   Wood, K.; Gallat, F.-X.; Otten, R.; van Heel, A. J.; Lethier, M.; van Eijck, L.; Moulin, M.; Haertlein, M.; Weik, M.; Mulder, F. A. A. Protein Surface and Core Dynamics Show Concerted Hydration-Dependent Activation. *Angew. Chem. Int. Ed.* **2013**, *52* (2), 665–668.

(29)   Smith, J. C. Protein Dynamics: Comparison of Simulations with Inelastic Neutron Scattering Experiments. *Q. Rev. Biophys.* **1991**, *24* (03), 227–291.

(30)   Wood, K.; Frölich, A.; Paciaroni, A.; Moulin, M.; Härtlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water: Direct Measurements by Neutron Scattering and MD Simulations. *J. Am. Chem. Soc.* **2008**, *130* (14), 4586–4587.

(31)   Bizzarri, A. R. Neutron Scattering and Molecular Dynamics Simulation: A Conjugate Approach to Investigate the Dynamics of Electron Transfer Proteins. *J. Phys. Condens. Matter* **2004**, *16* (6), R83–R110.

(32)   Tarek, M.; Tobias, D. J. The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-Scattering Experiments. *Biophys. J.* **2000**, *79* (6), 3244–3257.

(33)   Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. Solvent Mobility and the Protein'glass' Transition. *Nat. Struct. Mol. Biol.* **2000**, *7* (1), 34–38.

(34)   Kumar, P.; Buldyrev, S. V.; Becker, S. R.; Poole, P. H.; Starr, F. W.; Stanley, H. E. Relation between the Widom Line and the Breakdown of the Stokes–Einstein Relation in Supercooled Water. *Proc. Natl. Acad. Sci.* **2007**, *104* (23), 9575–9579.

(35)   Hong, L.; Glass, D.; Nickels, J.; Perticaroli, S.; Yi, Z.; Tyagi, M.; O'Neill, H.; Zhang, Q.; Sokolov, A.; Smith, J. Elastic and Conformational Softness of a Globular Protein. *Phys. Rev. Lett.* **2013**, *110* (2).

(36)   Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and A-Synuclein in Solution from NMR and Small Angle Scattering. *Structure* **2014**, *22* (2), 238–249.

(37)   Mylonas, E.; Hascher, A.; Bernadó, P.; Blackledge, M.; Mandelkow, E.; Svergun, D. I. Domain Conformation of Tau Protein Studied by Solution Small-Angle X-Ray Scattering†. *Biochemistry (Mosc.)* **2008**, *47* (39), 10345–10353.

(38)   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J Comput Chem* **2005**, *26* (16), 1781–1802.

(39)   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998**, *102* (18), 3586–3616.

(40)   Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J Comput Chem* **2004**, *25* (11), 1400–1415.

(41)   Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.

(42)   Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177–4189.

(43)   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613–4621.

(44)   Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Mol. Simul.* **1991**, *6* (1-3), 121–142.

(45)   Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990–2001.

(46)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(47)   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* (19), 8577–8593.

(48)   Frick, B.; Gonzalez, M. Five Years Operation of the Second Generation Backscattering Spectrometer IN16—a Retrospective, Recent Developments and Plans. *Phys. B Condens. Matter* **2001**, *301* (1–2), 8–19.

(49)   Luzar, A.; Chandler, D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.* **1996**, *76* (6), 928–931.

(50) Schiro, G.; Fichou, Y.; Gallat, F.-X.; Wood, K.; Gabel, F.; Moulin, M.; Härtlein, M.; Heyden, M.; Colletier, J.-P.; Orecchini, A.; et al. Translational Diffusion of Hydration Water Correlates with Functional Motions in Folded and Intrinsically Disordered Proteins. *Nat. Commun.* **2015**, *6*, 6490.

(51) Luise, A.; Falconi, M.; Desideri, A. Molecular Dynamics Simulation of Solvated Azurin: Correlation between Surface Solvent Accessibility and Water Residence Times. *Proteins Struct. Funct. Bioinforma.* **2000**, *39* (1), 56–67.

(52) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *79* (6), 2966–2974.

(53) Sterpone, F.; Stirnemann, G.; Laage, D. Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *J. Am. Chem. Soc.* **2012**, *134* (9), 4116–4119.

(54) Bagchi, K.; Roy, S. Sensitivity of Water Dynamics to Biologically Significant Surfaces of Monomeric Insulin: Role of Topology and Electrostatic Interactions. *J. Phys. Chem. B* **2014**, *118* (14), 3805–3813.

(55) Krishnan, M.; Kurkal-Siebert, V.; Smith, J. C. Methyl Group Dynamics and the Onset of Anharmonicity in Myoglobin. *J. Phys. Chem. B* **2008**, *112* (17), 5522–5533.

(56) Wood, K.; Tobias, D. J.; Kessler, B.; Gabel, F.; Oesterhelt, D.; Mulder, F. A. A.; Zaccai, G.; Weik, M. The Low-Temperature Inflection Observed in Neutron Scattering Measurements of Proteins Is Due to Methyl Rotation: Direct Evidence Using Isotope Labeling and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2010**, *132* (14), 4990–4991.

(57) Roh, J.; Novikov, V.; Gregory, R.; Curtis, J.; Chowdhuri, Z.; Sokolov, A. Onsets of Anharmonicity in Protein Dynamics. *Phys. Rev. Lett.* **2005**, *95* (3).

(58) Schiro, G.; Caronna, C.; Natali, F.; Cupane, A. Direct Evidence of the Amino Acid Side Chain and Backbone Contributions to Protein Anharmonicity. *J. Am. Chem. Soc.* **2010**, *132* (4), 1371–1376.

(59) Schiro, G.; Caronna, C.; Natali, F.; Cupane, A. Molecular Origin and Hydration Dependence of Protein Anharmonicity: An Elastic Neutron Scattering Study. *Phys. Chem. Chem. Phys.* **2010**, *12* (35), 10215–10220.

(60) Lee, A. L.; Wand, A. J. Microscopic Origins of Entropy, Heat Capacity and the Glass Transition in Proteins. *Nature* **2001**, *411* (6836), 501–504.

(61) Curtis, J. E.; Tarek, M.; Tobias, D. J. Methyl Group Dynamics as a Probe of the Protein Dynamical Transition. *J. Am. Chem. Soc.* **2004**, *126* (49), 15928–15929.

(62) Best, R. B.; Clarke, J.; Karplus, M. The Origin of Protein Sidechain Order Parameter Distributions. *J. Am. Chem. Soc.* **2004**, *126* (25), 7734–7735.

(63) Best, R. B.; Rutherford, T. J.; Freund, S. M. V.; Clarke, J. Hydrophobic Core Fluidity of Homologous Protein Domains: Relation of Side-Chain Dynamics to Core Composition and Packing†. *Biochemistry (Mosc.)* **2004**, *43* (5), 1145–1155.

(64) Tarek, M.; Tobias, D. Role of Protein-Water Hydrogen Bond Dynamics in the Protein Dynamical Transition. *Phys. Rev. Lett.* **2002**, *88* (13).

(65) Thirumalai, D.; Reddy, G.; Straub, J. E. Role of Water in Protein Aggregation and Amyloid Polymorphism. *Acc. Chem. Res.* **2012**, *45* (1), 83–92.

(66) Petkova, A. T.; Yau, W.-M.; Tycko, R. Experimental Constraints on Quaternary Structure in Alzheimer's B-Amyloid Fibrils. *Biochemistry (Mosc.)* **2006**, *45* (2), 498–512.

(67) Paravastu, A. K.; Leapman, R. D.; Yau, W.-M.; Tycko, R. Molecular Structural Basis for Polymorphism in Alzheimer's B-Amyloid Fibrils. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18349–18354.

7.2 Summary

Summary   The tau protein is an intrinsically disordered protein (IDP) that regulates microtubule activity in neurons and has been largely studied because of its implication in the Alzheimer disease. IDPs are proteins that lack a unique structure and rather constantly exchange between multiple conformations. Thus, the high flexibility of IDPs has made the study of their dynamics particularly interesting. For this purpose, the combination of experimental and computational techniques, such as molecular dynamics simulations, is a powerful approach. The use of hydrated powder in several experimental techniques has been very useful to study, for instance, the protein internal dynamics. Thus, there is demand for realistic in silico models of hydrated protein powders. In this study we present the first dynamical analysis of a protein powder of the IDP tau in the temperature range 20-300 K. By comparing with neutron scattering data, we identified the protein-water interface as the predominant feature determining the IDP dynamics. The so-called protein dynamical transition was shown to be attenuated, however not suppressed, in the parts of the protein that are not exposed to the solvent. In addition, we found similarities in the dynamical properties of the core of a globular protein and dry clusters formed by the IDP, despite their significant chemical differences. Thus, the dynamical features of proteins in hydrated powders originates mainly from their solvent-exposed residues, and we proposed that by measuring the dynamics of protein assemblies, such as aggregates, one might assess their state of hydration.

Chapter 8

General discussion, perspectives and concluding remarks

We have presented different studies aiming at characterizing the dynamical properties of the tau protein and its hydration water.

Chapter 3 reports the existence of common dynamical features in the hydration water of globular proteins and IDPs. In particular, we showed that the nature of hydration water motions that correlate with the protein dynamical transition (PDT) is similar for both classes of proteins, as exemplified by the tau protein for IDPs and the MBP for globular proteins. That the PDT correlates with an onset of translational motions in the solvent was suggested in a computational study more than a decade ago (Tarek and Tobias, 2002a). However, it had neither been confirmed experimentally nor proposed for the class of IDPs.

Recently, the view of water as a necessary component for protein function has been challenged by a solvent-free myoglobin protein covered with a polymer surfactant corona, which was found to be biologically active (Perriman et al., 2010). Interestingly, an elastic neutron scattering study, which has been published and is presented in appendix A, showed that large-amplitude motions are present in the myoblobin above the PDT, notwithstanding the absence of water. It suggests that molecules other than water could plastisize proteins, allowing them to be active. A very interesting perspective is to carry out quasi-elastic neutron scattering experiments on this system, with the purpose of characterizing the degrees of freedom that activate in the polymer across the PDT. This experiment will reveal the nature of the polymer motions that correlate with the PDT, thereby potentially generalizing the results of chapter 3 to non-aqueous protein coatings.

We further investigated the water dynamics around IDPs and globular proteins by comparing the collective dynamics of their hydration water. Indeed, while incoherent neutron scattering experiments, such as those reported in chapter 3, probed individual water dynamics, THz spectroscopy provides insight into the collective fluctuations of water molecules. The

experiments carried out on the MBP and tau proteins, reported in chapter 4, revealed differences in the hydration water dynamics of the two proteins. In particular, we showed that the tau protein influences two times less water than the MBP, in terms of collective THz fluctuations.

Whether or not this difference has biological implications, as it is suggested in other studies (Conti Nibali and Havenith, 2014), is an essential question, although the presented results only allow for speculations. Further experiments would be required to investigate this matter. On experiment could be to follow the THz spectra of hydration water during protein biological activity, as it has been previously reported for a metalloprotease (Grossman et al., 2011; Dielmann-Gessner et al., 2014). For the tau protein, it means to follow the hydration water dynamics during the binding and unbinding to microtubules. A related perspective is to compare the collective hydration water dynamics of functional and non-functional proteins. For instance, such an experiment was previously performed on the native and denatured lambda repressor protein λ^*_{6-85}, and showed a decrease in water perturbation when the protein is inactive (Ebbinghaus et al., 2008). In that perspective, it would be highly interesting to follow the collective water dynamics as the tau protein fibrillates into its amyloid form. This study would present two main strong points: (i) The neutron scattering investigation reported in chapter 5 revealed a change in the individual dynamics of the hydration water of the tau protein when it is fibrillated. This result suggests that a strong effect on the collective dynamics might be observed (as discussed in chapter 4). (ii) The fibrillation into amyloid fibers is a spontaneous phenomenon that requires inter-protein interactions. THz fluctuations of hydration water have been proposed to promote protein binding, *via* a mechanism of hydration funnel (Conti Nibali and Havenith, 2014). Thus, the notion of collective water dynamics as a key actor in protein fibrillation is a credible hypothesis, which could drastically change our understanding of aggregation processes, and open new lines of investigation to mitigate protein aggregation. The proposed experiments could be performed using kinetic absorption spectroscopy (Kim et al., 2008), which has been recently developed and used to study hydration dynamics during protein folding or enzymatic reaction(Kim et al., 2008; Grossman et al., 2011; Dielmann-Gessner et al., 2014). This technique allows to record the THz spectra of a protein solution with a ms time resolution. At the moment, the main barrier to carry out such an experiment is that the tau fibers are insoluble, which creates some experimental difficulties.

In addition to a qualitative comparison between the two proteins, the THz spectroscopy results reported in chapter 4 suggest a long-range disturbance of water dynamics around the proteins, on the orders of tens of Å. Long-range influence of collective water collective dynamics around proteins has been found in several other studies including (Ebbinghaus et al., 2007, 2008; Meister et al., 2013; King et al., 2014). These findings might seem at first sight in contradiction with other results on the structure and single-particle dynamics of hydration water, which showed a perturbation limited to the first or second hydration layer (Merzel and Smith, 2002; Pal et al., 2002; Jasnin et al., 2008). In reality, there is no paradox between a short-range disruption of the individual dynamics and a long-range influence of the collective motions, because those are different quantities.

