

MAS NMR Studies of Hierarchical Interplay in Protein Dynamics

Présentée le 29 avril 2021

Faculté des sciences de base
Laboratoire de résonance magnétique
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Baptiste BUSI

Acceptée sur proposition du jury

Prof. M. Chergui, président du jury
Prof. D. L. Emsley, directeur de thèse
Prof. T. Polenova, rapporteuse
Prof. M. Hong, rapporteuse
Prof. B. Fierz, rapporteur

Acknowledgements

First and foremost, I would like to thank all the members of my jury: Majed Chergui, (president of the jury), Mei Hong, Tatyana Polenova (external experts) and Beat Fierz (internal expert) for accepting to read and evaluate my thesis.

I want to especially thank my PhD director Lyndon Emsley for all of his support during my PhD. He was always available to provide scientific advice as well as countless corrections to my work. I cannot express in words how lucky I was to have such a great supervisor.

I also want to thank Nadia Gauljaux and Anne Lene Odegaard for their help with the administrative work, which I would not have been able to do on my own.

A great thanks also to Pascal Mieville, Anto Barisic and Vlado Janovic for their help with the spectrometers, especially when it was Friday evening and I needed their tools to fix the cryogenic systems. I would also like to acknowledge Emilie Baudat, for her help and for starting the ants project with Sean Mcgregor. I would also like to thank Sean for his ability to make the repetition of the same experiment 150 times in a day interesting with his great conversation.

I also want to thank, Basile Vuichoud, Jonas Milani, Jasmine Viger-Gravel, and Estel Canet for their warm welcome when I arrived first at the EPFL. I also want to thank Aurélien Bornet for the book exchanges and discussion, well needed to take a short pause from real science.

I also want to thank Arthur Pinon, Marie Le Guern-Herry, Andrae Thomazo and Thomas Crespo for their support even though they were focused on their own PhD. And a special thanks to Arthur for having been at my side for all the great decisions that are made at 4 am.

I also want to thank Dominick Kubicki, Brennan Walder, Gabriele Stevanato, Martin Balodis, Aditya Mishra, Anna Morales-Melgares, Pinelopi Moutzouri, Yu Rao, Bruno Simões de Almeida and Manuel Cordova, for their scientific discussions and the occasional beer/BBQ by the lake. I would also like to thank Jayasubba Reddy Yarava, for teaching me some of the technical aspects of NMR. A great thanks also to Albert Hofstetter, Federico Paruzzo, and

Claudia Avalos for their taste in party. A special thanks to Snaedis Björgvindsdóttir for her work in making our office the best and greatest of the whole building. And thanks again to Snaedis, Albert and Claudia for the climbing session that allowed me to keep my health during my PhD. A specific thanks to Claudia for pushing me to go running even during bouts of great laziness.

A special place in heaven for Andrea Bertarello, without whom this thesis would not have seen the light of day.

I also want to thank my parents for their undying support and their regular gift of home grown food and vegetables. And I want to also thanks my siblings, which have gone out of their way to find artistic postcards with sweets messages to lighten my days.

Lausanne, le 16 Novembre 2020

Abstract

Nuclear Magnetic Resonance (NMR) has been used to study biomolecules for several decades. Solid-State NMR, specifically with Magic Angle Spinning (MAS), is an emerging tool that yields rich information of insoluble or poorly soluble biomolecules. The dependence of the signal on local chemical environments allows a selective description of both position and behavior in different parts of biomolecules. More specifically several NMR relaxation observables (T_1 , T_2 , $T_{1\rho}$...) can give valuable information about the molecular level dynamics of the sample, such as the timescale or the amplitude of local motions.

Among the different classes of complex biomolecules, proteins are of particular interest due to their extreme variability in terms of function since virtually all known biological processes require the intervention of several specific proteins. And despite the enormous progresses that has been made to understand their behavior, there is still a long way to go before being able to predict accurately and easily their behavior.

In order to fully describe proteins, the concept of dynamism must be included. Proteins are indeed not static objects and their motions are a key component of their biological functions. Understanding the interplay between protein function and dynamics is currently one of the fundamental challenges of physical biology. One classical way of understanding dynamics is to observe the energy landscape. In order to have access to the distribution of the physical values describing this energy landscape, here we develop an approach based on the observation of the temperature and magnetic field dependence of T_1 . We consider that protein motions will roughly follow the Arrhenius law, and we use well established descriptions from Redfield theory to fit the temperature dependence of T_1 in order to extract activation energies.

The correlation between the model and our observables allows us to accurately describe aspects of the energy landscape driving protein dynamics. Specifically we measure relaxation at multiple magnetic field strengths, allowing us to better constrain the motional models, and to simultaneously evaluate the robustness and physical basis of the method. The data reveal backbone and sidechain motions exhibiting low and high-energy modes with temperature coefficients around $5 \text{ kJ}\cdot\text{mol}^{-1}$ and $25 \text{ kJ}\cdot\text{mol}^{-1}$. The results are compared to variable temperature molecular dynamics simulation of the crystal

lattice, providing further support for the interpretation of the experimental data in terms of molecular motion.

We then use the multi field variable temperature NMR longitudinal relaxation approach to determine the hierarchical activation energies of motions of four different proteins: two small globular proteins (GB1 and the SH3 domain of α -spectrin), an intrinsically disordered protein (the N-terminus of the nucleoprotein of the Sendai virus, Sendai Ntail), and an outer membrane protein (OmpG). We were able to identify similarities and differences in the average motions of the proteins. We find that the NMR relaxation properties of the four proteins do share similar features. The data characterizing average backbone motions are found to be very similar, the same for methyl group rotations, and similar activation energies are measured. The main observed difference occurs for the intrinsically disordered Sendai Ntail, where we observe much lower energy of activation for motions of protons associated with the protein-solvent interface as compared to the others. We also observe variability between the proteins regarding side chain ^{15}N relaxation of lysine residues, with a higher activation energy observed in OmpG. This hints at strong interactions with negatively charged lipids in the bilayer and provides a possible mechanistic clue for the ‘positive-inside’ rule for helical membrane proteins. Overall, these observations refine the understanding of the similarities and differences between hierarchical dynamics in proteins.

Finally, in another direction, we describe progress towards an approach to determine the fat distribution in ants, and the consequences for distribution of labor.

Keywords

Protein dynamics, nuclear magnetic resonance, activation energies, nuclear spin relaxation

Résumé

La Résonnance Magnétique Nucléaire ou RMN est utilisé pour étudier les biomolécules depuis plusieurs décennies. La RMN des solides, plus particulièrement avec une rotation à l'angle magique, est un outil permettant de donner de nombreuses informations pour les biomolécules insolubles ou peu soluble. La dépendance du signal de chaque noyau avec l'environnement chimique de celui-ci permet une description spécifique et donc précise de la position et du comportement des différentes parties composant les biomolécules. Plus précisément différents paramètres de relaxation (T_1 , T_2 , $T_{1\rho}$...) peuvent donner des informations importantes au sujet des dynamiques d'un échantillon donné, tel que l'amplitude ou le temps caractéristique de certain mouvement.

Parmi les biomolécules complexes, les protéines sont sujettes à un intérêt particulier de par leur extrême variabilité d'action. En effet la quasi-totalité des processus biologiques connu requière l'assistance de plusieurs protéines. Et malgré les progrès importants ayant été accomplis pour comprendre leurs comportements, de nombreux efforts sont encore nécessaire avant de pouvoir prédire et modéliser facilement et avec précision leurs actions.

Afin de décrire de manière complète les protéines, il faut considérer leur dynamique. En effet il ne s'agit pas d'objets statiques et leurs mouvements sont un composant clé de leurs fonctions biologique. Comprendre les liens entre dynamique et fonction est l'une des questions fondamentales de la biophysique. Une manière classique d'expliquer la dynamique est de regarder le paysage énergétique. Pour avoir accès à la distribution des constantes physiques décrivant le paysage énergétique, nous observons la variabilité du T_1 par rapport à la température et au champ magnétique. Nous considérons que les mouvements interne des protéines vont suivre approximativement une loi de type Arrhenius, et nous avons utilisé des descriptions reconnus et utilisés largement, issue de la « Redfield Theory » pour décrire l'influence de la température sur T_1 afin d'extraire les énergies d'activation.

Les corrélations entre le modèle et les observables nous permet de décrire précisément la forme du paysage énergétique qui guide la dynamique des protéines. Plus spécifiquement nous avons mesuré la relaxation à différentes forces de champ magnétique, ce qui permet de mieux définir le modèle, et pour en même temps évaluer la robustesse et les sous-basements physique de la méthode. Les données révèlent des mouvements du squelette de la protéine et des chaînes latérales avec des modes énergétiques haut et bas, avec des énergies d'activation aux alentours de $5 \text{ kJ}\cdot\text{mol}^{-1}$ et $25 \text{ kJ}\cdot\text{mol}^{-1}$. Les

résultats sont comparables à des simulations de dynamique moléculaire à différentes températures d'une structure cristalline de protéine. Ceci confirme l'interprétation des données expérimentales au sujet des mouvements moléculaires.

Nous avons utilisé la relaxation longitudinale de RMN à différentes températures et champs pour déterminer la hiérarchie des énergies d'activation de dynamique de quatre protéines différentes : deux petites protéines globulaires (GB1 et le domaine SH3 de l' α -spectrin), une protéine intrinsèquement désordonnée (la partie N-terminus de la nucléoprotéine du virus Sendai, Sendai Ntail) et une protéine de la membrane externe (OmpG). Nous avons été capables d'identifier les similarités et les différences dans la dynamique moyenne des protéines. Nous avons trouvé que les paramètres de relaxation pour les quatre protéines ont de nombreuses similarités. Les données décrivant les moyennes des dynamiques du squelette protéique sont très similaires, ainsi que les rotations des groupes méthyle. Des énergies d'activation similaires sont obtenues. La différence principale est observée pour la protéine désordonnée Sendai Ntail, où une plus faible énergie d'activation pour les mouvements des protons associés à l'interface protéine-solvant comparée aux autres protéines. Une variabilité est observée par relaxation de ^{15}N entre les différentes protéines au sujet des chaînes latérales des lysines, avec une énergie d'activation plus élevée chez OmpG. Ceci peut indiquer des interactions fortes avec les lipides chargés négativement dans la bicouche et propose un mécanisme possible pour la règle « positive-inside » des protéines membranaires hélicoïdales. Globalement, ces observations précisent la compréhension des points communs et des différences entre les différents mouvements hiérarchisés dans les protéines.

Finalement, dans un autre registre, nous décrivons le progrès vers une approche pour déterminer la distribution de tâches parmi les fourmis, et ses conséquences pour la division du travail.

Mots-clés

Dynamique des protéines, Résonance magnétique nucléaire, Énergie d'activation, Relaxométrie

Contents

Acknowledgements.....	vii
Abstract.....	x
Keywords	xi
Résumé	xii
Mots-clés.....	xiii
Introduction	17
1.1 Protein structure	17
1.1.1 Protein Structure and environment	18
1.1.2 Protein structure through the lens of NMR.....	20
1.2 Protein dynamics.....	24
1.2.1 Description of hierarchical protein dynamics at different scales using a biophysical approach	25
1.2.2 approach	25
1.2.3 Specificity of MAS-NMR to study the dynamics of proteins.....	30
1.3 References	37
Chapter 2 Analysis of the motional model and application to the model protein GB1	43
2.1 Relaxation and dynamics.....	44
2.1.1 Theoretical framework	44
2.1.2 Model free approach for temperature dependent measurements.....	47
2.2 Reproducibility and self consistency of the model	59
2.2.1 Method	59
2.2.2 Confidence intervals.....	62
2.2.3 Self-consistency of the model.....	66
2.3 Findings for the GB1 protein.....	69
2.3.1 Results	69
2.3.2 Conclusion.....	75

2.4	References	77
Chapter 3	Similarities and differences among protein dynamics	80
3.1	Introduction	80
3.2	Experimental.....	81
3.2.1	Samples.....	81
3.2.2	NMR experiments, measurements and activation energies:	87
3.3	Results	87
3.3.1	Relaxation Rates description	87
3.3.2	Activation Energies.....	99
3.4	Conclusions	103
3.5	References	104
Chapter 4	NMR analysis of lipid content and division of labour in ants	107
4.1	Ant biology.....	108
4.2	Ant society	110
4.3	Non-lethal corpulence and social network measurement	111
4.3.1	Lipid reserve measurements.....	111
4.3.2	Metabolic correlation	116
4.3.3	Tracking system.....	118
4.3.4	Behavioural/metabolic correlation	120
4.4	References	121
	Curriculum Vitae	123

Introduction

Proteins were shown to be an important part of biological life as soon as they were discovered, almost two centuries ago, and since then their role in virtually all living processes has been shown to be crucial.

The discovery in 1902 that proteins were formed by discrete ordered assemblies of amino acids was the first step in order to understand biological phenomena from the molecular point of view, a task that is far from being finished today. This description went forward in the early fifties with the proof that proteins were generally formed by a single chain of amino-acid residues.¹ This paved the way for an atomic-level approach to understanding the structure and the function of these biomolecules.

The first biochemical descriptions were focused on the composition of specific isolates of protein, but it was not until the advent of modern analytical techniques allowing us to probe the structure of proteins that our understanding of their mechanisms of action really made a gigantic step forward.²

From there, we will briefly mention the techniques and discoveries that paved the way to this study.

1.1 Protein structure

Proteins are identified by their chemical composition, i.e. large heteropolymers of amino acid residues. These large poly-peptides are essential for virtually all biological events. In order to understand the complex relationship between the ordering of amino acids and the almost infinite different functions that can be observed in nature, a description of the different scales used to observe and characterize the protein is essential. And even if not all information about a protein arises from the structure alone, this does often give great insight into the mechanisms of action of the protein. For example, structure of enzymes with different substrates or inhibitors can inform the enzymatic reaction mechanism taking place.

1.1.1 Protein Structure and environment

One of the first ways of describing proteins is to analyze each bond separately. The geometries that are allowed by the steric effects allow to constrain the relative orientations of the amino acids, giving a first description of the structures of proteins.³ This theoretical model is still used today, for example when numbering the Ramachandran outliers in a structure, as presented in Figure 1.

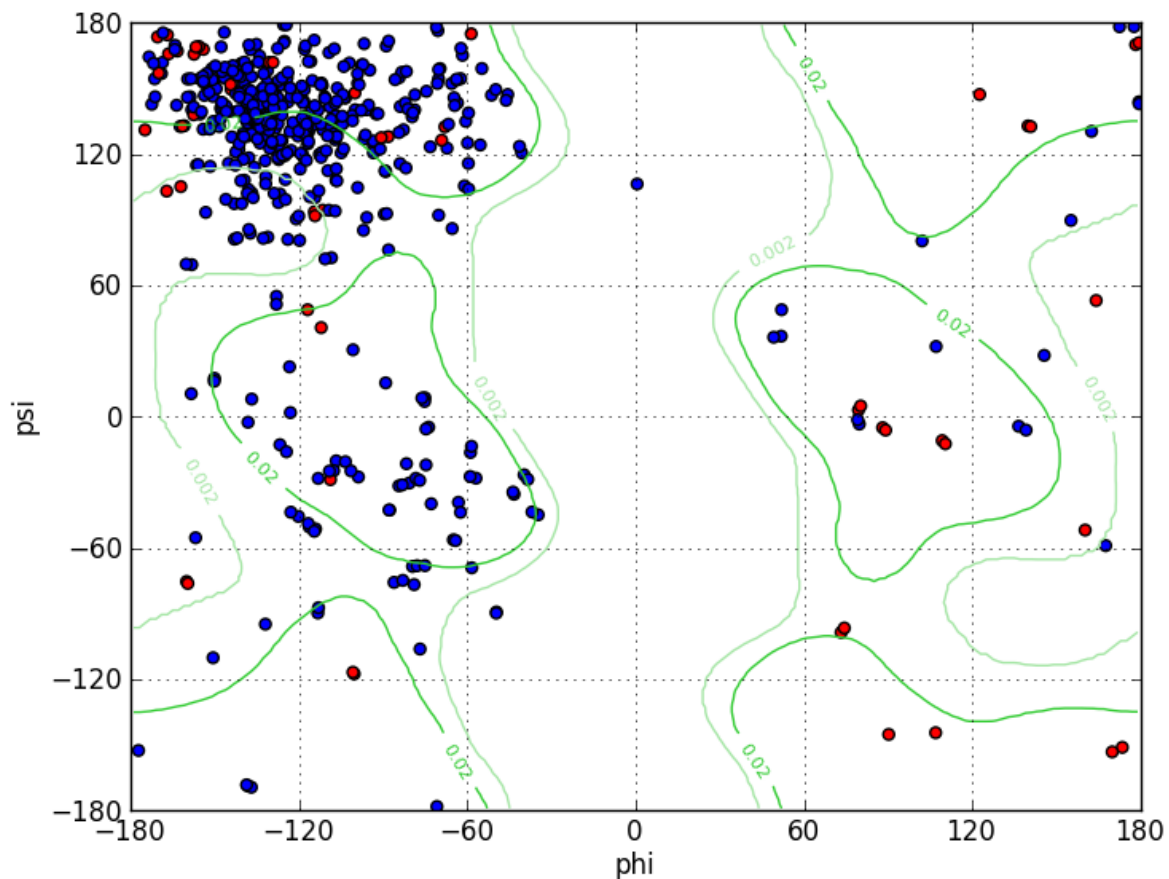


Figure 1: Representation of the angle phi ($N-C_{\alpha}$) et psi ($C_{\alpha}-C$) of the amino acids of the protein OmpG that will be studied in the chapter 3. The red dots represent the glycines and the green lines the probability contours for glycines observed in a large sample of proteins.

Single crystal X-ray diffraction is the leading technique when it comes to obtaining protein structures today. Most of the structures in the Protein data bank (PDB, rcsb.org) have been determined using X-ray diffraction methods (84% of protein structure described in 2019 in the structure databank PDB).

To illustrate the power of single crystal X-ray diffraction, one example can be the ion channels that are a key factor for cell homeostasis, signaling and many other biological processes. Snapshots of the channels in different conformations, or with different ligands, allow us to identify both the function

and the regulation of these proteins, as long as a crystalline sample can be obtained.⁴ Figure 2 shows an example of a 1.2 Å resolution structure of the immunoglobulin-binding B1 domain of protein G (GB1) mutant. This protein has been extensively studied and will be used as a model in the next chapter.

Numerous other methods such as Cryo-Electron Microscopy (cryo-EM)⁵⁻⁷ (13% of PDB structures in 2019), Fluorescence Resonance Energy Transfer (FRET)^{8,9} or mass spectroscopy are also incorporated in order to have a clearer picture of the three-dimensional atomic-level structure of proteins.

It is also possible to use other more indirect methods to infer structure activity relations, one example being analyzing evolution to determine contacts in proteins, using protein sequence coevolution. The concept is based on mutations that will compensate each other. For example, a core phenylalanine that is stabilized by a neighboring aliphatic side chain: if the phenylalanine changes to a polar residue due to a random mutation, then the protein can experience a loss of function due to drastic changes in structure. Gain of function can be achieved if the neighboring residues mutate to polar ones. Therefore, amino acids can change the evolutionary trajectories of their contacts in the protein. Statistical approaches using evolutionary models allow to determine such contacts, since those residues will be evolutionarily coupled.

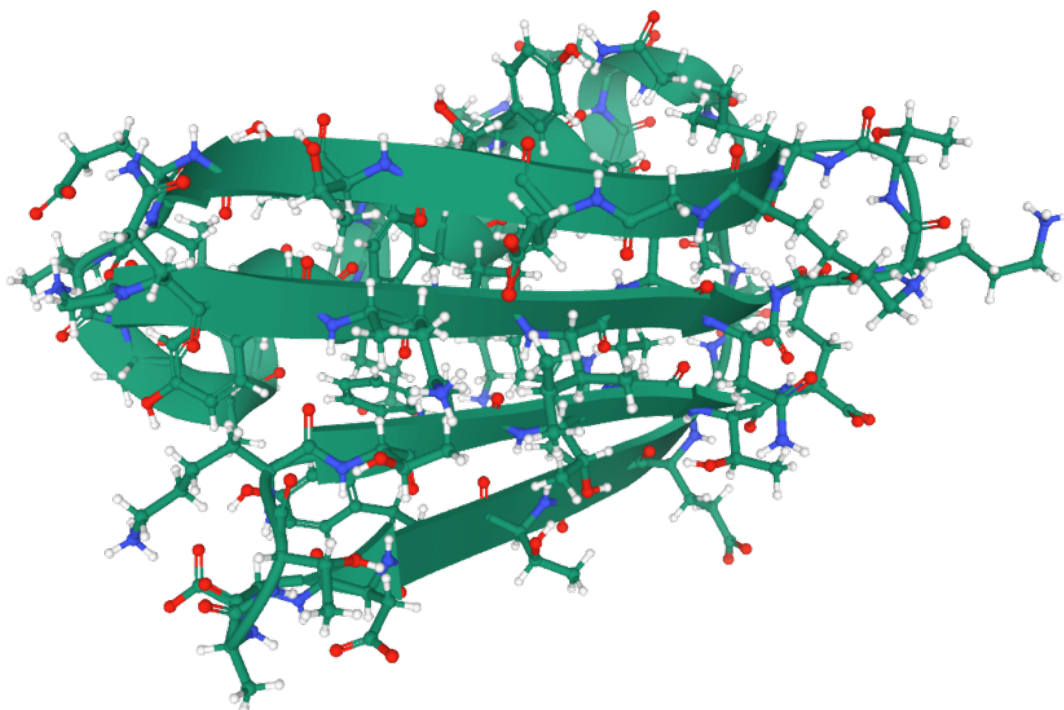


Figure 2 : Adapted from¹⁰. Crystal structure of a GB1 mutant destined to probe the ⁷⁷Se reaction to different chemical environments, as determined by single crystal X-ray diffraction. Image from RCSB PDB with Mol*¹¹

We will not go further into these other techniques, and will just highlight the importance of environment in protein structure, from ligand to hydration level.^{4, 12} The range and complexity of protein structure also opened the Pandora's box of protein folding, which will not be discussed in our analysis. Instead, we here focus on dynamics of proteins, in order to describe their flexibility.