As a matter of fact, speculatively, if a long-range influence of biomolecules on their environment were to exist, it seems reasonable to conjecture that it would happen through coherent motions (*i.e.* related to collective dynamics), which can propagate, rather than incoherent motions (single-particle dynamics). A very naive analogy to illustrate this speculation could be the comparison of a heater, which dissipates energy in the form of incoherent thermal motions, with a speaker, which dissipates energy in the form of coherent sound waves. If both devices were supplied with the same power, let us say 100 watt, one could hear the sound several hundred meters from the speaker but couldn't feel the heat at such a long distance from the heater. Although this analogy is truly naive in many aspects, it provides an intuitive notion that a long-range disturbance of a medium such as water is more likely to involve coherent perturbations.

Given the molecular crowding inside cells, an influence of water dynamics extending beyond 10 Å from macromolecule surfaces implies that most of the water in cells and bacteria is dynamically distinct from bulk water. Yet, this logical deduction from THz spectroscopy experiments remains to be shown experimentally. Orecchini and coworkers (Orecchini et al., 2012) have performed a coherent neutron scattering experiment on fully deuterated *E. coli*, which allowed them to probe the collective dynamics of intracellular water in the THz range. No significant differences were found with respect to bulk water dynamics and the authors concluded that the water inside cells is essentially bulk like. The results obtained during this thesis do not allow to explain the discrepancy between THz spectroscopy measurements suggesting that all the water in cells is dynamically perturbed, and the neutron scattering study from (Orecchini et al., 2012) that did not observe such a perturbation. One possible explanation could be that the changes of collective dynamics in hydration water compared to bulk water are too subtle (indeed, they result in only a few percent difference of THz absorption) to be distinguished from the noise on the neutron spectrometer BRISP (used in (Orecchini et al., 2012)). Further investigations are required. In particular, it would be interesting to perform Brillouin neutron spectroscopy on the same samples that were analyzed by THz spectroscopy in chaper 4, *i.e.* MBP and tau solutions.

In chapters 5 and 6, we reported several experiments on the protein and hydration water dynamics of aggregates, on the ns time scale. The hydration water mobility of the tau protein was found to be enhanced around the amyloid fibers, which we postulated to play a role in the fibrillation process *via* entropy effects. However, entropy-enthalpy and entropy-entropy compensations are hard to directly assess. One way to gain insights into the involvement of single-particle hydration water mobility in the fiber formation is to follow it during the fibrillation process, as it has been suggested above for collective water dynamics. However, as we showed in chapter 5, the dynamics of water at the surface of protein aggregates is likely to be heterogenous and, therefore, its study requires a space-resolved probe. This time- and spatially-resolved experiment could be performed by Overhauser dynamic nuclear polarization enhanced NMR spectroscopy (Ortony et al., 2011) or femtosecond fluorescent spectroscopy (Zhong et al., 2002, 2011). The first technique assesses the solvent dynamics by measuring magnetic dipolar interactions between the ^1H spin of water and the electron spin of a nitroxide radical-based label. The sec-

ond technique can probe the solvent dynamics *via* the fluorescence lifetime of tryptophan residues on the ps time scale, which depends on their environment.

Protein aggregates were found to have diverse dynamics, with no obvious relation with their aggregation state (see table 6.2). In addition, the MD simulation study presented in chapter 7 identifies the protein-water interface as a key factor contributing to the overall protein dynamics in hydrated powders. Thus, the latter finding suggests that the dynamical diversity observed among the aggregates reflects different hydration states. In particular, the type of exposed residues and the local topology of the water-protein interface is likely to be different from one aggregate to another, resulting in different protein dynamics. Therefore, we propose a picture where the protein structure modification occurring during the formation of aggregates is not directly responsible for the change of protein dynamics, but rather indirectly, through the modification of the water-protein interface. An interesting perspective aiming to explore this hypothesis would be to measure by neutron scattering perdeuterated aggregates hydrated in H_2O. These experiments would allow to probe the hydration water dynamics, whose variation along different aggregate types could be compared to the variation of protein dynamics reported in this thesis. These experiments would be however quite costly, given the high price of deuterated materials and the large quantities of protein required for neutron scattering.

Our team has been involved in a recently-started project that aims at understanding the aggregation of the eye-lens crystallin proteins. Crystallins are structural proteins that are present at high concentrations (over 300 mg/ml) in the eye lens. They are responsible for the transparency and refractive properties required for proper lens function. Because of a very little protein turnover, the aggregation of crystallins is critical and is responsible for the opacification of the eye lens, characterizing the cataract. This large research project, led by Prof. Douglas Tobias at UC Irvine, proposes to combine advanced computational techniques (all-atom MD, brownian dynamics and Monte Carlo simulations) and experimental techniques (NMR spectroscopy, THz spectroscopy, small-angle scattering and neutron scattering), in order to understand the aggregation of γ-crystallin. By using several cataract-related mutants, this project aims at characterizing the inter-protein interactions and the hydration water structural and dynamical features that are critical in the aggregation process. In particular, small angle X-ray scattering (SAXS) is used to reveal the nature and intensity of inter-protein interactions, as well as the structure factors of crowded crystallin solutions, which are used to validate and refine the simulations. Promising preliminary SAXS results have been obtained during this thesis, of which more details are provided in appendix B.

In summary, we have shown that the dynamical properties of protein hydration water depend on the types of proteins as well as on their aggregation state. This thesis proposes that the hydration water dynamics of IDPs is a rich observable at a molecular level, which provides insight into biological mechanisms, such as association with binding partners and aggregation. This is showcased by the study of the intrinsically disordered protein tau, as compared to the globular MBP, and by the examination of several protein aggregates. Rather than a precise description of the mechanisms linking hydration water dynamics and protein functions, this work provides strong motivations for deeper investigations of the

biological role of hydration water.

Paper : A Polymer Surfactant Corona Dynamically Replaces Water in Solvent-Free Protein Liquids and Ensures Macromolecular Flexibility and Activity

A Polymer Surfactant Corona Dynamically Replaces Water in Solvent-Free Protein Liquids and Ensures Macromolecular Flexibility and Activity

François-Xavier Gallat,[†,‡,§,⊥,◆] Alex P. S. Brogan,[‖,◆] Yann Fichou,[†,‡,§] Nina McGrath,[‖] Martine Moulin,[⊥,#] Michael Härtlein,[⊥,#] Jérôme Combet,[⊥] Joachim Wuttke,[⊗] Stephen Mann,[‖] Giuseppe Zaccai,[†,‡,§,⊥] Colin J. Jackson,[▽] Adam W. Perriman,[*,‖] and Martin Weik[*,†,‡,§,×]

[†]Institut de Biologie Structurale, Comissariat à l'Energie Atomique, F-38054 Grenoble, France
[‡]CNRS, UMR5075, F-38027 Grenoble, France
[§]Université Joseph Fourier, F-38000 Grenoble, France
[⊥]Institut Laue-Langevin, 6 rue Jules Horowitz, B.P. 156, 38042 Grenoble Cedex 9, France
[‖]Center for Organized Matter Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, U.K.
[#]ILL-EMBL Deuteration Laboratory, Partnership for Structural Biology, 38042 Grenoble Cedex 9, France
[⊗]Forschungszentrum Jülich, JCNS at FRM II, Lichtenbergstrasse 1, 85747 Garching, Germany
[▽]Research School of Chemistry, Australian National University, ACT, 0200, Australia
[×]ESRF, 6 rue Jules Horowitz, BP 220, 38043 Grenoble Cedex, France

Ⓢ Supporting Information

ABSTRACT: The observation of biological activity in solvent-free protein−polymer surfactant hybrids challenges the view of aqueous and nonaqueous solvents being unique promoters of protein dynamics linked to function. Here, we combine elastic incoherent neutron scattering and specific deuterium labeling to separately study protein and polymer motions in solvent-free hybrids. Myoglobin motions within the hybrid are found to closely resemble those of a hydrated protein, and motions of the polymer surfactant coating are similar to those of the hydration water, leading to the conclusion that the polymer surfactant coating plasticizes protein structures in a way similar to hydration water.

A t the surface of soluble proteins, hydration water forms a mobile hydrogen-bonding network and permits sufficient macromolecular flexibility for proteins to attain full activity.[1] A loss of flexibility, either through reduction in temperature[2−4] or loss of hydration water,[5,6] is known to impair protein function[7,8] because protein motions necessary to access different conformational states[9] are suppressed. The coupling of hydration-water dynamics and protein dynamics is currently the subject of controversial discussions. In particular, the question of whether hydration-water dynamics enslave protein motions in a unilateral way,[10] merely act as a plasticizer that provides conformational freedom for the protein,[11] or couple differently with different classes of biological macromolecules[12,13] is currently a matter of considerable interest. It is known that water can in certain cases be substituted as a solvent by organic compounds[14] and that a large number of proteins function as part of cellular membranes where they are often at least partially surrounded by lipids rather than water. However,

all functional proteins have, until recently, been thought to require solvent for function. This view has been challenged by the creation of solvent-free protein−polymer surfactant nano-hybrids (hereafter referred to as "hybrids").[15,16] Specifically, a myoglobin hybrid has been shown to be able to reversibly bind oxygen.[16] However, the physical basis for the activity of the hybrid is unknown; if dynamics underpin protein function and solvent plays an essential role in the dynamical behavior of protein molecules, how can a solvent-free protein be bio-logically active? In this context, the present study addresses two specific questions: do the dynamics of the protein moiety of the hybrid resemble the dynamics of a hydrated protein, and do the dynamics of polymer surfactant coating resemble hydration-water dynamics? Answering these questions by studying this unique experimental system has the potential to shed light on the interplay between protein solvation, dynamics, and function.

Incoherent neutron scattering (INS) is widely used to study protein dynamics and probes atomic motions on the ns−ps time scales and Å length scales.[17] We applied elastic INS (EINS) to study the dynamic behavior of a solvent-free protein−polymer surfactant nanohybrid for the first time. By deuterating either the protein or the polymer moiety of the hybrid, we were able to separately characterize the temperature-dependent atomic mean square displacements (MSD) of the polymer and protein motions, respectively. Our results indicate that the dynamics of myoglobin within the hybrid closely resemble those of hydrated myoglobin and that the dynamics of the polymer corona are similar to those of typical protein hydration water, leading to the conclusion that the polymer

coating can plasticize protein structures in a similar way to hydration water.

The hybrids used in this work were synthesized by electrostatically grafting anionic poly(ethylene oxide)-based polymer surfactant chains to the surface of myoglobin that has been cationized (C-Mb) with *N,N*-dimethyl-1,3-propanediamine (DMPA) (Figure 1a)[18] before extraction of solvent
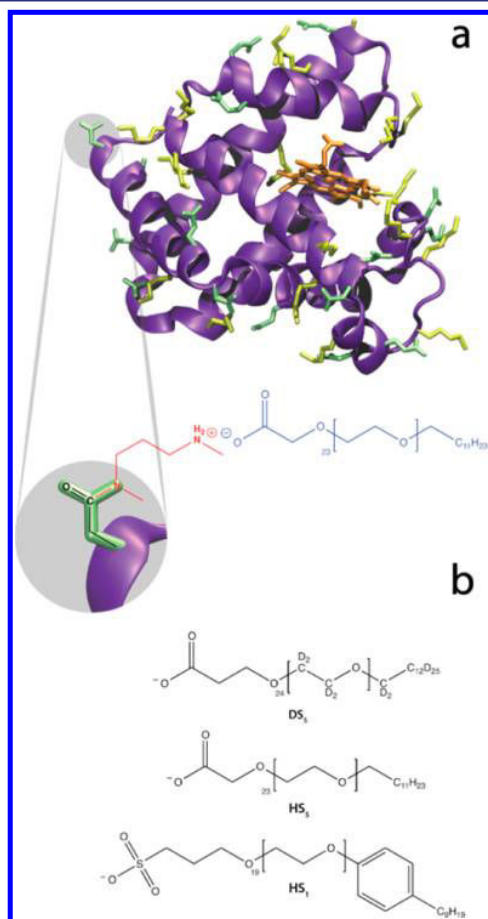

Figure 1. (a) Model illustrating the carbodiimide coupling of *N,N′*-dimethyl-1,3-propanediamine (DMPA; red) to an aspartic acid side chain (green) of myoglobin to yield cationized myoglobin (C-Mb), followed by electrostatic coupling with ethylene glycol ethoxylate lauryl ether (HS₅; blue) to give HMb/HS₅. After cationization there were ∼41 surfactant binding sites on C-Mb, including modified glutamic (green) and aspartic (green) acid as well as arginine (yellow) and lysine (yellow). (b) Anionic polymer surfactants: α-deuterated tridecanol propionic acid-terminated deuterated poly(ethylene glycol) (DS₅), ethylene glycol ethoxylate lauryl ether (HS₅), and poly-(ethylene glycol) 4-nonylphenyl 3-sulfopropyl ether (HS₁).