1.1.2 Protein structure through the lens of NMR

3% of the structures in the PDB (2019) have been determined by NMR spectroscopy. Solution state NMR as well as Magic Angle Spinning (MAS) NMR provide information about the atomic-level environment of nuclei. This local information can then be extrapolated to determine protein structures. The principles are explained below to illustrate the different information that is accessible through this method.

1.1.2.1 Basic NMR for studying proteins

The principle of NMR spectroscopy was described in 1938 by Rabi¹³, and demonstrated for condensed matter in 1946 by Bloch (liquids¹⁴) and Purcell (solids¹⁵). They were awarded the Nobel prize in physics for their discoveries. Here we will only talk about atoms with a spin $I = \frac{1}{2}$, since in this work the only nuclei that will be studied are ^{13}C , ^{15}N , and ^1H , which all have spin $I = \frac{1}{2}$.

To analyze those properties we use a quantum description of NMR with the Hamiltonian for a typical system of nuclear spin $I = \frac{1}{2}$ of the form:

$$H = H_z + H_{CS} + H_J + H_D \quad \text{Eq I 1}$$

With the H_z term describing the Zeeman interaction, the CS term describing the chemical shift, J the scalar coupling, and D the dipole-dipole coupling.¹⁶

For a spin I , in the presence of a strong magnetic field the energy levels of the spin states that are otherwise degenerate will be split into $2I + 1$ levels, so two states in our case. The low energy state has its component of magnetic moment aligned with the magnetic field while the high energy one has

its magnetic moment opposed to the magnetic field if γ is positive, and vice versa. This Zeeman effect is by far the strongest in NMR of diamagnetic spin $I = \frac{1}{2}$, and is at the basis of all our experiments. As for any spectroscopy, we will measure transitions between energy levels. For an isolated spin $\frac{1}{2}$ the energy levels accessible are $\pm \frac{1}{2} \hbar \gamma B_0$ with \hbar the reduced Planck constant, γ the gyromagnetic ratio of the nucleus and B_0 the strength of the magnetic field.

Therefore, the spin of such a nucleus would resonate with a frequency

$$\omega_0 = -\gamma B_0. \quad \text{Eq I 2}$$

However, more information than simply the nucleus type is accessible. The local electronic environment will influence the perceived field at the position of the nucleus. If the shielding tensor is σ the frequency will now be

$$\omega_0 = -\gamma(1 - \sigma)B_0. \quad \text{Eq I 3}$$

with the shielding constant being a consequence of neighboring electrons decreasing the field B_0 at the nucleus.

The nucleus with spin I has a magnetic moment μ_I according to

$$\mu = \hbar \gamma I \quad \text{Eq I 4}$$

Other nuclei that are close in space with a spin different from 0 will give rise to magnetic dipole-dipole interaction, called dipolar coupling in NMR.

Since the chemical shift and dipolar interactions are anisotropic, in solution, the rapid rotational tumbling will average these two interactions. In the case of the dipolar coupling it is averaged to zero. In the case of the chemical shift it is averaged to a single isotropic value.

NMR can also be done on solids. Spectra of static samples can be informative about the CSA tensors, but most solid-state NMR experiments today are made with magic-angle-spinning (MAS).

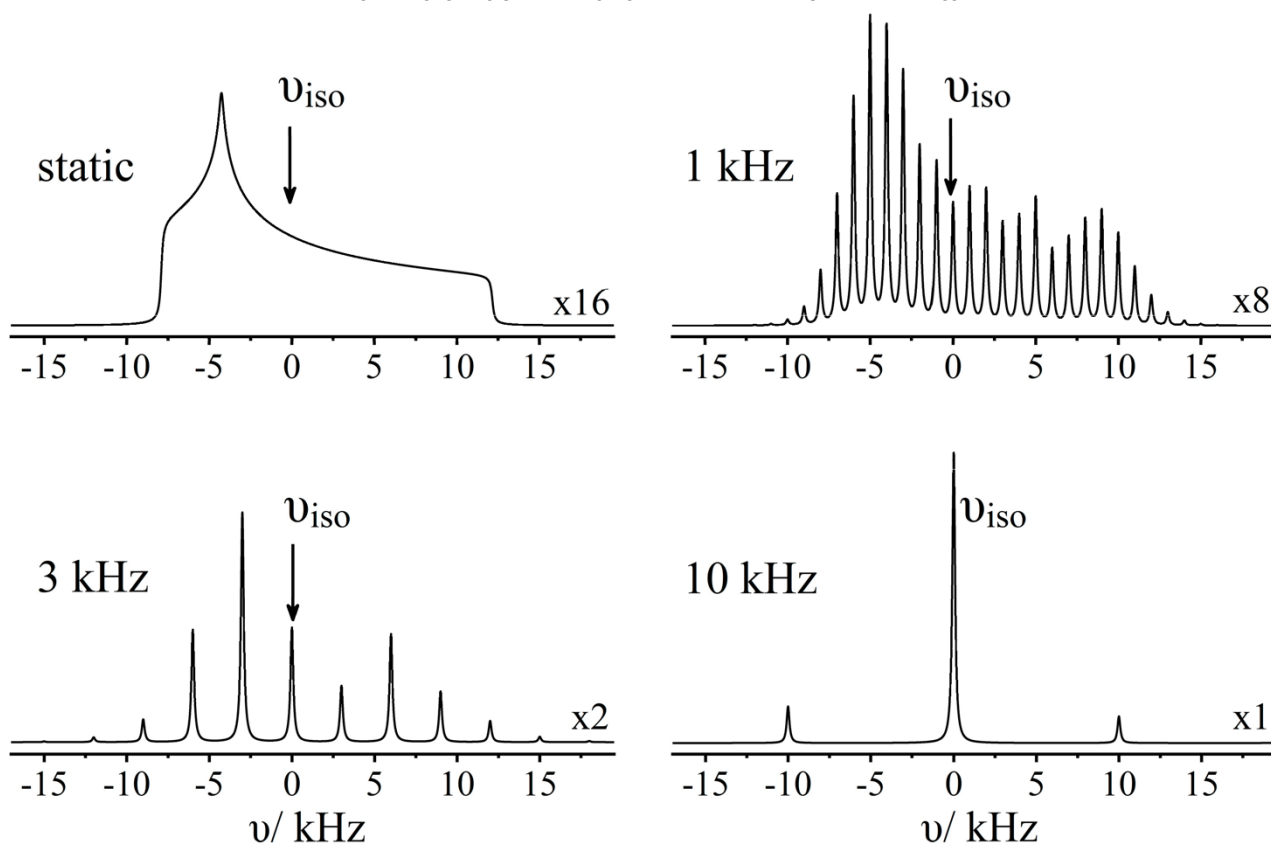


Figure 3: reproduced from ¹⁷. Simulation of the effect of MAS on the NMR signal of a $\frac{1}{2}$ spin nucleus with a broadening in static condition due to a large chemical shift anisotropy.

In MAS NMR, the interactions are averaged by a rapid physical rotation of the sample around an angle $\theta = 54.7^\circ$ (i.e. $3\cos^2\theta = 1$) with respect to the direction of the main magnetic field B_0 . This will average the second rank anisotropic part of the interactions to zero, leaving only the isotropic part of the interaction.

Other directly bonded nuclei will influence the frequency of the observed spin. The J-coupling is due to spin-spin interaction (both homo- and heteronuclear) where the up or down states of spins influence their neighbors. This will induce further splitting between the energy levels. In our experiments this will not be of great interest since at low temperature the broadening obscures this splitting, which is typically in the order of magnitude of 10-100 Hertz in proteins for ^1H , ^{13}C and ^{15}N . Furthermore, it does not contribute significantly to relaxation, which is the main observable that we will consider.

1.1.2.2 Structure of proteins using solution state NMR

As presented in the previous subsection, solution state NMR allows to discriminate between the different chemical environments of nuclei, and can therefore be used to gather information about the structure of molecules. This is also true for proteins, that can be studied in solution as long as their molecular weight does not exceed roughly 30 kDa. Above this range the molecular tumbling becomes too slow, leading to short T_2 , and the signals typically become too broad to be informative.

One of the major milestones that allowed determination of protein structure is the use of 2D spectroscopy, popularised by Richard Ernst and his team (discovered by Jean Jeneer), for which he was awarded the Nobel prize in 1991. This possibility to observe correlation signals between different nuclei on two (or more) axes simultaneously allows a much more precise description than single dimension experiments, due to the much higher resolution obtained.

This approach was used to fully take advantage of the nuclear Overhauser effect (NOE), which describes the transfer of polarization from one nucleus to another through the dipolar coupling. The very first protein structure was obtained on this basis using two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) through NMR in 1985 by Wüthrich and his team, for which he received the Nobel prize in 2002.¹⁸ Proton-proton distances extracted from NOE provided constraints for computational modelling, and allowed the research team to determine the atomic-level structure of the proteinase inhibitor IIA (a bovine pancreatic trypsin inhibitor) in solution.

A plethora of methods are today accessible to NMR spectroscopists to study protein structure in solution. With one clear advantage over single crystal X-ray diffraction, which still remains today the method of choice: proteins do not need to be crystalized to be observed, and therefore proteins with unstructured regions or that do not crystalize can be studied. Recently structures have also been determined for proteins inside cells.¹⁹

The first step in observing protein structure through NMR is the assignment of resonance in 2D (or up to 5D)²⁰ NMR spectra, where the backbone signals are assigned to the cross peaks obtained in the spectrum. Then in order to obtain the constraints necessary to model the protein, several methods can be used. As described above, NOESY spectra will give an indication of proximity for spatially close protons. It is important to note that these distance measurements are usually more qualitative than quantitative, since they do not only depend on the distance but also on the correlation time of the dipole-dipole interaction due to rotational tumbling. Furthermore, spin diffusion or other relaxation mechanisms can hinder the precision of the measurement. This potential imprecision is however not

a major problem in this approach because it is compensated for by the very large number of constraints obtained.

Using restraints is the basis of NMR structure determination. These can be distance restraints extracted from NOESY experiments, as described above, or angle restraints (the backbone ϕ and ψ angle) that can be extracted from coupling measurements or chemical shifts.²¹⁻²³ Orientational restraints can be obtained using residual dipolar splittings in anisotropic solutions. Paramagnetic relaxation enhancement (PRE) based on the change of T_1 or T_2 of the nuclei close to a stable radical or metal center is also very informative. This allows to probe long distances (15-25 Å) compared to the short (<6 Å) distances available through NOE experiments.²⁴

The constraints will then allow the creation of an ensemble of structures that are in agreement with the data. The ensemble can then be validated (or not) using statistical models such as WHAT IF (<https://swift.cmbi.umcn.nl/whatif/>) for example.²⁵ Today NMR structures of proteins typically can have a resolution on the order of 0.8/1 Å.⁹⁸

1.2 Protein dynamics

The ability to describe not only the sequence but also the three-dimensional structure of proteins is a key step in describing their functions. Nonetheless this description, even though it is essential to understand the biochemical reactions happening in living organisms, is not sufficient to fully describe them.