to a final hydration level (h) of 0.0035 g of H_2O per gram of myoglobin hybrid, which corresponded to ∼13 waters per hybrid molecule, or 4 waters per protein molecule by mass. This level of hydration is far below the several hundred water molecules needed to form a full hydration shell around small globular proteins such as myoglobin,[7] and is similar to that of a lyophilized protein powder after drying over P_2O_5 for several days.[19] As described previously, the protein moiety of these hybrids retains significant secondary structure, heme binding,

and oxygen binding/rebinding capacity[16] and can unfold and refold in response to changes in temperature.[20]

Hybrids of hydrogenated horse myoglobin or deuterated sperm whale myoglobin and the hydrogenated surfactants glycolic acid ethoxylate lauryl ether (HS₅) or poly(ethylene glycol) 4-nonylphenyl 3-sulfopropyl ether (HS₁) or α-deuterated tridecanol propionic acid-terminated deuterated poly(ethylene glycol) (DS₅) were synthesized (Figure 1b; see Materials and Methods in the Supporting Information). Expression and large-scale production of deuterated sperm whale myoglobin were carried out in the ILL-EMBL Deuteration Laboratory (Grenoble, France), and the synthesis of glycolic acid ethoxylate lauryl ether (HS₅) was done at the School of Chemistry, University of Bristol. The samples were lyophilized extensively and kept under nitrogen in a desiccator until required. Four different myoglobin/polymer hybrids were synthesized (Figure 1): (i) hydrogenated myoglobin/hydrogenated polymer HS₅ (HMb/HS₅), (ii) hydrogenated myoglobin/hydrogenated polymer HS₁ (HMb/HS₁), (iii) hydrogenated myoglobin/deuterated polymer DS₅ (HMb/DS₅), and (iv) deuterated myoglobin/hydrogenated polymer HS₅ (DMb/HS₅). The polymer surfactant occupancies of HMb/HS₅, HMb/HS₁, and DMb/HS₅ were approximately 100% (*i.e.*, approximately 41 polymer surfactant chains per C-Mb molecule), while HMb/DS₅ had 43% occupancy (18 polymer surfactant chains per C-Mb). Control samples of hydrogenated horse myoglobin powder, hydrated with D_2O at 0.43 g D_2O/g Mb (HMb/D_2O), and hydrogenated horse myoglobin, dried over P_2O_5 (HMb-dry), were also prepared.

Given that the polymer surfactant used in this study (HS₅) differed slightly from those described previously (HS₁[16] and HS₂[20]), the samples containing HS₅ were characterized to ensure they retained the same characteristics as the hybrids previously constructed. Synchrotron radiation circular dichroism (SRCD) experiments performed on HMb/HS₅ at 25 °C yielded spectra that are consistent with previous work on HMb/HS₂, confirming that, although there is a reduction in the level of α-helical structure (as compared to hydrated myoglobin), the secondary structure of the protein is still intact (Figure S1). Results from thermal denaturation SRCD experiments (Figure S2) performed on HMb/HS₅ are also largely identical to those of HMb/HS₂; *i.e.*, reversible unfolding of the protein is observed with a half-denaturation temperature ($T_{1/2}$) of 163 °C and a refolding efficiency of 95% from 155 °C (Figure S3). Finally, equilibrium gas binding experiments performed on HMb/HS₅ are consistent with previous work on HMb/HS₁,[16] showing reversible dioxygen binding with an oxygen affinity ($P_{1/2}$) of 3.4(0.2) torr with no cooperativity (Figure S4).

EINS experiments were carried out on the backscattering spectrometers IN16[21] (Institut Laue-Langevin, Grenoble, France) and SPHERES[22] (Jülich Centre for Neutron Science at FRMII, Garching, Germany). Incoherent neutron scattering from biological samples is dominated by the signal from hydrogen nuclei, whose incoherent scattering cross section is ∼40 times larger than those of deuterium and other atoms in the sample.[17] Consequently, for the HMb/HS₁ sample, 26% of the EINS signal is from the protein and 74% from the polymer surfactant HS₁. In the case of the HMb/HS₅ sample, 23% of the signal arises from the protein and 77% from the polymer surfactant HS₅. For the DMb/HS₅ sample, 11% of the signal is from protein and 89% is from the polymer surfactant. Thus, in these three samples, the signal is dominated by the polymer

contribution. In contrast, for HMb/DS$_5$, 91% of the signal is from the protein contribution and 9% from the polymer surfactant HS$_5$. Finally, in HMb/D$_2$O (HMb-dry), 98% (100%) of the signal is from protein and 2% (0%) from the D$_2$O hydration water (see Materials and Methods for detailed contributions to EINS signal).

The temperature dependence of the EINS was measured from 20 to 300 K for all samples, and atomic MSD (Figure 2) were obtained by linear regressions of the logarithm of the elastic intensity as a function of the scattering vector Q^2 (Figure SI Methods 10). Experiments on the fully hydrogenated hybrids HMb/HS$_5$ and HMb/HS$_1$ provide MSD averaged over both the Mb and the polymer moieties (Figure 2a). Up to ~120 K, MSD increase linearly due to harmonic motions. At ~120 K, the MSD start to increase at a higher rate, due to methyl rotations entering the experimental window of the neutron spectrometer.[11,24−26] A dynamical transition[3] is observed at temperatures ~250 K, marking the onset of anharmonic motions. A similar dynamical transition is also apparent in the HMb/D$_2$O sample (Figure 2a), which has been linked to the onset of biological activity in myoglobin.[8] We conclude that the dynamical behavior of the hybrid is similar to that of hydrated myoglobin and is largely independent of the nature of the polymer employed (HS$_1$[16] or HS$_5$).

The polymer surfactant occupancy of the HMb/DS$_5$ sample was reduced due to an unknown reason related to polymer deuteration. Nevertheless, the temperature dependence of its MSD qualitatively and quantitatively resembles the MSD of a hydrated protein, and is significantly different from those of a dry myoglobin powder (Figure 2b). In particular, the MSD go through a dynamical transition at ~250 K that is absent in the dry protein and are of a similar absolute value at 300 K to those of the HMb/D$_2$O sample (Figure 2b). Flash photolysis experiments have shown that a heme-bound ligand cannot leave the protein after photodissociation if myoglobin is either hydrated below $h = 0.15$,[6] embedded in a solid,[2] or below 230 K.[2] Thus, the anharmonic equilibrium fluctuations present in the hybrids above 250 K (Figure 2a,b) provide a physical explanation for the observed reversible oxygen binding activity in the hybrids (Figure S4). We suspect that the MSD of the protein moiety of a HMb/DS$_5$ sample with full polymer occupancy would be even closer to those of HMb/D$_2$O. However, we cannot exclude the possibility that the reduction in MSD of HMb/DS$_5$ with respect to HMb/D$_2$O is a result of the polymer surfactant molecules providing a somewhat less plasticizing environment compared to aqueous solvent.

Polymer surfactant motions in the hybrid were probed with the DMb/HS$_5$ sample. The MSD are almost identical to those of the fully hydrogenated hybrids and are qualitatively similar to the MSD of hydration water on the surface of deuterated maltose binding protein[23] (Figure 2c). Dynamics on the ns−ps time scale of the polymer surfactant coating and of hydration water are thus very similar, suggesting that the corona dynamically replaces water in these hybrids and creates a local environment in which proteins can be biologically active. On a microscopic scale, it remains unclear how the partially hydrophobic polymer surfactant interacts with the protein surface. In contrast to water molecules that hydrogen-bond with polar groups on the protein surface or cluster around apolar groups, the polymer surfactant can also engage in hydrophobic and electrostatic interactions. Electrostatic interactions *via* the DMPA linker might contribute to the dynamical coupling between the polymer surfactant and protein moieties
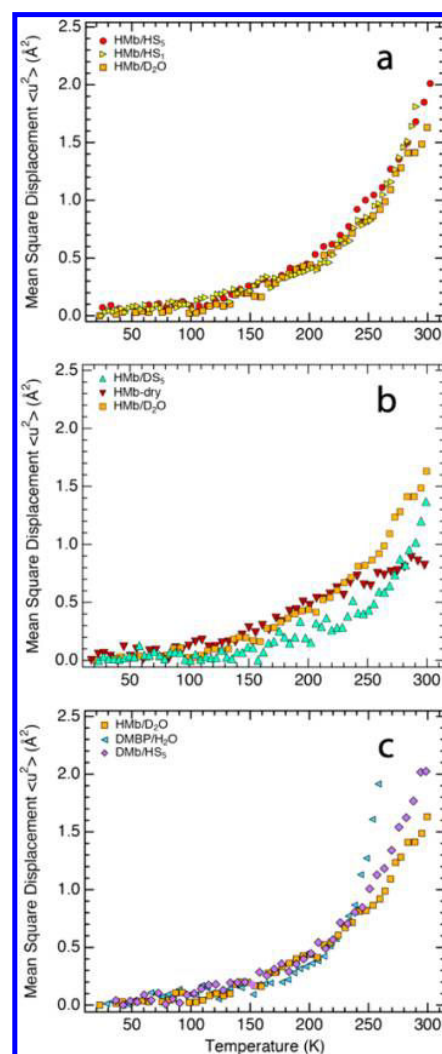

Figure 2. Atomic mean square displacements (MSD) as a function of temperature of solvent-free myoglobin−polymer surfactant hybrids and of their components. (a) Cationized, hydrogenated horse myoglobin complexed with the hydrogenated polymer surfactant HS$_1$ (HMb/HS$_1$; yellow triangles) or with the hydrogenated polymer surfactant HS$_5$ (HMb/HS$_5$; red circles), hydrogenated horse myoglobin in a powder hydrated at 0.43 g D$_2$O/g myoglobin[13] (HMb/D$_2$O; orange squares). (b) Cationized, hydrogenated horse myoglobin complexed with the deuterated polymer surfactant DS$_5$ (HMb/DS$_5$; green triangles), hydrogenated dry Mb powder (HMb-dry; red triangles), and HMb/D$_2$O[13] (orange squares). (c) Cationized, deuterated sperm whale myoglobin complexed with hydrogenated polymer surfactant HS$_5$ (DMb/HS$_5$; violet diamonds), hydration water in a powder of deuterated maltose binding protein (MBP) hydrated at 0.44 g H$_2$O/g MBP (blue triangles),[23] and HMb/D$_2$O[13] (orange squares). MSD were extracted in the range $0.20 < Q^2 < 1.40$ Å^{-2} for the HMb/HS$_5$, HMb/DS$_5$, and HMb/D$_2$O samples, in the range $0.18 < Q^2 < 1.33$ Å^{-2} for the HMb-dry sample, and in the range $0.20 < Q^2 < 1.13$ Å^{-2} for the DMb/HS$_5$ sample. Published neutron data of DMBP/H$_2$O[23] were re-examined for MSD extraction in the Q^2-range $0.19 < Q^2 < 1.13$ Å^{-2}. All data were collected on IN16[21] and SPHERES[22] and were processed in exactly the same way.

and might themselves modify the dynamics of the protein. The decreased number of hydrogen-bonding donors and acceptors in the polymer surfactant corona within a hybrid as compared

to those in the first water layer in fully hydrated proteins might lead to more electrostatic and polar interactions among surface side chains. As a result, the number of hydrogen-bonding partners per amino acid side chain might be similar in polymer surfactant-coated and hydration-water-covered proteins. Molecular dynamics simulations of a solvent-free protein−polymer surfactant hybrid might shed some light on microscopic details of polymer surfactant−protein interactions.

The observed similarities between the MSD of the polymer surfactant coating in these hybrids and protein hydration water highlight the importance of flexibility in the molecular matrix encapsulating biologically active proteins. We propose that this similarity is the basis of ligand-binding activity in solvent-free liquid myoglobins. Our results indicate that electrostatically linking polymer surfactant molecules to the surface of myoglobin produces a novel class of solvent-free biomolecular liquids in which the attached coronal layer functionally and dynamically substitutes for a hydration shell of water molecules and ensures macromolecular flexibility and conformational dynamics. The ability to replace the solvation shell of proteins without impairment of molecular dynamics could have considerable potential for the design and development of new functional biohybrid nanomaterials for use in anhydrous environments and other technologically challenging milieux.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Mb deuteration, purification, and characterization; synthesis and characterization of anionic polymer surfactants; Mb functionalization and polymer grafting; neutron scattering experiments; SRCD spectroscopy and thermal denaturation thermodynamics; oxygen binding assays. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

weik@ibs.fr; chawp@bristol.ac.uk

Author Contributions

◆F.-X.G. and A.P.S.B. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Ball, P. *Chem. Rev.* **2008**, *108*, 74.
(2) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355.
(3) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754.
(4) Zaccai, G. *Science* **2000**, *288*, 1604.
(5) Lehnert, U.; Reat, V.; Weik, M.; Zaccai, G.; Pfister, C. *Biophys. J.* **1998**, *75*, 1945.
(6) Kleinert, T.; Doster, W.; Post, F.; Settles, M. Hydration Effects on Protein Function: The Kinetics of Ligand Binding to Myoglobin. In *Proceedings of the Italian Physical Society*; Palma, M. U., Palma-Vittorelli, M. B., Parak, F., Eds.; Italian Physical Society: Bologna, 1993; Vol. 43, p 127.
(7) Rupley, J. A.; Careri, G. In *Advances in Protein Chemistry*; Anfinsen, C. B., Richards, F. M., Edsall, J. T., Eisenberg, D. S., Eds.; Academic Press: New York, 1991; Vol. 41, p 37.
(8) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205.
(9) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.
(10) Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroe, I. R.; Swenson, J.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129.
(11) Doster, W.; Settles, M. *Biochim. Biophys. Acta* **2005**, *1749*, 173.
(12) Khodadadi, S.; Roh, J. H.; Kisliuk, A.; Mamontov, E.; Tyagi, M.; Woodson, S. A.; Briber, R. M.; Sokolov, A. P. *Biophys. J.* **2010**, *98*, 1321.
(13) Gallat, F.-X.; Laganowski, A.; Wood, K.; Gabel, F.; van Eijck, L.; Wuttke, J.; Moulin, M.; Härtlein, M.; Eisenberg, D.; Colletier, J.-P.; Zaccai, G.; Weik, M. *Biophys. J.* **2012**, *103*, 129.
(14) Klibanov, A. M. *Nature* **2001**, *409*, 241.
(15) Perriman, A. W.; Colfen, H.; Hughes, R. W.; Barrie, C. L.; Mann, S. *Angew. Chem., Int. Ed.* **2009**, *48*, 6242.
(16) Perriman, A. W.; Brogan, A. P. S.; Colfen, H.; Tsoureas, N.; Owen, G. R.; Mann, S. *Nat. Chem.* **2010**, *2*, 622.
(17) Gabel, F.; Bicout, D.; Lehnert, U.; Tehei, M.; Weik, M.; Zaccai, G. *Q. Rev. Biophys.* **2002**, *35*, 327.
(18) Perriman, A. W.; Mann, S. *ACS Nano* **2011**, *5*, 6085.
(19) Dolman, M.; Halling, P. J.; Moore, B. D.; Waldron, S. *Biopolymers* **1997**, *41*, 313.
(20) Brogan, A. P. S.; Siligardi, G.; Hussain, H.; Perriman, A. W.; Mann, S. *Chem. Sci.* **2012**, *3*, 1839.
(21) Frick, B.; Gonzalez, M. *Physica B: Condensed Matter* **2001**, *301*, 8.
(22) Wuttke, J.; Budwig, A.; Drochner, M.; Kämmerling, H.; Kayser, F.-J.; Kleines, H.; Ossovyi, V.; Pardo, L. C.; Prager, M.; Schneider, G. J.; Schneider, H.; Staringer, S.; Richter, D. *Rev. Sci. Instrum.* **2012**, in press.
(23) Wood, K.; Frolich, A.; Paciaroni, A.; Moulin, M.; Hartlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. *J. Am. Chem. Soc.* **2008**, *130*, 4586.
(24) Roh, J. H.; Novikov, V. N.; Gregory, R. B.; Curtis, J. E.; Chowdhuri, Z.; Sokolov, A. P. *Phys. Rev. Lett.* **2005**, *95*, 038101.
(25) Schiro, G.; Caronna, C.; Natali, F.; Cupane, A. *J. Am. Chem. Soc.* **2010**, *132*, 1371.
(26) Wood, K.; Tobias, D. J.; Kessler, B.; Gabel, F.; Oesterhelt, D.; Mulder, F. A.; Zaccai, G.; Weik, M. *J. Am. Chem. Soc.* **2010**, *132*, 4990.

Appendix B

SAXS experiments on γ-S crystallin

Crystallins are structural proteins that are present at high concentrations (over 300 mg/ml) in the eye lens. They are responsible for the transparency and refractive properties required for proper lens function. Because of a very little protein turnover, the aggregation of crystallins is critical and is responsible for the opacification of the eye lens, characterizing the cataract. A recently-started research project, led by Prof.Douglas Tobias at UC Irvine, proposes to combine advanced computational techniques (all-atom MD, brownian dynamics and Monte Carlo simulations) and experimental techniques (NMR spectroscopy, THz spectroscopy, small-angle scattering and neutron scattering), in order to understand the aggregation of γ-crystallins. By using several cataract-related mutants, this project aims at characterizing the inter-protein interactions and the hydration water structural and dynamical features that are critical in the aggregation process. In the framework of this project, I have applied small angle X-ray scattering (SAXS) on the mutant G18V and wild type (WT) γ-crystallin, in order to measure the nature of inter-protein interactions, as well as the structure factors of crowded crystallin solutions. The latter is used to refine and validate the Monte Carlo simulations.

The experiments have been carried out on the beamline BM29 at the European Synchrotron Radiation Facility (ESRF), in Grenoble. The first step was to find the proper experimental conditions (buffer, temperature, exposure time, sample flowing, etc.) allowing to record a correct form factor (FF) at low concentration. Both WT and G18V were found to be very sensitive to radiation damage. By flowing freshly-prepared samples through the beam and by reducing the beam intensity, the measured WT FF was found to be in good agreement with the theoretical FF, which was calculated from the NMR structure (PDB entry 2M3T), using the software CRYSOL (figure B.1).

Concentrated solutions of WT γ-crystallin were then measured up to about 200 mg/ml. The scattering curves from the protein solutions at different concentrations are shown in figure B.2. From the intensities extrapolated to S=0 (note that S is the scattering vector, often denoted Q), at different concentrations, one can obtain the second viral coefficient.
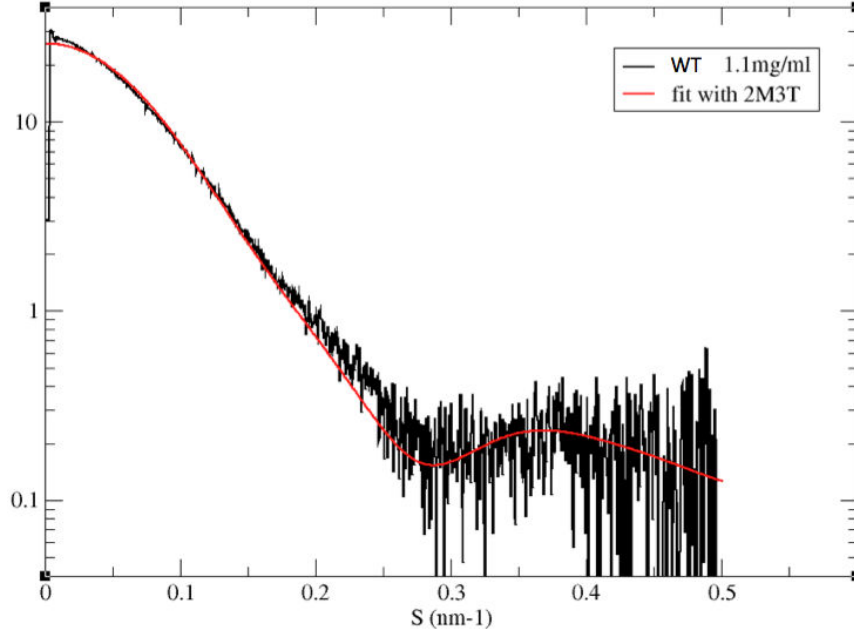
Figure B.1: Scattering curve from WT γ-crystallin measured at 1.1 mg/ml (black curve), compared with the theoretical scattering curve of the NMR structure 2M3T (red curve). The agreement is reasonable, although not perfect.

In particular, we used the following formula to fit the data :

$$\frac{I(S = 0, c)}{c} = I(0,0) \times \frac{1}{1 + 2cA_2M + 3c^2A_3N_a + ...} \quad , \tag{B.1}$$

where A_i is the ith virial coefficient, also commonly expressed as $B_i = A_iM^2/N_a$, I(S=0,c) is the scattered intensity at S=0 from a protein solution at a concentration c. Two sets of measurements provided A_2 values of 2 $\times 10^{-5}$ and 6 $\times 10^{-5}$ $cm^3.mol.g^{-2}$. Although the values are quite different (probably due to concentration determination errors, which have been corrected since then), they are all positive, indicating the existence of repulsive interactions between the protein molecules.

By dividing the scattering curves by the form factor (which is the scattering intensity of a low-concentration solution), one can obtain the structure factor of the solution at a given concentration. Preliminary structure factors, which are already being used to optimize simulation models, are shown in figure B.3.

The cataract-related mutant G18V is by definition prone to aggregation, and we haven't been able so far to measure good data on G18V solutions. Now that the experimental procedures have been optimized, the next stage is to acquire better quality data on WT and G18V, which will be used as references to optimize Monte Carlo simulations. The perspective of this work is to gather data on several cataract-related mutants, with the
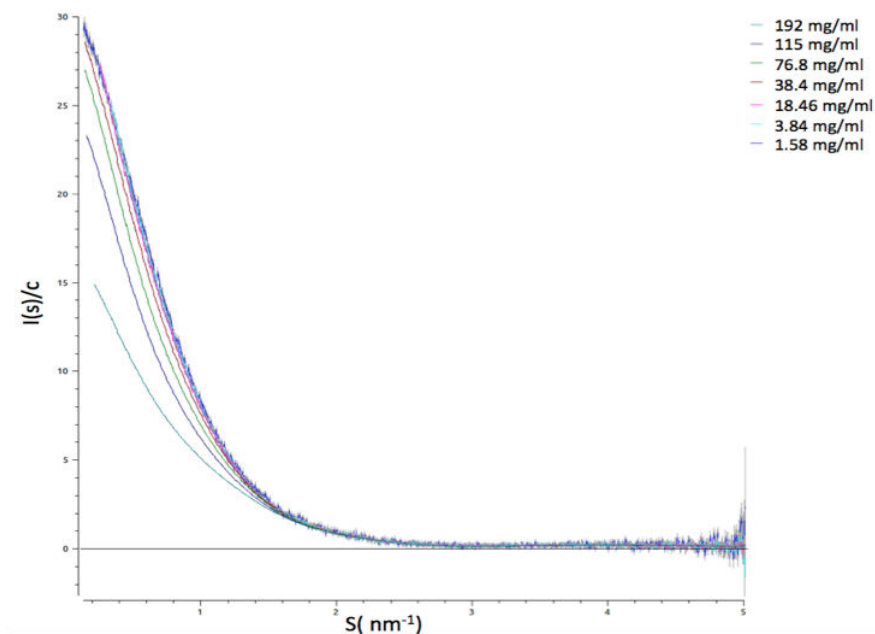
Figure B.2: Scattering curves at different concentrations of WT γ-crystallin.

purpose of comparing their inter-particle interactions and the structure factors. Another aspect of this project is to measure the single-particle and collective dynamics of the hydration water of different γ-crystallin mutants, by neutron scattering and THz spectroscopy, respectively. The long term goal is to reveal whether or not the aggregation propensity of γ-crystallins can be related to their interactions with the surrounding water.
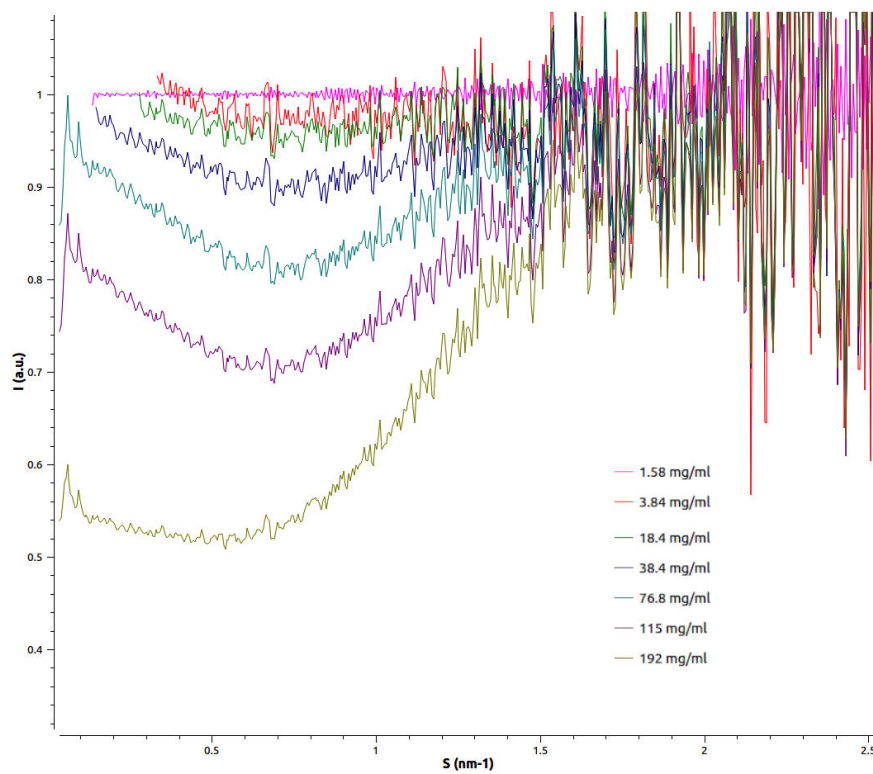
Figure B.3: Structure factors of WT solutions, obtained by dividing the scattering curve at a given concentration by the scattering curve at the lowest concentration (1.6 mg/ml).