The behavior of proteins, and especially their numerous interactions, can only be understood as highly dynamic processes. From single ligand binding to multimeric cooperativity, all these processes require a description of a fourth dimension: time.²⁷ Dynamics in proteins are extremely diverse, since they can be driven by chemical energy or not, be pH or temperature dependent, require or be inhibited by small molecules and ions, etc...²⁸⁻³¹. This is why most studies focus on specific motions, or specific probes, in specific proteins, in order to simplify the problem. There is an extensive literature about the use of different techniques to characterize protein motions.^{30, 32-41} These approaches allow us to gain information about the specific motions of the studied proteins, but also allow the emergence of the set of underlying rules describing those motions.

The focus on the underlying properties of protein dynamics has obviously been widely studied. Mössbauer spectroscopy or differential scanning calorimetry for example can be used to identify motions of a system as a function of temperature, and therefore can discriminate the different types of motion in proteins.⁴²⁻⁴⁵

A description of the thermodynamic landscape of proteins would allow us to perfectly predict their dynamics. This is one of the core objectives of the previously cited methods. This approach has been proven to be successful in cases using predictions from molecular dynamics simulations where the thermodynamic landscape regarding specific conformations is computed.⁴⁶⁻⁴⁹ This powerful and rapidly evolving technique suffers nonetheless from some limitations. Long time scales (more than a millisecond) and large ensembles (like the ribosome) are impossible to compute efficiently. Moreover, the force fields that have been developed are still not a ‘perfect’ description of the real forces and lead to inaccuracies.⁵⁰ Determining the landscape experimentally would obviously be much more satisfactory.

1.2.1 Description of hierarchical protein dynamics at different scales using a biophysical approach

1.2.2 approach

As mentioned above, despite large differences in function and structure between different proteins, the dynamic properties of proteins are thought to share common features. The so-called universal dynamic properties include “transitions” in global dynamics as a function of temperature, often ascribed to be due to interactions between protein components and solvent⁵¹.

Observing these universal properties has been a longstanding focus, and several methods have been used as probes, including neutron scattering,^{52, 53} Mössbauer spectroscopy^{51, 54}, Terahertz spectroscopy⁵⁵, dielectric spectroscopy,⁵¹ DSC⁵⁶ and molecular dynamics simulations^{57, 58}. However, this has still not led to a consensus picture of the dynamic transitions and their origins, since different methods detect different, often contradictory, processes^{59, 60}. This situation may be in part due to varying sample conditions for the studies, which were required to isolate contributions from different processes, but since the timescales of internal dynamics of proteins spans several orders of magnitude,

it is intrinsically a difficult subject to study since most techniques are only sensitive to a restricted timescale.^{61, 62}

Recently, Lewandowski et al. showed how multiple NMR observables can be used to simultaneously identify dynamics occurring on different timescales and in different parts of the system by measuring relaxation rates over temperatures from 105 to 280 K in crystalline GB1.⁶³ The analysis of these multiple probes, reporting unambiguously on the different structural components of this complex system, allowed them to determine, from a single sample, a complete and coherent picture of the dynamic processes across the whole temperature range. It also allowed them to rationalize observations from other techniques. Their findings supported strong coupling between protein and solvent dynamics above 160 K, with motions in the solvent, then the protein side chains, and then the protein backbone being activated sequentially as temperature increases.

The key to generating this complete picture was the use of different ^1H , ^{13}C and ^{15}N relaxation rates⁶⁴ to distinguish between motional modes in a type-specific manner, affecting bulk solvent, hydration water, protein side chains and the backbone (see Figure 4). In the following two chapters we will build on this approach.

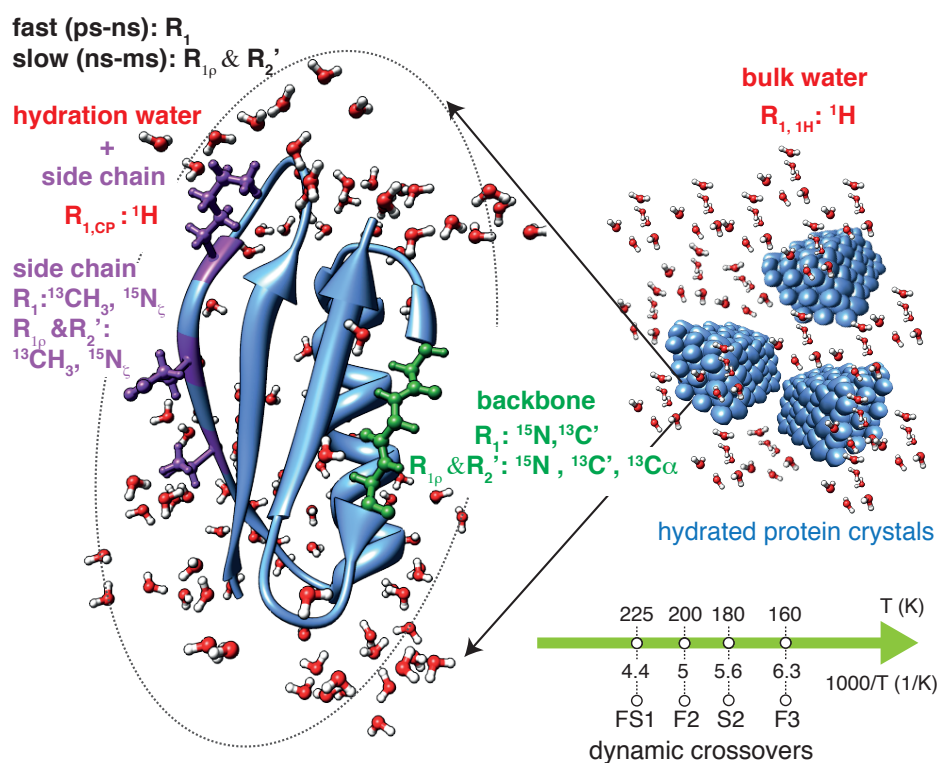


Figure 4: reproduced from ⁶³. Location of the motions and the relaxation rates that are most sensitive to those motions, together with an illustration of the dynamic transitions identified by the change in the

relaxation trends (green arrow in the bottom right corner). The rates indicated in green, purple and red font report respectively on backbone, side chain and solvent motions. The numbers at the top and the bottom of the arrow indicate respectively approximate transition temperatures (units of K) and inverse temperature (units of 10^{-3} 1/K). Labels FS1, F2, S2 and F3 indicate the transitions. Those transition represent the temperatures at which the motions driving relaxations change for a given observable.

In order to further set the scene, in this section we will review some of the methods used for the study of protein dynamics, as well as their interactions with water.⁵⁰ We will describe the consensus with a large range of different techniques in order to be able to test and compare the results obtained with NMR, to avoid mistaking technical biases for meaningful information about proteins.

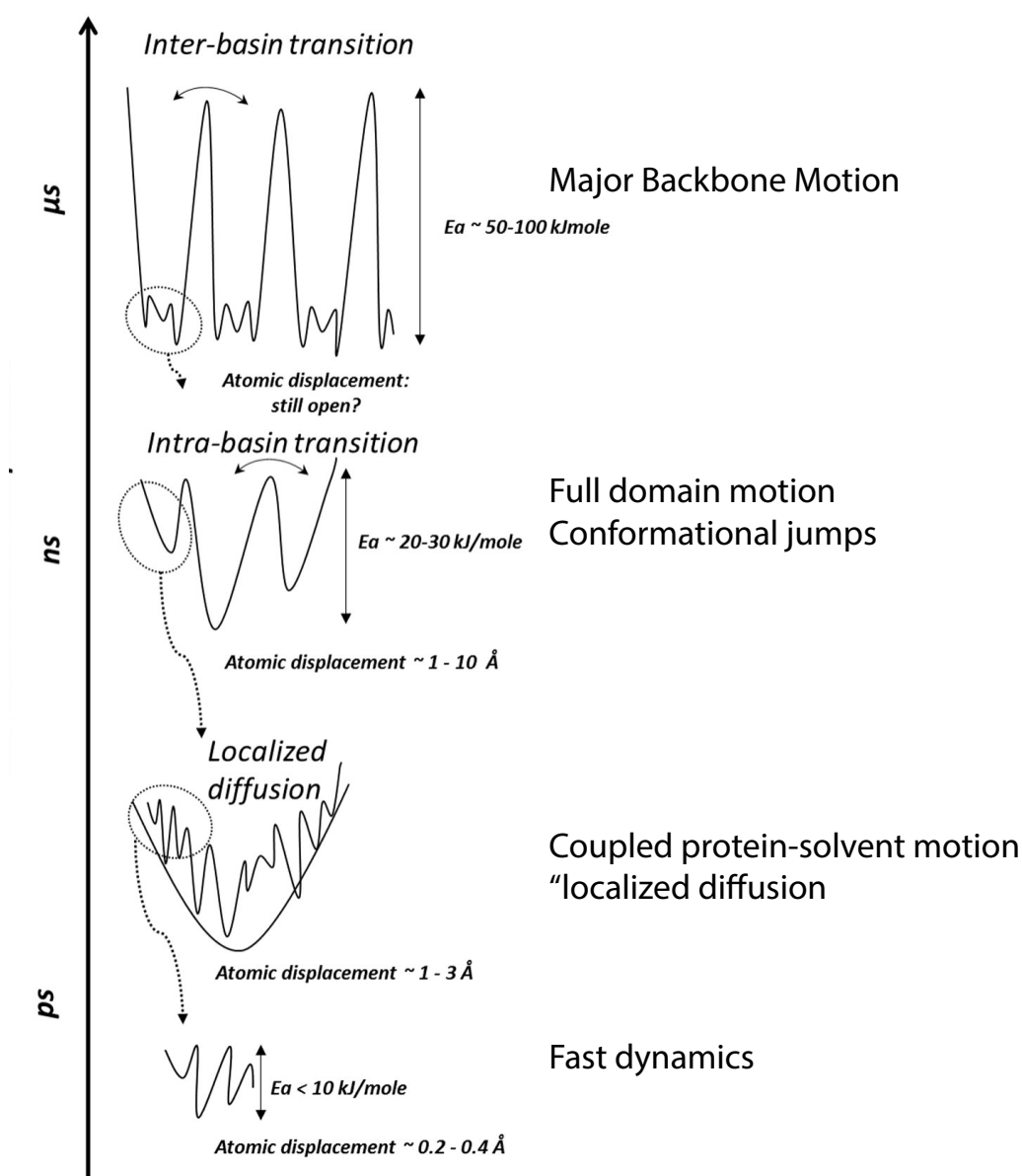


Figure 5: Adapted from From⁶². The vertical axis is the typical timescale at room temperature. The first panel represents inter basin transitions that are due to large structural changes. The second panel represents intra

basin transitions which are mostly attributed to side chain/domain motions. The third panel represents the motions of protein and solvent that are coupled together and are analogous to a diffusion in a potential well, aka localized diffusion. The last panel represents fast dynamics which typically have a low energy of activation and small amplitude.