Appendix C

Résumé de la thèse en français

L'eau, plus qu'un objet d'étude scientifique, est depuis la nuit des temps une préoccupation centrale de toutes civilisations. Dans son livre *H_2O : une biographie de l'eau*, Philip Ball nous transporte à travers les âges et nous raconte ce qui est connu, ou inconnu, des caractéristiques de cette substance. Au temps de l'Égypte antique, l'eau, en plus de permettre de se nourrir et de s'abreuver, avait un pouvoir de guérison ; les corps malades étaient ainsi plongés dans des baignoires d'eau sacrée. Nous retrouvons également cette substance parmi les quatre éléments fondamentaux qui, durant plus de deux millénaires, ont été associés à la création de tous matériaux constituant le monde. Il n'est donc pas étonnant que l'eau ait souvent été entourée de bien des fantasmes et légendes. Encore aujourd'hui, l'eau à une forte valeur symbolique, si bien que même la communauté scientifique à toutes les difficultés à en débattre objectivement, comme en témoignent par exemple les affaires de la mémoire de l'eau ou de l'eau polymérisée. Au XVIIIème siècle, le chimiste Antoine Laurent de Lavoisier fut l'un des pionniers dans l'étude de l'eau, et démontra que celle-ci est un composé de deux éléments : l'oxygène et l'hydrogène. L'eau n'était donc plus un élément fondamental. Depuis, l'eau a montré des propriétés physico-chimiques étonnantes qui fait d'elle une substance unique. Son rôle dans un contexte biologique est, comme présenté dans l'introduction, essentiel et seulement partiellement connu. Ce travail de thèse contribuera, je l'espère, à une meilleure compréhension de ce domaine.

Les travaux présentés dans ce manuscrit ont principalement été réalisés à l'Institut de Biologie Structurale (IBS), sous la direction du Docteur Martin Weik, et sont le fruit d'une intense collaboration avec les groupes du Professeur Douglas Tobias à l'université de Californie à Irvine (Etats-Unis) et du Professeur Martina Havenith à l'université de la Ruhr (Allemagne). Cet appendice suit le plan du manuscrit et résume les résultats essentiels de chaque chapitre. Pour plus de détails, il convient de se référer aux chapitres complets, rédigés en anglais.

C.1 Introduction

L'eau est une composante essentielle à tous les systèmes biologiques connus aujourd'hui et elle est pour cette raison parfois appelée la *matrice de la vie*. Au niveau moléculaire, l'eau joue un rôle clef dans une large variété de processus, tels que la liaison de ligands, la stabilité de l'ADN et des protéines, le transfert d'électrons et de protons, etc. Eau et macromolécules sont en étroite collaboration et en permanence sous l'influence l'un de l'autre. D'une part, l'eau à proximité des macromolécules, appelée eau d'hydratation, est perturbée par ces dernières et possède des propriétés dynamiques et structurales distinctes de l'eau libre (le terme d'eau libre se réfère à de l'eau pure qui ne connaît aucune influence de la part d'autres composés). D'autre part, la présence d'eau autour des protéines leur confère les propriétés dynamiques et structurales indispensables à leur bon fonctionnement. En particulier, les protéines hydratées connaissent une augmentation de leur dynamique autour de 240 K, augmentation appelée transition dynamique, et qui n'est pas observée en l'absence d'eau. Cette transition correspond à l'activation de mouvements de larges amplitudes nécessaires à la fonction de la protéine. Dans le protéome, les protéines intrinsèquement désordonnées (IDPs) forment une classe de protéines récemment découverte ne possédant pas de structure tridimensionnelle unique et clairement définie. Cette diversité de structures possibles pour une protéine rend leur dynamique et leurs interactions avec l'eau différentes des protéines dites globulaires (structure 3D unique et définie). Les IDPs ont attiré une attention toute particulière du fait que leur agrégation s'avère impliquée dans une large variété de maladies, y compris les maladies neurodégénératives. La protéine tau, qui a fait l'objet de cette thèse, est une IDP qui régule l'activité des microtubules dans les neurones. Tau est particulièrement étudiée car sa fibrillation (agrégation en forme de fibres) en fibres de type amyloïde est l'une des marques caractéristiques de la maladie d'Alzheimer.

Ce projet de thèse vise à établir si les propriétés dynamiques de l'eau présentent une pertinence biologique et si elles pourraient être utilisées pour mieux comprendre et pour moduler certains processus biologiques. Dans ce but, nous avons utilisé la diffusion incohérente de neutrons, la spectroscopie térahertz (THz), la diffusion de rayons X aux petits angles, et des simulations de dynamique moléculaire (MD), pour étudier la dynamique de l'eau d'hydratation autour de la protéine désordonnée tau, comparée à une protéine structurée, ainsi qu'autour de protéines fonctionnelles et agrégées.

En particulier, après avoir introduit les techniques utilisées dans le **chapitre 2**, ce manuscrit propose de répondre aux questions suivantes :

Chapitre 3 Quelles sont les caractéristiques dynamiques de l'eau d'hydratation qui permettent aux protéines de fonctionner? Cette question a été abordée en combinant diffusion quasi-élastique de neutrons et simulations MD pour étudier l'eau d'hydratation de l'IDP tau et d'une protéine globulaire, la *maltose binding protein* (MBP).

Chapitre 4 Est-ce que la dynamique collective de l'eau d'hydratation diffère selon la nature de la protéine? Ce point a été abordé en utilisant la spectroscopie THz sur les protéines tau et MBP.

Chapitre 5 Quels changements sont observés dans la dynamique d'eau d'hydratation lorsqu'une protéine est à l'état de fibres amyloïdes? Cette problématique a été examinée en étudiant la protéine tau sous formes native et fibrillaire, par diffusion de neutrons et simulations MD.

Chapitre 6 Les agrégats de protéines possèdent-ils une signature dynamique, et si oui, est-elle reliée à l'interaction eau-protéine? Cette question a été explorée par l'étude en diffusion élastique de neutrons de plusieurs protéines dans différents états d'agrégation.

Chapitre 7 Peut-on utiliser la simulation MD pour modéliser une poudre de protéine intrinsèquement dépliée, et que peut-on apprendre de ce modèle? Ce problème a été examiné en simulant une poudre hydratée des protéines tau et MBP.

.

C.2 Méthodes biophysiques

Les travaux présentés dans ce manuscrit sont essentiellement basés sur l'utilisation des techniques énumérées ci dessous.

- **La diffusion incohérente (quasi-)élastique de neutrons**, qui permet de quantifier la dynamique des atomes d'hydrogène présents dans un échantillon, à l'échelle de la pico-nanoseconde.

- **La diffusion aux petits angles**, qui permet de mesurer une forme globale de particules en solution ainsi que les interactions inter-particules.

- **La spectroscopie THz**, qui donne la possibilité de mesurer la dynamique collective de l'eau d'hydratation.

- **Les simulations de dynamique moléculaire**, qui permettent de simuler à l'échelle atomique l'évolution temporelle d'un système, comme par exemple une protéine entourée de molécules d'eau, en intégrant les équations de mécanique classique.

Les équations sous-jacentes à l'utilisation de ces méthodes sont explicitées dans le chapitre 2. En figure C.1 est présentée un diagramme temporel sur lequel sont placées différentes techniques biophysiques, ainsi que certains processus biologiques.
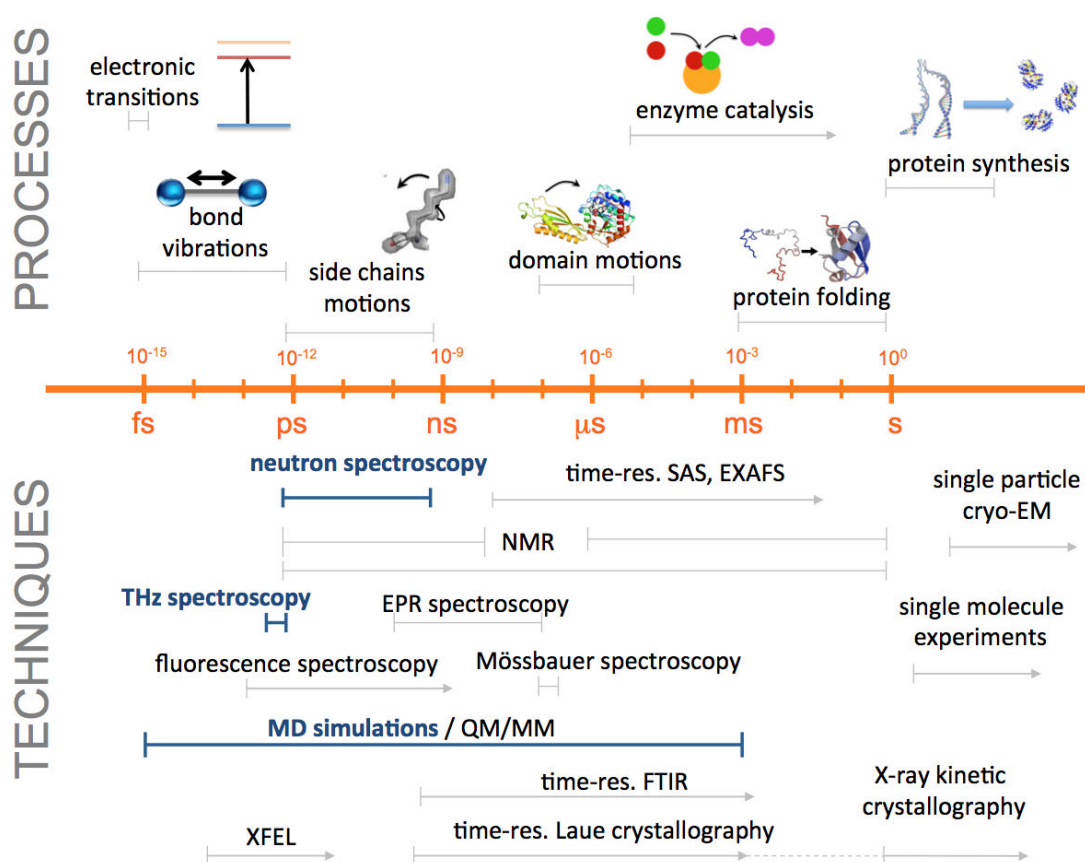
Figure C.1: Différents processus biologiques et techniques biophysiques placés sur une échelle temporelle. Les techniques majoritairement utilisées durant cette thèse sont indiquées en bleu.

C.3 Résumé des travaux effectués

C.3.1 Nature des mouvements de l'eau d'hydratation impliqués dans la transition dynamique de la protéine

L'eau est la matrice naturelle des macromolécules biologiques et est essentielle pour l'activité moléculaire des cellules. Malgré plus d'une décennie d'études sur le couplage entre dynamique de l'eau et dynamique des protéines, celui-ci reste toujours controversé. Dans ce travail, nous avons combiné deuteration de protéines, diffusion quasi-élastique de neutrons, et simulations de dynamique moléculaire, pour examiner la nature des mouvements de l'eau d'hydratation des protéines tau et MBP. Cette étude a été menée dans une gamme de température allant de 200 à 300 K, couvrant ainsi la transition dynamique attendue dans la protéine autour de 240 K. Le signal quasi-élastique a été « fitté » avec un modèle où les molécules d'eau subissent soit un mouvement de translation, soit de rotation, ou restent

immobiles. Les résultats des expériences et des simulations s'accordent pour montrer une augmentation significative de la composante translationnelle de l'eau à la température de la transition dynamique (240 K), indépendamment du type de protéines (voir figure 4 et 5 du chapitre 3). Par conséquent nous proposons de généraliser le constat que la diffusion translationnelle de l'eau d'hydratation permet l'existence des mouvements de large amplitude de la protéine, indispensables à la fonction biologique de cette dernière à l'ensemble du protéome.

C.3.2 Comparaison de la dynamique collective de l'eau d'hydratation des protéines Tau et MBP

L'étude précédente a porté sur l'étude des mouvements individuels (incohérents) de l'eau autour de l'IDP tau et de la protéine globulaire MBP. La dynamique d'un système se caractérise également par le mouvement des molécules les unes par rapport aux autres. Cette dynamique est dite collective et peut être mesurée, par exemple, par spectroscopie THz. Dans ce travail, nous avons caractérisé par spectroscopie THz les protéines tau et MBP en solution. D'une part, nous avons observé pour chacune des deux protéines une influence à longue portée sur la dynamique collective de l'eau environnante (de l'ordre de plusieurs dizaines d'Å depuis la surface de la protéine). D'autre part, nous avons montré que le volume d'eau autour de la protéine tau influencé par cette dernière est deux fois moins important que celui influencé par la MBP. Nous suggérons que ce résultat peut avoir une conséquence sur la fonction biologique des IDPs.