Figure 5 displays an overview of the hierarchical motions present in internal protein dynamics. Fast dynamics are mostly due to vibrations and are of small amplitude and displacement. Localized diffusion refers to the coupled motion in both protein and solvent. Conformational jumps represent the exchange between different conformers, mostly from side chains in protein. Major backbone motions display the highest energies of activation and are due to torsion of the backbone angles.

Fast dynamics: Neutron scattering can extract information from both local and collective motions. As seen by Neutron scattering, fast (ps at room temperature) motions have a small amplitude (<1 Å) and an energy of activation of 2-5 kJ/mol. Collective vibrations on that timescale are also observed. The side chains have extremely heterogeneous motional timescale at RT, from 10 s of ps to μ s with energies of activation of 10-20 kJ/mol. Water has a strong impact on fast protein dynamics. The hydration of a protein will increase the amplitude of picosecond motions in a protein at high temperature, and decrease it at low temperature.⁶⁵

Coupled protein-solvent motions: These motions analogous to a localized diffusion are also very heterogeneous. It is observable notably by terahertz spectroscopy on the whole protein, even in parts not directly in contact with the solvent. The typical amplitude is estimated to be 1-3 Å, and the typical timescale are 10-50 ps at room temperature. This specific motion seems to be necessary for protein activity since those motions and protein activity are strongly correlated.^{52, 53} The correlation between protein activity and 1-3 Å 10-50 ps motions is probably due to localized diffusion enabling other motions, since solvent dynamics is typically much faster than the other motions in a protein. It is important to note that in this picture there is no discrete jump between different conformers, but rather continuous diffusion in a constrained position.⁶⁵⁻⁶⁷ Water-protein coupling has therefore been postulated to be well suited for analysis by models including friction-like parameters, which could lead to new types of models suitable for long timescale MD while still retaining accuracy.⁶² Hydration water behaves differently for intrinsically disordered proteins (IDP) compared to membrane proteins as reported by neutron scattering experiments.⁶⁸ IDPs present stronger coupling between the protein and water compared to structured proteins. For all proteins there is nonetheless a broad distribution of couplings depending on the specific surface interactions.

Side-chain motions: Conformational jumps of large scale observed with neutron scattering and dielectric spectroscopy (1-10 Å) display energies of activation in the 20-40 kJ/mol range and have a timescale of ns at room temperature.^{62, 69} Neutron spin echo spectroscopy can be used to probe those motions that are of high importance for proteins that rely on large structural changes for their activity. For example, the catalytic activity of a kinase that relies on 50 ns timescale large amplitude motions.⁷⁰

Methyl group rotations: In such dynamics, methyl rotation is generally included. Both MD simulation and neutron scattering can characterize methyl rotation.⁶⁷ One interesting feature is that the methyl group rotation due to its hydrophobicity does not seem to be related to the solvent,⁶⁶ in that it does not depend on the presence (or absence) of hydration water around the protein. We notice that the fastest methyl groups in both lysozyme and myoglobin are near their active sites.^{71, 72} All other protein dynamics except some specific hydrophobic motions depend on the solvent.⁶²

Backbone motions: By analyzing with precision different parts of the peptide plane with NMR, it is possible to further identify the relative dynamics of protein backbone.⁷³ The current slow limit for neutron scattering is represented by timescales higher than hundreds of ns. Similarly, most MD methods will be in the same timescale, which makes the study of slow dynamics difficult in biomolecules. Nonetheless, torsional rotations happening in the microsecond timescale at room temperature in backbones have been identified, using MAS NMR mostly, since having information about microsecond timescales has proven to be a challenge in proteins.⁷⁴ Backbone motions could be related to the glass transition in protein, since they show a drastic change of behavior at T_g , the transition temperature.⁴⁴

Processes with even longer timescales can be analysed by other methods.^{8, 75 66} FRET methods are particularly useful for in cell characterization since the probes do not require a preparation as drastic as many other techniques, and can be done in vivo.

The influence of the solvent: It is speculated that hydration water acts as an enabler of protein dynamics, and that proteins will displace neighboring water molecules as they move, which is in itself the definition of a coupling.⁶² Surprisingly larger scale motions are thought to have an even larger hydration dependence than small/fast ones, even if they are not directly coupled to the solvent.

Water can be either trapped inside the protein or adsorbed to the protein surface. The trapped water is generally very stable.⁶⁴ Differences between bulk and hydration water have been highlighted by

dielectric relaxation measurements, showing that protein will prevent diffusion of the hydration water.

⁷⁶ Some hydrogen bonds in the interfacial water can be 1000 times longer lived than in bulk.⁶⁵

According to tryptophane photoexcitation experiments, which possess a ps resolution, after a distance larger than 7 Å from the protein, there is no influence detectable of the protein, and therefore no more hydration shell.⁷⁷ The influence of the protein on water is not uniform and depends on the local interactions. The hydration shell is still very dynamic, but less than the bulk water.⁶⁴ Hydration water stays dynamic at very low temperature, and will not crystallise but instead forms a glass.⁵⁹

Dynamics in presence of other biological solvents has also been studied, with neutron scattering among other methods.⁷⁸ With various results such as glycerol protecting the proteins from cold and trehalose from the heat. Liquid-liquid phase separation is also observed in vivo in the case of several membrane-less organelles.⁷⁹

1.2.3 Specificity of MAS-NMR to study the dynamics of proteins

To have information about the thermodynamic landscape we need to have an probe of motion. The use of relaxometry in NMR is very well suited for this purpose.

To use relaxometry a semi classical derivation of the master equation (a matrix or set of equations describing the time evolution of a system), was established by Wangness, Bloch and Redfield.^{14, 80} For different models of the dynamics involved a spectral density function can be estimated. In the case of the model-free approach,^{81, 82} which assumes that a motion can be described with two parameters, the rate of dynamics and an order parameter describing the spatial restriction of the motion, by assuming a monoexponential decay for the correlation function the spectral density function is

$$J(\omega) = (1 - S^2) \frac{5}{2} \frac{\tau_c}{1 + (\omega\tau_c)^2} \quad \text{Eq I5}$$

with the two parameters, τ_c characterizing the timescale of the motion, and S^2 the amplitude of the motion. The order parameter S will be 0 if the motion is isotropic (no spatial restriction) and 1 if there is no motion. The dependence of S with the spatial restriction of its specific motion has been previously examined.⁸²

Longitudinal relaxation (R_1 , the return of magnetisation to equilibrium along the z axis) of individual nuclei can easily be observed, and is the observable of choice.⁸³ Note that spin diffusion can homogenize the properties of neighboring atoms.⁸⁴ This is especially true at low temperature, and can be attenuated by high spinning speeds.⁸⁵ Longitudinal relaxation is most sensitive to motion on ns timescales.

Heteronuclear cross relaxation (NOE) is due to transfer of polarization between different nuclei. The information extracted from NOEs have a high redundancy with T_1 ($R_1 = 1/T_1$) parameters.⁸⁶

Transverse relaxation ($R_2 = 1/T_2$) parameters are sensitive to motion occurring with a longer timescale.⁸³ Unfortunately, anisotropic interaction being incompletely averaged out leads to difficulties in quantitative interpretation of those data, and the observable in solids is usually referred to as T_2' .⁸³ Moreover, the Redfield approximation might not hold in the slow motion regime. A spin lock field can also be used to measure transverse relaxation ($T_{1\rho}$). Generally different lock amplitudes are used to better analyse the motion.⁸⁷ Cross relaxation, which is due to interferences between different interaction is also used.⁷³ Cross relaxation is mainly due to dipolar and CSA interaction, or between two different CSA interaction.⁸⁷ Finally it is also possible to probe the relaxation dispersion using CPMG RD experiments.³⁸

In principle MAS NMR allows to study virtually any protein without the restriction of crystallization. Other preparation methods such as sedimentation or incorporation into lipids bilayers can be used to obtain a solid sample for MAS studies.

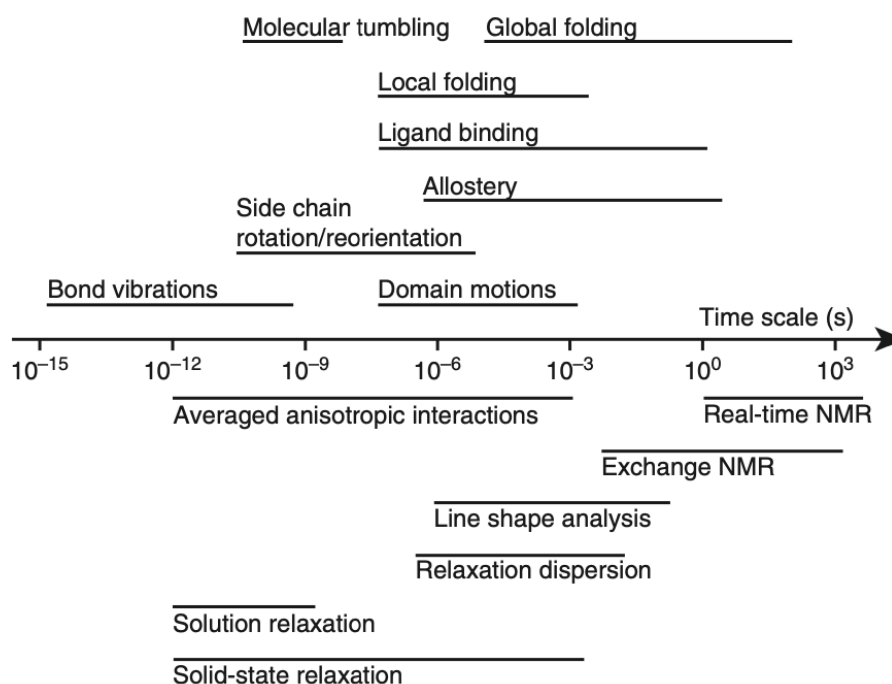


Figure 6: From⁸⁸ top panel represents the different timescales at room temperature for motions in protein, bottom panel, the experiments sensitivity to different types dynamics.

The relaxation parameters are thus highly sensitive to motions over a very wide range of timescales, and the equations describing the relation between motion and relaxation have been extensively studied and validated.⁸⁸⁻⁹⁰ Solid state NMR has also an advantage compared to the more widely used solution state NMR, which is the possibility to analyze samples at a much lower temperature (i.e. frozen), since there is no reliance on the tumbling of molecules. This comes at a price of lower resolution than solution state NMR. It is therefore possible to analyse the motions over a large range of temperatures, thus allowing us to probe both the low-energy and high-energy motions directly, that could otherwise be masking each other at room temperature as described in Figure 5.^{61, 63} Due to the site-specific resolution, NMR can be used to extract local information.