C.3.3 Mobilité de l'eau d'hydratation de la protéine tau dans sa forme native et fibrillaire

La protéine humaine intrinsèquement désordonnée tau peut former des structures fibreuses, composées d'un corps amyloïde rigide et d'une partie périphérique désordonnée, appelée "fuzzy coat" (voire figure 1 du chapitre 5). L'accumulation de ces fibres dans le cerveau est l'une des marques caractéristiques de la maladie d'Alzheimer. Les mécanismes de formation de ces fibres, et en particulier le rôle que pourrait jouer l'eau, sont très peu connus. En combinant diffusion de neutrons et simulations de dynamique moléculaire, nous avons étudié la dynamique de l'eau d'hydratation des fibres amyloïdes de la protéine tau. En comparaison avec la protéine native, l'eau d'hydratation des fibres s'est avérée être plus mobile, comme le montre une diffusion translationnelle plus importante des molécules d'eau ainsi qu'une augmentation des déplacements carrés moyens (voir figure 3 du chapitre 5). En utilisant le peptide $^{306}VQIVYK^{311}$ comme modèle pour le corps de la fibre, les simulations permettent d'attribuer cette augmentation de dynamique aux molécules d'eau se trouvant spécifiquement autour du "fuzzy coat". Nous suggérons que cette hausse de mobilité de l'eau d'hydratation après fibrillation puisse promouvoir la formation de fibres, par des effets de compensations thermodynamiques. De plus, une augmentation du coefficient de diffusion de d'eau autour de la forme pathologique de tau pourrait ouvrir des perspectives d'amélioration du diagnostic de la maladie d'Alzheimer par IRM de diffusion.

C.3.4 Une diversité dynamique au sein des agrégats de protéines

Dans certaines conditions physico-chimiques, les protéines peuvent s'agréger et devenir pathologiques. Ces agrégats sont responsables des maladies dites conformationnelles, qui incluent les maladies d'Alzheimer, de Parkinson ou encore le diabète de type 2. Bien que de nombreuses études se soient intéressées à la structure de ces agrégats, très peu de travaux ont porté sur la caractérisation de leur dynamique. Dans ce projet, nous avons étudié par diffusion élastique de neutrons la dynamique interne de plusieurs protéines (le lysozyme, la MBP, l'hexapeptide VQIVYK, et la protéine tau complète), dans différents états d'agrégation. En particulier, nous avons comparé la dynamique des protéines natives avec celle d'agrégats non structurés, dits amorphes, et d'agrégats structurés par des feuillets β inter-moléculaires, appelés amyloïdes. Les résultats des différentes mesures sont résumés dans le tableau C.1. Comme le montre ce tableau, la dynamique des protéines ne semble pas être directement reliée à leur état d'agrégation. En effet, pour un même type d'agrégat, on observe une augmentation, diminution, ou similarité de sa dynamique par rapport à la protéine monomérique. Il est intéressant de noter que la dynamique des formes agrégées est en majorité plus importante que celle des protéines natives. Cela contraste avec l'image intuitive des agrégats comme étant des structures rigides et peu mobiles. Pour pouvoir interpréter plus en détails ces résultats, nous nous proposons dans la section suivante de modéliser des poudres de protéines hydratées. L'étude *in silico* des poudres hydratées des protéines tau et MBP vise à une meilleure compréhension de leur dynamique à l'échelle moléculaire.

	Taux d'hydratation (g D2O/g de protéine)	Agrégats amyloïdes	Agrégats amorphes
Lysozyme	0.4	+	=
MBP	0.4	ND	+
VQIVYK	0.4	=	ND
htau40	0.4	=	ND
Concanavalin A (Schiro et al., 2012)	0.2	+	-

Table C.1: Comparaison de la dynamique des différents agrégats au dessus de la transition dynamique (c'est à dire dans l'intervalle de température 260-300 K), par rapport à la forme monomérique de chaque protéine. Les symboles + , - , ou = signifient que la dynamique de l'agrégat considéré est supérieure, inférieure ou égale à celle du monomère. L'abréviation ND indique l'absence de mesures.

C.3.5 Modèle d'une poudre hydratée de la protéine tau

Les poudres hydratées présentent l'avantage d'empêcher la translation/rotation globale de la protéine, tout en conservant sa dynamique interne (qui correspond essentiellement aux mouvements des chaines latérales). Pour cette raison, elles sont largement utilisées dans divers techniques expérimentales telles que la diffusion de neutrons ou la spectroscopie par

résonance magnétique nucléaire. Dans cette étude, nous avons modélisé les poudres des protéines tau et MBP hydratées à 0.4 gramme d'eau par gramme de protéines (voir les photos des boites de simulation en figure C.2). En comparant les dynamiques des protéines obtenues *in silico* avec les résultats expérimentaux, ainsi qu'en analysant différentes parties des protéines, nous concluons que les propriétés de l'interface eau-protéine sont déterminantes pour la dynamique globale de la protéine. De plus, nous observons que la transition dynamique, qui est supprimée dans le cas des poudres sèches, est présente dans les parties de protéine (hydratée) qui ne sont pas en contact avec l'eau. Cela suggère que les molécules d'eau se trouvant à la surface de la protéine ont un effet dynamique qui se propage dans toute la protéine.
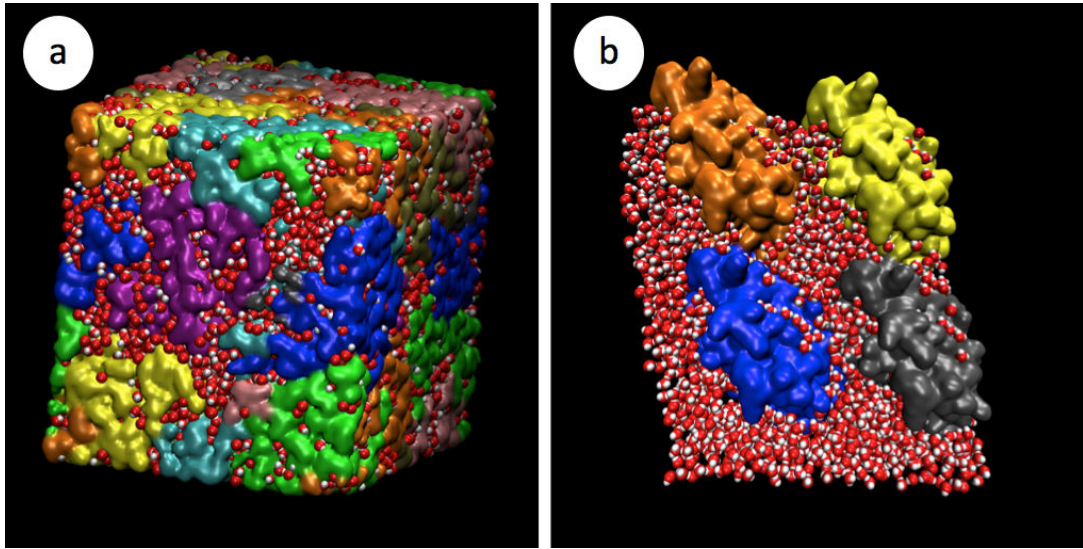
Figure C.2: Capture des boites de simulation des poudres hydraté de la protéine tau (a) et MBP (b).

C.4 Conclusion

Ce manuscrit de thèse présente une caractérisation dynamique de l'eau d'hydratation de la protéine intrinsèquement désordonnée tau, dans sa forme native et fibrillaire. Nous avons montré que, malgré des points communs dans la dynamique individuelle de l'eau d'hydratation des IDPs et des protéines globulaires, des différences significatives existent dans leurs dynamiques collectives. De plus, une diversité dynamique a été observée au sein des agrégats de protéines, que nous attribuons à des différences dans l'interface eau-agrégats. Enfin, nous avons mis en évidence des différences de mobilité de l'eau autour de la protéine tau, lorsque celle-ci est sous forme fonctionnelle ou pathologique. Nous suggérons que ces différences puissent jouer un rôle dans la formation des fibres amyloïdes.

En conclusion, ce travail avance l'hypothèse que la dynamique de l'eau d'hydratation est

une propriétés biologique importante, qui pourrait permettre de mieux comprendre certains processus moléculaires. Plutôt que de décrire avec précision les mécanismes liant dynamique de l'eau et fonction des protéines, ce travail incite à explorer davantage le rôle biologique de l'eau d'hydratation.

L'une des perspectives qui ressort de ces travaux est de mesurer la dynamique (individuelle ou collective) de l'eau d'hydratation en fonction du temps. En effet, le suivi des propriétés dynamiques de l'eau lors de processus biologiques, telles que la liaison et la fibrillation de protéines, pourrait permettre de mieux comprendre l'implication de l'eau dans l'activité des macromolécules. Cela nécessite l'utilisation de techniques expérimentales avancées possédant une résolution temporelle de l'ordre de la µs à la ms.

Bibliography

Amos, L. A. (2004). Microtubule structure and its stabilisation. *Organic & Biomolecular Chemistry*, 2(15):2153.

Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E., Sauke, T. B., Shyamsunder, E., and Young, R. D. (1985). Protein states and proteinquakes. *Proceedings of the National Academy of Sciences*, 82(15):5000–5004. PMID: 3860839.

Arnaudov, L. N. and de Vries, R. (2005). Thermally induced fibrillar aggregation of hen egg white lysozyme. *Biophysical Journal*, 88(1):515–526.

Astbury, W. T., Dickinson, S., and Bailey, K. (1935). The x-ray interpretation of denaturation and the structure of the seed globulins. *Biochemical Journal*, 29(10):2351.

Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., and Gunsalus, I. C. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–5373.

Autenrieth, F., Tajkhorshid, E., Schulten, K., and Luthey-Schulten, Z. (2004). Role of water in transient cytochrome c2 docking. *The Journal of Physical Chemistry B*, 108(52):20376–20387.

Bagchi, B. (2005). Water dynamics in the hydration layer around proteins and micelles. *Chemical Reviews*, 105(9):3197–3219.

Baldwin, A. J., Knowles, T. P. J., Tartaglia, G. G., Fitzpatrick, A. W., Devlin, G. L., Shammas, S. L., Waudby, C. A., Mossuto, M. F., Meehan, S., Gras, S. L., Christodoulou, J., Anthony-Cahill, S. J., Barker, P. D., Vendruscolo, M., and Dobson, C. M. (2011). Metastability of native proteins and the phenomenon of amyloid formation. *Journal of the American Chemical Society*, 133(36):14160–14163.

Baldwin, R. L. (2014). Dynamic hydration shell restores kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proceedings of the National Academy of Sciences*, 111(36):13052–13056.

Ball, P. (2008). Water as an active constituent in cell biology. *Chemical Reviews*, 108(1):74–108.

Ballatore, C., Lee, V. M., and Trojanowski, J. Q. (2007). Tau-mediated neurodegeneration in alzheimer's disease and related disorders. *Nature Reviews Neuroscience*, 8(9):663–672.

Barillari, C., Taylor, J., Viner, R., and Essex, J. W. (2007). Classification of water molecules in protein binding sites. *Journal of the American Chemical Society*, 129(9):2577–2587. PMID: 17288418.

Becker, T., Hayward, J. A., Finney, J. L., Daniel, R. M., and Smith, J. C. (2004). Neutron frequency windows and the protein dynamical transition. *Biophysical Journal*, 87(3):1436–1444.

Bee, M. (1988). *Quasielastic neutron scattering*.

Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271.

Bergner, A., Heugen, U., Brundermann, E., Schwaab, G., Havenith, M., Chamberlin, D. R., and Haller, E. E. (2005). New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water. *Review of Scientific Instruments*, 76(6):063110.

Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using Small-Angle x-ray scattering. *Journal of the American Chemical Society*, 129(17):5656–5664.

Breiten, B., Lockett, M. R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., Bowers, C. M., Heroux, A., Krilov, G., and Whitesides, G. M. (2013). Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc*, 135(41):15579–84.

Brust, R., Lukacs, A., Haigney, A., Addison, K., Gil, A., Towrie, M., Clark, I. P., Greetham, G. M., Tonge, P. J., and Meech, S. R. (2013). Proteins in action: Femtosecond to millisecond structural dynamics of a photoactive flavoprotein. *Journal of the American Chemical Society*, 135(43):16168–16174.

Caliskan, G., Briber, R. M., Thirumalai, D., Garcia-Sakai, V., Woodson, S. A., and Sokolov, A. P. (2006). Dynamic transition in tRNA is solvent induced. *Journal of the American Chemical Society*, 128(1):32–33.

Chen, S., Liu, L., Fratini, E., Baglioni, P., Faraone, A., and Mamontov, E. (2006). Observation of fragile-to-strong dynamic crossover in protein hydration water. *Proceedings of the National Academy of Sciences*, 103(24):9012–9016. PMID: 16751274.

Cheung, M. S., García, A. E., and Onuchic, J. N. (2002). Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences*, 99(2):685–690. PMID: 11805324.

Chiti, F. and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75(1):333–366.

Colombo, M. F., Rau, D. C., and Parsegian, V. A. (1992). Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science (New York, N.Y.)*, 256(5057):655–659. PMID: 1585178.

Conti Nibali, V. and Havenith, M. (2014). New insights into the role of water in biological function: Studying solvated biomolecules using terahertz absorption spectroscopy in conjunction with molecular dynamics simulations. *Journal of the American Chemical Society*, 136(37):12800–12807.