MD simulation and NMR can be used conjointly to study proteins (or nucleic acids)^{49, 86} In particular here using a comparison of different observables from MD could allow us to validate the motional models.^{63, 91}

The same conformational changes can be observed with NMR as well as with neutron scattering or dielectric spectroscopy. Motions with 1–10 Å of amplitude and an energy of activation in the range of 20–40 kJ/mol were observed. Those amplitude and energy of activation were confirmed by MD simulation.^{67, 69, 92}

Protein dynamics techniques were mostly first developed in solution NMR. These techniques have then been adapted to analysis of solid samples, in order to extract information about site specific motions including their timescale, amplitude and anisotropy.⁷⁴ For example analyzing ^{13}C relaxation data gives information on site specific side chain motions.⁹³ Even if resolution is sometimes a challenge in solid state NMR, site specific dynamics can be obtained, at least on room temperature samples, with an amino-acid specific estimation of the timescale and anisotropy.^{87, 94}

Deuterium wideline spectroscopy was also widely used to probe motion in biological systems in the early days of solid state NMR.⁸⁷

In order to comment on the precision of MAS NMR to probe dynamics, results have been compared to validate the concept and to observe the differences of the behavior of proteins in solid state (crystallized, sedimented...) and in solution. Studies compared protein side chain dynamics of small peptides by looking at ^2H methyl relaxation in solution and in solid state. The authors highlighted the similarity of motions between the two preparations, even if some details differ, due to different sample preparations and instruments. The similarities can be seen in Figure 7.⁹⁵ The dynamics in proteins are not believed to fundamentally change in crystals or solutions, as long as the hydration level is high enough. Most of the observed differences to the dynamics in solution and in crystals seem to be located in the loops that might be more constrained by crystal packing effects.⁸⁷ The solid (MAS) and solution comparison of SH3 also highlights the similarities of behavior in both environments. The pertinence of the model free analysis is also demonstrated for solid state in their approach, with monoexponential decay (thanks to an ingenious labelling scheme).³⁵

Rotational double echo resonance (REDOR-type) experiments allow to measure order parameters directly (describing the rigidity of the system) of specific residues (with ns and ps resolution, and an estimation of the amplitude of the associated motion), and again those results are in concordance with the results found in solution.⁹⁶ Due to the absence of molecular tumbling on the ns timescale, more dynamics are accessible with solid state NMR than with solution state. Different relaxation parameters, or lineshape analysis (specifically with quadrupoles like deuterium) allow to access the question of protein dynamics with different observables.^{87, 97}

There are specific motions and effects that should not be ignored in order to produce accurate results. A 10 μ s timescale rocking motion in protein crystals is observed at room temperature, it is specific to the crystalline phase and must be addressed in slow dynamics studies in solid state.^{98 99} This seems to have homogeneous timescale (30-50 μ s) even if the amplitude varies significantly.⁹⁰ Moreover Proton Driven Spin Diffusion (PDSD) and direct spin diffusion is strong (in protonated samples and/or for ^{13}C uniform labeling) especially at low spinning speeds and low temperatures, and will average the relaxation rates between spatially close nuclei.⁸⁸ ^{15}N $T_{1\rho}$ relaxation shows small amplitude motions (1 μ s timescale) across proteins. Larger amplitudes are found at specific sites, but the timescale seems uniform.¹⁰⁰ These results are in concordance with observation from other techniques.¹⁰¹ Like in solution, exchange between different states for a protein can be probed, even with very low populations. Moreover, the timescale of transitions and energy landscape describing the slow exchange can be obtained.¹⁰²

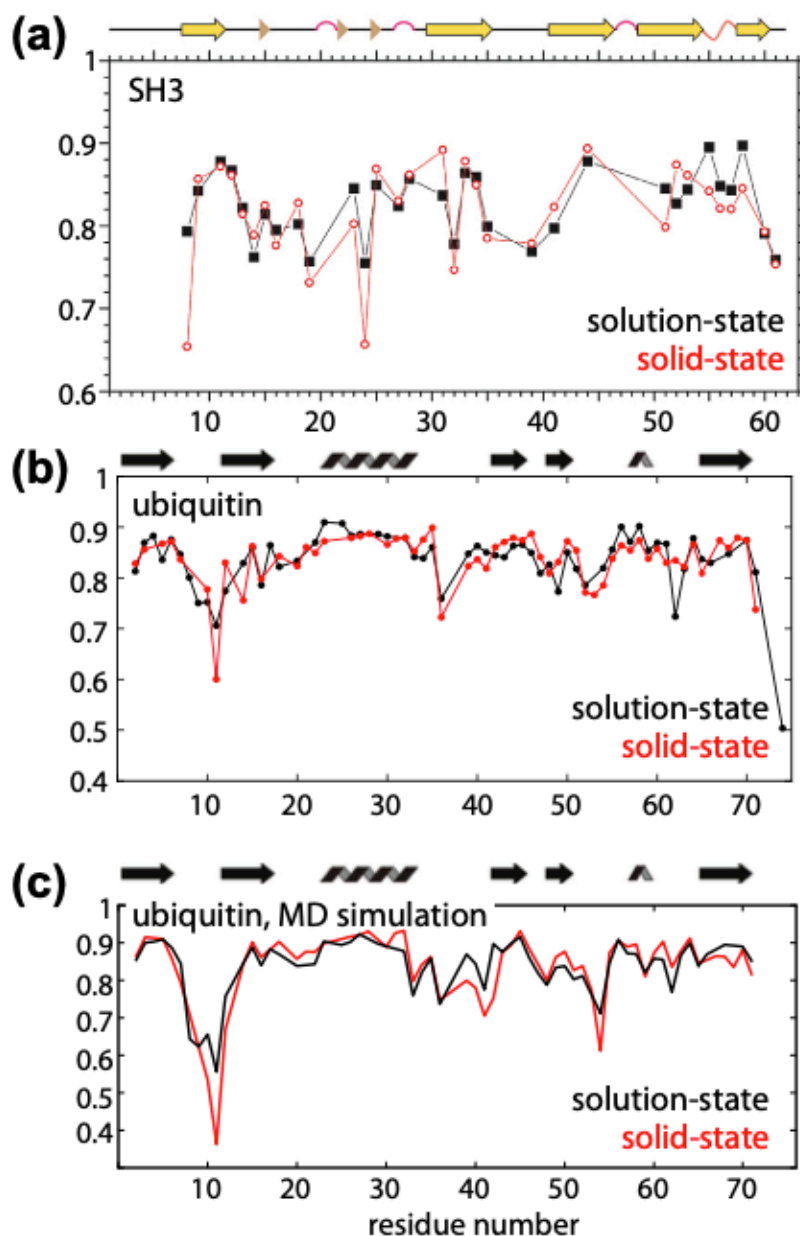


Figure 7: From^{74, 87} order parameters of backbone N-H both in solution and in solid state. SH3(a) and Ubiquitin (b,c). data in c) have been obtained by MD simulation. The general trends are the same in both conditions.

One of the advantages of MAS is the absence of molecular tumbling, which masks any motion with a timescale slower than the molecular tumbling itself (on the order of 10 s of ns for small globular proteins studied in solution). Therefore slow motions (ns-ms) can be determined with a site specific resolution using MAS NMR.¹⁰³ This example gives a quantitative description of the only 4 residues (for GB1) which exhibit significant motions in the 40 μ s to 0.1 ms range. More than just timescales, it is also possible to have access to the anisotropy of motions.

In MAS the dipolar coupling is not irreversibly averaged like in solution, and can be probed directly by MAS NMR. This therefore allows access to both the amplitude and the anisotropy of motions in a specific residue, if the structure is known.^{88, 104}

MAS NMR allows also to probe a larger temperature range and therefore to better describe the thermodynamic landscape. For example, spin lattice relaxation of ^2H at different fields has been analyzed to identify motional timescales, and their dependence on temperature.⁹³ At low T , (below 250 K) ^2H T_1 relaxation is almost field independent indicating that motions have a timescale faster than ns. Moreover, motions in the time range of 0.1-1 μs only induce ^2H relaxation above this temperature (250 K). This was described as a dynamic glass transition, but the term (not the observation) is refuted in the literature.^{59, 105, 106} Since this rapid shift in observed relaxation rates is due to a second order transition (motions being more or less efficient to induce relaxation as a function of temperature), we are here looking at a progressive transition and not an abrupt one.

The specificity of the thermodynamic landscape can be probed for different classes of proteins.

In fibrils, difficult to seen by solution state NMR, three different motional timescales can be observed at room temperature by measuring R_1 and $R_{1\rho}$ at different fields. These three types (different timescales) of motions can be detected with site specific resolution.¹⁰⁷

In membrane proteins there is evidence for two different timescales at room temperature with 10-100 ps local motions.⁹⁵ Those motions are more important at the contact with the lipidic phase, and slower collective motions with a timescale of 10-100 ns are more important near the water solvent.¹⁰⁸

Large assemblies and their interactions are also better suited to MAS NMR compared to solution, due to the limitation of the tumbling with the size. Large complexes, and protein-protein contacts and interaction can be probed by PREs in the solid state.¹⁰⁹ Internal motions of big assemblies can also be detected by probing the flexibility among large timescales, as has been described for the nucleosome.¹¹⁰

For MAS NMR, crystalline samples (or highly ordered samples) are still generally better than amorphous ones, which is well adapted for the study of big assemblies like virus capsids or fibrils.⁹⁹ Moreover extracting information from protons in solid state with a high precision is still hard due to poor resolution of most spectra in the proton dimension.¹¹¹

As described above, water plays a key role in protein dynamics,^{40, 62, 65, 67, 112} and protein dynamics occurs in a wide range of timescale and amplitudes, with the influence of water being critical.⁶² Moreover, it was shown that MD simulation taking in account explicitly the role of water were better when it comes to replication of NMR measurements.¹¹³ Water can be separated into water close to the protein, and bulk water, with different motions detectable by solid-state NMR. One should notice that in protonated water, the water seen to be in contact with the protein extends further than the hydration shell, and that signal from bulk water might be significant while probing hydration water.¹¹² Site specific water interactions can be probed, and those interactions highlight the role of the hydration shell to modulate protein dynamics.³⁰ Other solvents also have their importance, for example liquid-liquid phase separation is correlated to conformational dynamics, and work is done outside of “classically” hydrated proteins.¹¹⁴

Here, we build on this initial study to further develop our understanding of hierarchical dynamics in proteins.

In chapter 2 we will use the approach presented by Lewandowski et al.⁶³ and extend the experiment to relaxation at different fields. This allows us to validate the relaxation model used for a specific motion, and to better constrain better the physical description of the hierarchically organized motions.

In chapter 3 we will compare the behavior of different classes of proteins, soluble globular proteins, intrinsically disordered proteins (IDP) and membrane proteins, using multiple field variable temperature spin lattice relaxation measurements.

The chapter 4 presents a different subject (metabolomic studies of ants) and will be treated separately.

1.3 References

1. Sanger, F.; Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem J* **1951**, *49* (4), 463-81.
2. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958**, *181* (4610), 662-6.
3. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol* **1963**, *7* (1), 95-&.
4. Ye, S.; Li, Y.; Chen, L.; Jiang, Y., Crystal structures of a ligand-free MthK gating ring: insights into the ligand gating mechanism of K⁺ channels. *Cell* **2006**, *126* (6), 1161-73.
5. Ishida, H.; Matsumoto, A., Free-Energy Landscape of Reverse tRNA Translocation through the Ribosome Analyzed by Electron Microscopy Density Maps and Molecular Dynamics Simulations. *Plos One* **2014**, *9* (7).