Daniel, R. M., Smith, J. C., Ferrand, M., Héry, S., Dunn, R., and Finney, J. L. (1998). Enzyme activity below the dynamical transition at 220 k. *Biophysical Journal*, 75(5):2504–2507.

Dielmann-Gessner, J., Grossman, M., Conti Nibali, V., Born, B., Solomonov, I., Fields, G. B., Havenith, M., and Sagi, I. (2014). Enzymatic turnover of macromolecules generates long-lasting protein–water-coupled motions beyond reaction steady state. *Proceedings of the National Academy of Sciences*, page 201410144.

Dolman, M., Halling, P. J., Moore, B. D., and Waldron, S. (1997). How dry are anhydrous enzymes? measurement of residual and buried 18O-labeled water molecules using mass spectrometry. *Biopolymers*, 41(3):313–321.

Doster, W. (2008). The dynamical transition of proteins, concepts and misconceptions. *European Biophysics Journal*, 37(5):591–602.

Doster, W., Busch, S., Gaspar, A. M., Appavou, M. S., Wuttke, J., and Scheer, H. (2010). Dynamical transition of protein-hydration water. *Phys Rev Lett*, 104(9):098101.

Doster, W., Cusack, S., and Petry, W. (1989). Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature*, 337(6209):754–756. PMID: 2918910.

Doster, W., Kleinert, T., Post, F., and Settles, M. (1993). Hydration effects on protein function: the kinetics of ligand binding to myoglobin. *Proceedings of the Italian Physical Society*, 43:127–130.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., et al. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1):26–59.

Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., Brown, C. J., et al. (2000). Intrinsic protein disorder in complete genomes. *Genome Informatics Series*, page 161–171.

Dyson, H. J. (2011). Expanding the proteome: disordered and alternatively folded proteins. *Quarterly Reviews of Biophysics*, 44(04):467–518.

Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3):197–208.

Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Gruebele, M., Leitner, D. M., and Havenith, M. (2008). Protein sequence- and pH-Dependent hydration probed by terahertz spectroscopy. *Journal of the American Chemical Society*, 130(8):2374–2375.

Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Heugen, U., Gruebele, M., Leitner, D. M., and Havenith, M. (2007). An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences*, 104(52):20749–20752.

Ellis, R. J. and Minton, A. P. (2003). Cell biology: Join the crowd. *Nature*, 425(6953):27–28.

Ermler, U., Fritzsch, G., Buchanan, S. K., and Michel, H. (1994). Structure of the photosynthetic reaction centre from rhodobacter sphaeroides at 2.65 a resolution: cofactors and protein-cofactor interactions. *Structure (London, England: 1993)*, 2(10):925–936. PMID: 7866744.

Fenimore, P. W., Frauenfelder, H., McMahon, B. H., and Young, R. D. (2004). Bulk-solvent and hydration-shell fluctuations, similar to α- and β-fluctuations in glasses, control protein motions and functions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(40):14408–14413. PMID: 15448207.

Ferrand, M., Dianoux, A. J., Petry, W., and Zaccaï, G. (1993). Thermal motions and function of bacteriorhodopsin in purple membranes: effects of temperature and hydration studied by neutron scattering. *Proceedings of the National Academy of Sciences*, 90(20):9668–9672. PMID: 8415760.

Fichou, Y., Schirò, G., Gallat, F., Laguri, C., Moulin, M., Combet, J., Zamponi, M., Härtlein, M., Picart, C., Mossou, E., Lortat-Jacob, H., Colletier, J., Tobias, D. J., and Weik, M. (2015). Hydration water mobility is enhanced around tau amyloid fibers. *Proceedings of the National Academy of Sciences*, 112(20):6365–6370. PMID: 25918405.

Francisco, W. A., Wille, G., Smith, A. J., Merkler, D. J., and Klinman, J. P. (2004). Investigation of the pathway for inter-copper electron transfer in peptidylglycine alpha-amidating monooxygenase. *Journal of the American Chemical Society*, 126(41):13168–13169. PMID: 15479039.

Frick, B. and Gonzalez, M. (2001). Five years operation of the second generation backscattering spectrometer IN16—a retrospective, recent developments and plans. *Physica B: Condensed Matter*, 301(1–2):8–19.

Fuxreiter, M., Mezei, M., Simon, I., and Osman, R. (2005). Interfacial water as a "hydration fingerprint" in the noncognate complex of BamHI. *Biophysical Journal*, 89(2):903–911. PMID: 15894630 PMCID: PMC1366640.

Gallat, F. (2011). *Dynamique des protéines et de la couche d'hydratation étudiée par diffusion de neutrons et méthodes biophysiques complémentaires*. PhD thesis.

Gallat, F., Brogan, A. P. S., Fichou, Y., McGrath, N., Moulin, M., Härtlein, M., Combet, J., Wuttke, J., Mann, S., Zaccai, G., Jackson, C. J., Perriman, A. W., and Weik, M. (2012a). A polymer surfactant corona dynamically replaces water in Solvent-Free protein liquids and ensures macromolecular flexibility and activity. *Journal of the American Chemical Society*, 134(32):13168–13171.

Gallat, F., Laganowsky, A., Wood, K., Gabel, F., van Eijck, L., Wuttke, J., Moulin, M., Härtlein, M., Eisenberg, D., Colletier, J., Zaccai, G., and Weik, M. (2012b). Dynamical coupling of intrinsically disordered proteins and their hydration water: Comparison with folded soluble and membrane proteins. *Biophysical Journal*, 103(1):129–136.

Garcia, A. E. and Onuchic, J. N. (2003). Folding a protein in a computer: An atomic description of the folding/unfolding of protein a. *Proceedings of the National Academy of Sciences*, 100(24):13898–13903. PMID: 14623983.

Goedert, M., Klug, A., and Crowther, R. A. (2006). Tau protein, the paired helical filament and alzheimer's disease. *Journal of Alzheimer's Disease*, 9(0):195–207.

Goodsell, D. S. (2009). Escherichia coli. *Biochemistry and Molecular Biology Education*, 37(6):325–332.

Greenwald, J. and Riek, R. (2010). Biology of amyloid: Structure, function, and regulation. *Structure*, 18(10):1244–1260.

Grossman, M., Born, B., Heyden, M., Tworowski, D., Fields, G. B., Sagi, I., and Havenith, M. (2011). Correlated structural kinetics and retarded solvent dynamics at the metallo-protease active site. *Nature Structural & Molecular Biology*, 18(10):1102–1108.

Halle, B. (2004). Protein hydration dynamics in solution: a critical survey. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1448):1207–1328. PMID: 15306377 PMCID: PMC1693401.

Harano, Y. and Kinoshita, M. (2004). Large gain in translational entropy of water is a major driving force in protein folding. *Chemical Physics Letters*, 399(4–6):342–348.

Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172):964–972.

Heyden, M., Sun, J., Funkner, S., Mathias, G., Forbert, H., Havenith, M., and Marx, D. (2010). Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *Proceedings of the National Academy of Sciences*, 107(27):12068–12073.

Heyden, M. and Tobias, D. J. (2013). Spatial dependence of Protein-Water collective Hydrogen-Bond dynamics. *Physical Review Letters*, 111(21):218101.

Jahn, T. R. and Radford, S. E. (2005). The yin and yang of protein folding: The yin and yang of protein folding. *FEBS Journal*, 272(23):5962–5970.

Jasnin, M., Eijck, L. v., Koza, M. M., Peters, J., Laguri, C., Lortat-Jacob, H., and Zaccai, G. (2010). Dynamics of heparan sulfate explored by neutron scattering. *Physical Chemistry Chemical Physics*, 12(14):3360–3362.

Jasnin, M., Moulin, M., Haertlein, M., Zaccai, G., and Tehei, M. (2008). Down to atomic-scale intracellular water dynamics. *EMBO reports*, 9(6):543–547.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935.

Khodadadi, S., Pawlus, S., and Sokolov, A. P. (2008). Influence of hydration on protein dynamics: Combining dielectric and neutron scattering spectroscopy data. *The Journal of Physical Chemistry B*, 112(45):14273–14280.

Kim, S. J., Born, B., Havenith, M., and Gruebele, M. (2008). Real-Time detection of Protein-Water dynamics upon protein folding by terahertz absorption spectroscopy. *Angewandte Chemie International Edition*, 47(34):6486–6489.

King, J. T., Arthur, E. J., Brooks, C. L., and Kubarych, K. J. (2014). Crowding induced collective hydration of biological macromolecules over extended distances. *Journal of the American Chemical Society*, 136(1):188–194.

Kornblatt, J. A. and Hoa, G. H. (1990). A nontraditional role for water in the cytochrome c oxidase reaction. *Biochemistry*, 29(40):9370–9376. PMID: 2174258.

Krishnan, M., Kurkal-Siebert, V., and Smith, J. C. (2008). Methyl group dynamics and the onset of anharmonicity in myoglobin. *The Journal of Physical Chemistry B*, 112(17):5522–5533.

Lagi, M., Chu, X., Kim, C., Mallamace, F., Baglioni, P., and Chen, S. (2008). The Low-Temperature dynamic crossover phenomenon in protein hydration water: Simulations vs experiments. *The Journal of Physical Chemistry B*, 112(6):1571–1575.

Leitner, D. M., Gruebele, M., and Havenith, M. (2008). Solvation dynamics of biomolecules: modeling and terahertz experiments. *HFSP journal*, 2(6):314–323.

Linderstrom-Lang, K. I. and Schellman, J. (1959). *The enzymes*. Academic press, New York.

Lukyanov, K. A., Chudakov, D. M., Lukyanov, S., and Verkhusha, V. V. (2005). Photoactivatable fluorescent proteins. *Nature Reviews Molecular Cell Biology*, 6(11):885–890.

Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., Singru, P. S., Nilsson, K. P. R., Simon, R., Schubert, D., Eisenberg, D., Rivier, J., Sawchenko, P., Vale, W., and Riek, R. (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science*, 325(5938):328–332. PMID: 19541956.

Malisauskas, M., Zamotin, V., Jass, J., Noppe, W., Dobson, C. M., and Morozova-Roche, L. A. (2003). Amyloid protofilaments from the calcium-binding protein equine lysozyme: formation of ring and linear structures depends on pH and metal ion concentration. *Journal of Molecular Biology*, 330(4):879–890. PMID: 12850154.

Mandelkow, E. M. and Mandelkow, E. (2012). Biochemistry and cell biology of tau protein in neurofibrillary degeneration. *Cold Spring Harb Perspect Med*, 2(7):a006247.

Meister, K., Ebbinghaus, S., Xu, Y., Duman, J. G., DeVries, A., Gruebele, M., Leitner, D. M., and Havenith, M. (2013). Long-range protein–water dynamics in hyperactive insect antifreeze proteins. *Proceedings of the National Academy of Sciences*, 110(5):1617–1622. PMID: 23277543.

Merzel, F. and Smith, J. C. (2002). Is the first hydration shell of lysozyme of higher density than bulk water? *Proceedings of the National Academy of Sciences*, 99(8):5378–5383. PMID: 11959992.

Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001). Studying excited states of proteins by NMR spectroscopy. *Nature Structural & Molecular Biology*, 8(11):932–935.

Niehues, G., Heyden, M., Schmidt, D. A., and Havenith, M. (2011). Exploring hydrobicity by THz absorption spectroscopy of solvated amino acids. *Faraday Discussions*, 150:193.

Orecchini, A., Sebastiani, F., Jasnin, M., Paciaroni, A., De Francesco, A., Petrillo, C., Moulin, M., Haertlein, M., Zaccai, G., and Sacchetti, F. (2012). Collective dynamics of intracellular water in living cells. *Journal of Physics: Conference Series*, 340(1):012091.

Ortony, J. H., Cheng, C., Franck, J. M., Kausik, R., Pavlova, A., Hunt, J., and Han, S. (2011). Probing the hydration water diffusion of macromolecular surfaces and interfaces. *New Journal of Physics*, 13(1):015006.

Ostermann, A., Waschipky, R., Parak, F. G., and Nienhaus, G. U. (2000). Ligand binding and conformational motions in myoglobin. *Nature*, 404(6774):205–208.

Pal, S. K., Peon, J., and Zewail, A. H. (2002). Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proceedings of the National Academy of Sciences*, 99(4):1763–1768.

Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z., and Wolynes, P. G. (2004). Water in protein structure prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10):3352–3357. PMID: 14988499.

Perriman, A. W., Brogan, A. P. S., Cölfen, H., Tsoureas, N., Owen, G. R., and Mann, S. (2010). Reversible dioxygen binding in solvent-free liquid myoglobin. *Nature Chemistry*, 2(8):622–626.

Pieper, J., Hauss, T., Buchsteiner, A., Baczyński, K., Adamiak, K., Lechner, R. E., and Renger, G. (2007). Temperature- and Hydration-Dependent protein dynamics in photosystem II of green plants studied by quasielastic neutron scattering†. *Biochemistry*, 46(40):11398–11409.