6. Leelananda, S. P.; Lindert, S., Iterative Molecular Dynamics-Rosetta Membrane Protein Structure Refinement Guided by Cryo-EM Densities. *J Chem Theory Comput* **2017**, *13* (10), 5131-5145.
7. Parmar, M.; Rawson, S.; Scarff, C. A.; Goldman, A.; Dafforn, T. R.; Muench, S. P.; Postis, V. L. G., Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim Biophys Acta Biomembr* **2018**, *1860* (2), 378-383.
8. Koinuma, S.; Takeuchi, K.; Wada, N.; Nakamura, T., cAMP-induced activation of protein kinase A and p190B RhoGAP mediates down-regulation of TC10 activity at the plasma membrane and neurite outgrowth. *Genes Cells* **2017**, *22* (11), 953-967.
9. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat Methods* **2008**, *5* (6), 507-516.
10. Chen, Q. Q.; Xu, S. P.; Lu, X. Y.; Boeri, M. V.; Pepelyayeva, Y.; Diaz, E. L.; Soni, S. D.; Allaire, M.; Forstner, M. B.; Bahnson, B. J.; Rozovsky, S., Se-77 NMR Probes the Protein Environment of Selenomethionine. *J Phys Chem B* **2020**, *124* (4), 601-616.
11. Sehnal, D.; Rose, A. S.; Koča, J.; Burley, S. K.; Velankar, S., Mol*: towards a common library and tools for web molecular graphics. In *Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data*, Eurographics Association: Brno, Czech Republic, 2018; pp 29–33.
12. Veneziano, R.; Rossi, C.; Chenal, A.; Brenner, C.; Ladant, D.; Chopineau, J., Synthesis and characterization of tethered lipid assemblies for membrane protein reconstitution (Review). *Biointerphases* **2017**, *12* (4), 04E301.
13. Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P., A new method of measuring nuclear magnetic moment. *Phys Rev* **1938**, *53* (4), 318-318.
14. Wangsness, R. K.; Bloch, F., The Dynamical Theory of Nuclear Induction. *Phys Rev* **1953**, *89* (4), 728-739.
15. Pound, R. V.; Purcell, E. M., Measurement of Magnetic Resonance Absorption by Nuclear Moments in a Solid. *Phys Rev* **1946**, *69* (11-1), 681-681.
16. Keeler, J., *Understanding NMR Spectroscopy*. Wiley: 2011.
17. Knitsch, R.; Brinkkotter, M.; Wiegand, T.; Kehr, G.; Erker, G.; Hansen, M. R.; Eckert, H., Solid-State NMR Techniques for the Structural Characterization of Cyclic Aggregates Based on Borane-Phosphane Frustrated Lewis Pairs. *Molecules* **2020**, *25* (6).
18. Williamson, M. P.; Havel, T. F.; Wuthrich, K., Solution Conformation of Proteinase Inhibitor-lia from Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry. *J Mol Biol* **1985**, *182* (2), 295-315.
19. Ikeya, T.; Guntert, P.; Ito, Y., Protein Structure Determination in Living Cells. *International Journal of Molecular Sciences* **2019**, *20* (10).
20. Zerko, S.; Byrski, P.; Włodarczyk-Pruszyński, P.; Gorka, M.; Ledolter, K.; Masliah, E.; Konrat, R.; Kozminski, W., Five and four dimensional experiments for robust backbone resonance assignment of large intrinsically disordered proteins: application to Tau3x protein. *Journal of Biomolecular Nmr* **2016**, *65* (3-4), 193-203.
21. Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M., Protein structure determination from NMR chemical shifts. *P Natl Acad Sci USA* **2007**, *104* (23), 9615-9620.
22. Hudaky, P.; Perczel, A., Toward direct determination of conformations of protein building units from multidimensional NMR experiments VI. Chemical shift analysis of his to gain 3D structure and protonation state information. *J Comput Chem* **2005**, *26* (13), 1307-1317.
23. Sgourakis, N. G.; Lange, O. F.; DiMaio, F.; Andre, I.; Fitzkee, N. C.; Rossi, P.; Montelione, G. T.; Bax, A.; Baker, D., Determination of the Structures of Symmetric Protein Oligomers from NMR Chemical Shifts and Residual Dipolar Couplings. *Journal of the American Chemical Society* **2011**, *133* (16), 6288-6298.
24. Battiste, J. L.; Wagner, G., Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry-Us* **2000**, *39* (18), 5355-5365.
25. Vriend, G., What If - a Molecular Modeling and Drug Design Program. *J Mol Graphics* **1990**, *8* (1), 52-&.
26. Laulumaa, S.; Kursula, P., Sub-Atomic Resolution Crystal Structures Reveal Conserved Geometric Outliers at Functional Sites. *Molecules* **2019**, *24* (17).
27. Weber, G., Energetics of ligand binding to proteins. *Adv Protein Chem* **1975**, *29*, 1-83.
28. Tompa, K.; Banki, P.; Bokor, M.; Kamasa, P.; Gyorgy, L.; Tompa, P., Interfacial Water at Protein Surfaces: Wide-Line NMR and DSC Characterization of Hydration in Ubiquitin Solutions. *Biophys J* **2009**, *96* (7), 2789-2798.
29. Draganski, A. R.; Friedman, J. M.; Ludescher, R. D., Solvent-Slaved Dynamic Processes Observed by Tryptophan Phosphorescence of Human Serum Albumin. *Biophys J* **2017**, *112* (5), 881-891.
30. Fenwick, R. B.; Oyen, D.; Dyson, H. J.; Wright, P. E., Slow Dynamics of Tryptophan-Water Networks in Proteins. *Journal of the American Chemical Society* **2018**, *140* (2), 675-682.
31. Jansson, H.; Bergman, R.; Swenson, J., Role of Solvent for the Dynamics and the Glass Transition of Proteins. *The Journal of Physical Chemistry B* **2011**, *115* (14), 4099-4109.
32. Imbardelli, D.; Chidichimo, G., An Investigation of the Dynamics of Ch3i in a Nematic Mesophase by Nmr Linewidth Analysis. *Chem Phys Lett* **1987**, *134* (1), 96-101.

33. Imbardelli, D.; Chidichimo, G.; Longeri, M.; Bucci, P.; Saupe, A., Molecular-Dynamics Investigation by Measurement of Nmr Linewidth of Partially Oriented Molecules in Nematic Mesophases. *J Magn Reson* **1990**, *90* (3), 514-521.
34. Mittermaier, A.; Varani, L.; Muhandiram, D. R.; Kay, L. E.; Varani, G., Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J Mol Biol* **1999**, *294* (4), 967-79.
35. Agarwal, V.; Xue, Y.; Reif, B.; Skrynnikov, N. R., Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: a similarity revealed. *J Am Chem Soc* **2008**, *130* (49), 16611-21.
36. Choy, W. Y.; Shortle, D.; Kay, L. E., Side chain dynamics in unfolded protein states: an NMR based ²H spin relaxation study of Δ 131 Δ . *J Am Chem Soc* **2003**, *125* (7), 1748-58.
37. Daley, M. E.; Sykes, B. D., Characterization of threonine side chain dynamics in an antifreeze protein using natural abundance ¹³C NMR spectroscopy. *J Biomol NMR* **2004**, *29* (2), 139-50.
38. Esadze, A.; Li, D. W.; Wang, T.; Bruschweiler, R.; Iwahara, J., Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear ¹H-¹⁵N NMR spectroscopy. *J Am Chem Soc* **2011**, *133* (4), 909-19.
39. Kannan, S.; Zacharias, M., Role of tryptophan side chain dynamics on the Trp-cage mini-protein folding studied by molecular dynamics simulations. *Plos One* **2014**, *9* (2), e88383.
40. Qin, Y.; Zhang, L.; Wang, L.; Zhong, D., Observation of the Global Dynamic Collectivity of a Hydration Shell around Apomyoglobin. *J Phys Chem Lett* **2017**, *8* (6), 1124-1131.
41. Tugarinov, V.; Kay, L. E., Relaxation rates of degenerate ¹H transitions in methyl groups of proteins as reporters of side-chain dynamics. *J Am Chem Soc* **2006**, *128* (22), 7299-308.
42. Ekimoto, T.; Odagaki, T.; Yoshimori, A., Separation of dynamics in the free energy landscape. *Aip Conf Proc* **2008**, *982*, 211-214.
43. Idiyatullin, D.; Nesmelova, I.; Daragan, V. A.; Mayo, K. H., Heat capacities and a snapshot of the energy landscape in protein GB1 from the pre-denaturation temperature dependence of backbone NH nanosecond fluctuations. *J Mol Biol* **2003**, *325* (1), 149-162.
44. Jansson, H.; Swenson, J., The protein glass transition as measured by dielectric spectroscopy and differential scanning calorimetry. *Bba-Proteins Proteom* **2010**, *1804* (1), 20-26.
45. Young, R. D.; Frauenfelder, H.; Fenimore, P. W., Mossbauer Effect in Proteins. *Phys Rev Lett* **2011**, *107* (15).
46. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
47. Demir, K.; Kilic, N.; Dudak, F. C.; Boyaci, I. H.; Yasar, F., The investigation of the secondary structure propensities and free-energy landscapes of peptide ligands by replica exchange molecular dynamics simulations. *Mol Simulat* **2014**, *40* (13), 1015-1025.
48. Hu, G. D.; Xu, S. C.; Wang, J. H., Characterizing the Free-Energy Landscape of MDM2 Protein-Ligand Interactions by Steered Molecular Dynamics Simulations. *Chem Biol Drug Des* **2015**, *86* (6), 1351-1359.
49. Mollica, L.; Baías, M.; Lewandowski, J. R.; Wylie, B. J.; Sperling, L. J.; Rienstra, C. M.; Emsley, L.; Blackledge, M., Atomic-Resolution Structural Dynamics in Crystalline Proteins from NMR and Molecular Simulation. *J Phys Chem Lett* **2012**, *3* (23), 3657-3662.
50. Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E., Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* **2012**, *41*, 429-52.
51. Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Strope, I. R.; Swenson, J.; Young, R. D., A unified model of protein dynamics. *P. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129-34.
52. Doster, W.; Cusack, S.; Petry, W., Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **1989**, *337* (6209), 754-756.
53. Lipps, F.; Levy, S.; Markelz, A. G., Hydration and temperature interdependence of protein picosecond dynamics. *Phys Chem Chem Phys* **2012**, *14* (18), 6375-6381.
54. Parak, F.; Formanek, H., Untersuchung des Schwingungsanteils und des Kristallgitterfehleranteils des Temperaturfaktors in Myoglobin durch Vergleich von Mössbauer-absorptionsmessungen mit Röntgenstrukturdaten. *Acta Crystallogr. A* **1971**, *27*, 573-578.
55. Knab, J. R.; Chen, J. Y.; He, Y. F.; Markelz, A. G., Terahertz measurements of protein relaxational dynamics. *P IEEE* **2007**, *95* (8), 1605-1610.
56. Jansson, H.; Bergman, R.; Swenson, J., Role of solvent for the dynamics and the glass transition of proteins. *J. Phys. Chem. B* **2011**, *115*, 4099-109.
57. Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M., Solvent mobility and the protein 'glass' transition. *Nat Struct Biol* **2000**, *7* (1), 34-38.
58. Tarek, M.; Tobias, D. J., Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys Rev Lett* **2002**, *88* (13).
59. Doster, W., The dynamical transition of proteins, concepts and misconceptions. *Eur Biophys J Biophys* **2008**, *37* (5), 591-602.