Plotkin, S. S. and Onuchic, J. N. (2002a). Understanding protein folding with energy landscape theory part i: Basic concepts. *Quarterly Reviews of Biophysics*, 35(02):111–167.

Plotkin, S. S. and Onuchic, J. N. (2002b). Understanding protein folding with energy landscape theory part II: quantitative aspects. *Quarterly Reviews of Biophysics*, 35(03):205–286.

Rasmussen, B. F., Stock, A. M., Ringe, D., and Petsko, G. A. (1992). Crystalline ribonuclease a loses function below the dynamical transition at 220 k. *Nature*, 357(6377):423–424.

Robinson, C. R. and Sligar, S. G. (1993). Molecular recognition mediated by bound water: A mechanism for star activity of the restriction endonuclease EcoRI. *Journal of Molecular Biology*, 234(2):302–306.

Roh, J., Curtis, J., Azzam, S., Novikov, V., Peral, I., Chowdhuri, Z., Gregory, R., and Sokolov, A. (2006). Influence of hydration on the dynamics of lysozyme. *Biophysical Journal*, 91(7):2573–2588.

Roh, J., Novikov, V., Gregory, R., Curtis, J., Chowdhuri, Z., and Sokolov, A. (2005). Onsets of anharmonicity in protein dynamics. *Physical Review Letters*, 95(3).

Rojas Quijano, F. A., Morrow, D., Wise, B. M., Brancia, F. L., and Goux, W. J. (2006). Prediction of nucleating sequences from amyloidogenic propensities of Tau-Related peptides. *Biochemistry*, 45(14):4638–4652.

Ross, C. A. and Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *, Published online: 01 July 2004; | doi:10.1038/nm1066*, 10:S10–S17.

Ross, C. A. and Poirier, M. A. (2005). What is the role of protein aggregation in neurodegeneration? *Nature Reviews Molecular Cell Biology*, 6(11):891–898.

Royer, W. E., Pardanani, A., Gibson, Q. H., Peterson, E. S., and Friedman, J. M. (1996). Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proceedings of the National Academy of Sciences*, 93(25):14526–14531. PMID: 8962085.

Rupley, J. A. and Careri, G. (1991). Protein hydration and function. *Advances in Protein Chemistry*, 41:37–172. PMID: 2069077.

Sanchez-Medina, C., Sekhar, A., Vallurupalli, P., Cerminara, M., Muñoz, V., and Kay, L. E. (2014). Probing the free energy landscape of the Fast-Folding gpW protein by relaxation dispersion NMR. *Journal of the American Chemical Society*, 136(20):7444–7451.

Schiro, G., Caronna, C., Natali, F., and Cupane, A. (2010a). Direct evidence of the amino acid side chain and backbone contributions to protein anharmonicity. *Journal of the American Chemical Society*, 132(4):1371–1376.

Schiro, G., Caronna, C., Natali, F., and Cupane, A. (2010b). Molecular origin and hydration dependence of protein anharmonicity: an elastic neutron scattering study. *Physical Chemistry Chemical Physics*, 12(35):10215–10220.

Schiro, G., Caronna, C., Natali, F., Koza, M. M., and Cupane, A. (2011). The "Protein dynamical transition" does not require the protein polypeptide chain. *The Journal of Physical Chemistry Letters*, pages 2275–2279.

Schiro, G., Fichou, Y., Gallat, F., Wood, K., Gabel, F., Moulin, M., Härtlein, M., Heyden, M., Colletier, J., Orecchini, A., Paciaroni, A., Wuttke, J., Tobias, D. J., and Weik, M. (2015). Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins. *Nature Communications*, 6:6490.

Schiro, G., Vetri, V., Frick, B., Militello, V., Leone, M., and Cupane, A. (2012). Neutron scattering reveals enhanced protein dynamics in concanavalin a amyloid fibrils. *The Journal of Physical Chemistry Letters*, 3(8):992–996.

Schotte, F., Lim, M., Jackson, T. A., Smirnov, A. V., Soman, J., Olson, J. S., Phillips, G. N., Wulff, M., and Anfinrud, P. A. (2003). Watching a protein as it functions with 150-ps Time-Resolved x-ray crystallography. *Science*, 300(5627):1944–1947. PMID: 12817148.

Seki, Y., Tomizawa, T., Khechinashvili, N. N., and Soda, K. (2002). Contribution of solvent water to the solution x-ray scattering profile of proteins. *Biophysical Chemistry*, 95(3):235–252.

Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007). *Methods in Molecular Biophysics: Structure, Dynamics, Function.* Cambridge University Press.

Sibille, N. and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochemical Society Transactions*, 40(5):955–962.

Sugase, K., Dyson, H. J., and Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–1025.

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. F. (1997). Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *Journal of Molecular Biology*, 273(3):729–739.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S., and Zaccai, G. (1998). Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences*, 95(5):2267–2272. PMID: 9482874.

Szent-Gyorgyi, A. (1979). *Cell-associated water.* Academic press, New York.

Tarek, M. and Tobias, D. (2002a). Role of Protein-Water hydrogen bond dynamics in the protein dynamical transition. *Physical Review Letters*, 88(13).

Tarek, M. and Tobias, D. J. (2002b). Single-Particle and collective dynamics of protein hydration water: A molecular dynamics study. *Physical Review Letters*, 89(27).

Tehei, M., Franzetti, B., Madern, D., Ginzburg, M., Ginzburg, B. Z., Giudici-Orticoni, M., Bruschi, M., and Zaccai, G. (2004). Adaptation to extreme environments: macromolecular dynamics in bacteria compared in vivo by neutron scattering. *EMBO reports*, 5(1):66–70.

Thirumalai, D., Reddy, G., and Straub, J. E. (2012). Role of water in protein aggregation and amyloid polymorphism. *Accounts of Chemical Research*, 45(1):83–92.

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in biochemical sciences*, 27(10):527–533.

Turoverov, K. K., Kuznetsova, I. M., and Uversky, V. N. (2010). The protein kingdom extended: Ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Progress in Biophysics and Molecular Biology*, 102(2–3):73–84.

Uversky, V. N. (2014). Introduction to intrinsically disordered proteins (IDPs). *Chemical Reviews*, 114(13):6557–6560.

Uversky, V. N. and Dunker, A. K. (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1804(6):1231–1264.

Uversky, V. N. and Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1698(2):131–153.

Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: Introducing the d2 concept. *Annual Review of Biophysics*, 37(1):215–246. PMID: 18573080.

Uversky, V. N., Oldfield, C. J., Midic, U., Xie, H., Xue, B., Vucetic, S., Iakoucheva, L. M., Obradovic, Z., and Dunker, A. K. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, 10(Suppl 1):S7. PMID: 19594884.

Vega, C. and Abascal, J. L. F. (2011). Simulating water with rigid non-polarizable models: a general perspective. *Physical Chemistry Chemical Physics*, 13(44):19663.

Wood, K., Frölich, A., Paciaroni, A., Moulin, M., Härtlein, M., Zaccai, G., Tobias, D. J., and Weik, M. (2008). Coincidence of dynamical transitions in a soluble protein and its hydration water: Direct measurements by neutron scattering and MD simulations. *Journal of the American Chemical Society*, 130(14):4586–4587.

Wood, K., Plazanet, M., Gabel, F., Kessler, B., Oesterhelt, D., Tobias, D. J., Zaccai, G., and Weik, M. (2007). Coupling of protein and hydration-water dynamics in biological membranes. *Proceedings of the National Academy of Sciences*, 104(46):18049–18054.

Wood, K., Tobias, D. J., Kessler, B., Gabel, F., Oesterhelt, D., Mulder, F. A. A., Zaccai, G., and Weik, M. (2010). The Low-Temperature inflection observed in neutron scattering measurements of proteins is due to methyl rotation: Direct evidence using isotope labeling and molecular dynamics simulations. *Journal of the American Chemical Society*, 132(14):4990–4991.

Wuttke, J., Budwig, A., Drochner, M., Kammerling, H., Kayser, F. J., Kleines, H., Ossovyi, V., Pardo, L. C., Prager, M., Richter, D., Schneider, G. J., Schneider, H., and Staringer, S. (2012). SPHERES, julich's high-flux neutron backscattering spectrometer at FRM II. *Rev Sci Instrum*, 83(7):075109.

Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., and

Obradovic, Z. (2007). Functional anthology of intrinsic disorder. i. biological processes and functions of proteins with long disordered regions. *Journal of proteome research*, 6(5):1882–1898. PMID: 17391014 PMCID: PMC2543138.

Xu, L., Kumar, P., Buldyrev, S. V., Chen, S., Poole, P. H., Sciortino, F., and Stanley, H. E. (2005). Relation between the widom line and the dynamic crossover in systems with a liquid–liquid phase transition. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16558–16562. PMID: 16267132.

Yi, Z., Miao, Y., Baudry, J., Jain, N., and Smith, J. C. (2012). Derivation of Mean-Square displacements for protein dynamics from elastic incoherent neutron scattering. *Journal of Physical Chemistry B*.

Zaccai, G. (1987). Structure and hydration of purple membranes in different conditions. *Journal of Molecular Biology*, 194(3):569–572.

Zaccai, G. (2000). How soft is a protein? a protein dynamics force constant measured by neutron scattering. *Science*, 288(5471):1604–1607.

Zhong, D., Pal, S. K., and Zewail, A. H. (2011). Biological water: A critique. *Chemical Physics Letters*, 503(1):1–11.

Zhong, D., Pal, S. K., Zhang, D., Chan, S. I., and Zewail, A. H. (2002). Femtosecond dynamics of rubredoxin: Tryptophan solvation and resonance energy transfer in the protein. *Proceedings of the National Academy of Sciences*, 99(1):13–18.

Résumé

Les protéines qui ne possèdent pas de structure unique dans leur forme fonctionnelle constituent la classe des protéines intrinsèquement dépliées (IDPs). Ces dernières sont ubiquitaires dans une cellule et sont connues pour former des agrégats impliqués dans une large variété de maladies. Malgré leurs conformations étendues qui résultent en une large interface avec l'eau environnante, très peu d'informations sont connues sur l'interaction des IDPs avec l'eau. L'eau est parfois appelée la *matrice de la vie* car elle est indispensable à la plupart des processus biologiques, tels que le repliement, la stabilité ou l'activité des protéines. La protéine tau est une IDP qui régule la dynamique de croissance des microtubules dans les neurones, et dont la fibrillation en fibres de type amyloïde est l'une des marques caractéristiques de la maladie d'Alzheimer. Ce projet de thèse se propose d'explorer l'importance biologique de la dynamique de l'eau autour des IDPs. Nous combinons des méthodes expérimentales et computationnelles, incluant la diffusion incohérente de neutrons, la spectroscopie terahertz, la diffusion de rayons X aux petits angles, et les simulations de dynamique moléculaire, dans le but d'étudier la dynamique de l'eau d'hydratation de la protéine tau, dans ses formes native et fibrillaire. Pour les IDPs comme pour les protéines globulaires, il est montré que la diffusion translationnelle de l'eau d'hydratation permet l'existence des mouvements de larges amplitudes de la protéine, indispensables à la fonction biologique de cette dernière. En comparant avec la forme native, nous mettons aussi en évidence une augmentation de la mobilité de l'eau d'hydratation de la forme fibrillaire de tau. Nous proposons que cette augmentation joue un rôle dans la formation des fibres. De plus, l'étude de la dynamique collective de l'eau d'hydratation montre que la protéine tau influence un volume d'eau deux fois moindre qu'une protéine globulaire équivalente, ce qui pourrait être impliqué dans son mécanisme de liaison avec un partenaire. En conclusion, en étudiant les propriétés dynamiques de l'eau autour des IDPs, ces travaux de thèse suggèrent que la dynamique de l'eau d'hydratation pourrait jouer un rôle fondamental dans les mécanismes de liaison et de fibrillation des IDPs.

Abstract

Proteins that do not have a well-defined structure in their functional state are referred to as intrinsically disordered proteins (IDPs). IDPs are ubiquitous in biological cells and their aggregation is involved in many diseases. The extended conformations of IDPs result in a large water interface, yet, interactions between IDPs and water are only scarcely documented. Water has been termed the *matrix of life* because it is essential for a variety of molecular processes, including protein folding, stability, and activity. The IDP tau regulates microtubule activity in neurons and is known to form amyloid fibers that are one of the hallmarks of Alzheimer disease. In this PhD thesis, the biological relevance of water dynamics around IDPs is addressed. We combine computational and experimental approaches, including all-atom MD simulations, incoherent neutron scattering, terahertz spectroscopy and small angle X-ray scattering, to study the hydration water dynamics of the tau protein in its native and fibrillated states. Firstly, a translational diffusion of hydration water molecules is found to be essential for biologically relevant dynamics of both IDPs and globular proteins. Secondly, compared to monomers, we find an enhancement of hydration water mobility around tau amyloid fibers that is suggested to play a role in fiber formation. Finally, the investigation of collective water dynamics reveals that the tau protein influences about two times less water molecules than a globular protein, which might be involved in tau's binding mechanisms. In conclusion, this piece of work investigated the dynamical properties of water around IDPs and suggests that the hydration water dynamics might play fundamental roles in binding and aggregation of IDPs.