60. Fenimore, P. W.; Frauenfelder, H.; Magazù, S.; McMahon, B. H.; Mezei, F.; Migliardo, F.; Young, R. D.; Stroe, I.; Magazu, S.; McMahon, B. H.; Mezei, F.; Migliardo, F.; Young, R. D.; Stroe, I., Concepts and Problems in Protein Dynamics. *Chem Phys* **2013**.
61. Smith, A. A.; Ernst, M.; Meier, B. H., Because the Light is Better Here: Correlation-Time Analysis by NMR Spectroscopy. *Angew Chem Int Edit* **2017**, *56* (44), 13590-13595.
62. Khodadadi, S.; Sokolov, A. P., Atomistic details of protein dynamics and the role of hydration water. *Bba-Gen Subjects* **2017**, *1861* (1), 3546-3552.
63. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct observation of hierarchical protein dynamics. *Science* **2015**, *348* (6234), 578-581.
64. Zhong, D. P.; Pal, S. K.; Zewail, A. H., Biological water: A critique. *Chem Phys Lett* **2011**, *503* (1-3), 1-11.
65. Chong, S. H.; Ham, S., Dynamics of Hydration Water Plays a Key Role in Determining the Binding Thermodynamics of Protein Complexes. *Sci Rep* **2017**, *7* (1), 8744.
66. Nickels, J. D.; O'Neill, H.; Hong, L.; Tyagi, M.; Ehlers, G.; Weiss, K. L.; Zhang, Q.; Yi, Z.; Mamontov, E.; Smith, J. C.; Sokolov, A. P., Dynamics of protein and its hydration water: neutron scattering studies on fully deuterated GFP. *Biophys J* **2012**, *103* (7), 1566-1575.
67. Hong, L.; Smolin, N.; Lindner, B.; Sokolov, A. P.; Smith, J. C., Three Classes of Motion in the Dynamic Neutron-Scattering Susceptibility of a Globular Protein. *Phys Rev Lett* **2011**, *107* (14).
68. Gallat, F. X.; Laganowsky, A.; Wood, K.; Gabel, F.; van Eijck, L.; Wuttke, J.; Moulin, M.; Hartlein, M.; Eisenberg, D.; Colletier, J. P.; Zaccai, G.; Weik, M., Dynamical Coupling of Intrinsically Disordered Proteins and Their Hydration Water: Comparison with Folded Soluble and Membrane Proteins. *Biophys J* **2012**, *103* (1), 129-136.
69. Smolin, N.; Biehl, R.; Kneller, G. R.; Richter, D.; Smith, J. C., Functional domain motions in proteins on the ~1-100 ns timescale: comparison of neutron spin-echo spectroscopy of phosphoglycerate kinase with molecular-dynamics simulation. *Biophys J* **2012**, *102* (5), 1108-1117.
70. Inoue, R.; Biehl, R.; Rosenkranz, T.; Fitter, J.; Monkenbusch, M.; Radulescu, A.; Farago, B.; Richter, D., Large Domain Fluctuations on 50-ns Timescale Enable Catalytic Activity in Phosphoglycerate Kinase. *Biophys J* **2010**, *99* (7), 2309-2317.
71. Krishnan, M.; Kurkal-Siebert, V.; Smith, J. C., Methyl Group Dynamics and the Onset of Anharmonicity in Myoglobin. *The Journal of Physical Chemistry B* **2008**, *112* (17), 5522-5533.
72. 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.
73. Fischer, M. W. F.; Zeng, L.; Majumdar, A.; Zuiderweg, E. R. P., Characterizing semilocal motions in proteins by NMR relaxation studies. *P Natl Acad Sci USA* **1998**, *95* (14), 8016-8019.
74. Haller, J. D.; Schanda, P., Amplitudes and time scales of picosecond-to-microsecond motion in proteins studied by solid-state NMR: a critical evaluation of experimental approaches and application to crystalline ubiquitin. *Journal of Biomolecular Nmr* **2013**, *57* (3), 263-280.
75. Orjuela Leon, A. C.; Marwosky, A.; Arand, M., Evidence for a complex formation between CYP2J5 and mEH in living cells by FRET analysis of membrane protein interaction in the endoplasmic reticulum (FAMPIR). *Arch Toxicol* **2017**, *91* (11), 3561-3570.
76. Settles, M.; Doster, W., Anomalous diffusion of adsorbed water: a neutron scattering study of hydrated myoglobin. *Faraday Discussions* **1996**, *103* (0), 269-279.
77. Pal, S. K.; Peon, J.; Zewail, A. H., Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution. *P Natl Acad Sci USA* **2002**, *99* (4), 1763-1768.
78. Sakai, V. G.; Khodadadi, S.; Cicerone, M. T.; Curtis, J. E.; Sokolov, A. P.; Roh, J. H., Solvent effects on protein fast dynamics: implications for biopreservation. *Soft Matter* **2013**, *9* (22), 5336-5340.
79. Lau, Y.; Oamen, H. P.; Caudron, F., Protein Phase Separation during Stress Adaptation and Cellular Memory. *Cells-Basel* **2020**, *9* (5).
80. Redfield, A. G., On the Theory of Relaxation Processes. *Ibm J Res Dev* **1957**, *1* (1), 19-31.
81. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *Journal of the American Chemical Society* **1982**, *104* (17), 4559-4570.
82. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *Journal of the American Chemical Society* **1982**, *104* (17), 4546-4559.
83. J. Cavanagh, W. J. F., A. G. I. Palmer, N. J. Skelton, M. Rance, *Protein NMR Spectroscopy*. 2006.
84. Suter, D.; Ernst, R. R., Spin Diffusion in Resolved Solid-State Nmr-Spectra. *Phys Rev B* **1985**, *32* (9), 5608-5627.
85. Lewandowski, J. R.; Dumez, J. N.; Akbey, U.; Lange, S.; Emsley, L.; Oschkinat, H., Enhanced Resolution and Coherence Lifetimes in the Solid-State NMR Spectroscopy of Perdeuterated Proteins under Ultrafast Magic-Angle Spinning. *J Phys Chem Lett* **2011**, *2* (17), 2205-2211.
86. Fisette, O.; Lague, P.; Gagne, S.; Morin, S., Synergistic Applications of MD and NMR for the Study of Biological Systems. *J Biomed Biotechnol* **2012**.

87. Schanda, P.; Ernst, M., Studying Dynamics by Magic-Angle Spinning Solid-State NMR Spectroscopy: Principles and Applications to Biomolecules. *Prog Nucl Magn Reson Spectrosc* **2016**, *96*, 1-46.
88. Lamley, J. M.; Lewandowski, J. R., Relaxation-Based Magic-Angle Spinning NMR Approaches for Studying Protein Dynamics. *Emagres* **2016**, *5* (3), 1423-1433.
89. Choy, W. Y.; Kay, L. E., Model selection for the interpretation of protein side chain methyl dynamics. *J Biomol NMR* **2003**, *25* (4), 325-33.
90. Krushelnitsky, A.; Gauto, D.; Camargo, D. C. R.; Schanda, P.; Saalwachter, K., Microsecond motions probed by near-rotary-resonance R (1 rho) N-15 MAS NMR experiments: the model case of protein overall-rocking in crystals. *Journal of Biomolecular Nmr* **2018**, *71* (1), 53-67.
91. Lamley, J. M.; Lougher, M. J.; Sass, H. J.; Rogowski, M.; Grzesiek, S.; Lewandowski, J. R., Unraveling the complexity of protein backbone dynamics with combined (13)C and (15)N solid-state NMR relaxation measurements. *Phys Chem Chem Phys* **2015**, *17* (34), 21997-2008.
92. Buck, M.; Boyd, J.; Redfield, C.; MacKenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M., Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry-Us* **1995**, *34* (12), 4041-55.
93. Daragan, V. A.; Mayo, K. H., A simple approach to analyzing protein side-chain dynamics from 13C NMR relaxation data. *J Magn Reson* **1998**, *130* (2), 329-34.
94. Giraud, N.; Bockmann, A.; Lesage, A.; Penin, F.; Blackledge, M.; Emsley, L., Site-specific backbone dynamics from a crystalline protein by solid-state NMR spectroscopy. *J Am Chem Soc* **2004**, *126* (37), 11422-3.
95. Reif, B.; Xue, Y.; Agarwal, V.; Pavlova, M. S.; Hologne, M.; Diehl, A.; Ryabov, Y. E.; Skrynnikov, N. R., Protein side-chain dynamics observed by solution- and solid-state NMR: comparative analysis of methyl 2H relaxation data. *J Am Chem Soc* **2006**, *128* (38), 12354-5.
96. Schanda, P.; Meier, B. H.; Ernst, M., Quantitative Analysis of Protein Backbone Dynamics in Microcrystalline Ubiquitin by Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **2010**, *132* (45), 15957-15967.
97. Hologne, M.; Chen, Z. J.; Reif, B., Characterization of dynamic processes using deuterium in uniformly H-2, C-13, N-15 enriched peptides by MAS solid-state NMR. *J Magn Reson* **2006**, *179* (1), 20-28.
98. Kurauskas, V.; Izmailov, S. A.; Rogacheva, O. N.; Hessel, A.; Ayala, I.; Woodhouse, J.; Shilova, A.; Xue, Y.; Yuwen, T. R.; Coquelle, N.; Colletier, J. P.; Skrynnikov, N. R.; Schanda, P., Slow conformational exchange and overall rocking motion in ubiquitin protein crystals. *Nat Commun* **2017**, *8*.
99. Lewandowski, J. R.; Sein, J.; Blackledge, M.; Emsley, L., Anisotropic Collective Motion Contributes to Nuclear Spin Relaxation in Crystalline Proteins. *Journal of the American Chemical Society* **2010**, *132* (4), 1246-+.
100. Lakomek, N. A.; Penzel, S.; Lends, A.; Cadalbert, R.; Ernst, M.; Meier, B. H., Microsecond Dynamics in Ubiquitin Probed by Solid-State N-15 NMR Spectroscopy R-1 rho Relaxation Experiments under Fast MAS (60-110 kHz). *Chem-Eur J* **2017**, *23* (39), 9425-9433.
101. Copperman, J.; Dinpajooh, M.; Beyerle, E. R.; Guenza, M. G., Universality and Specificity in Protein Fluctuation Dynamics. *Phys Rev Lett* **2017**, *119* (15).
102. Gauto, D. F.; Hessel, A.; Rovo, P.; Kurauskas, V.; Linser, R.; Schanda, P., Protein conformational dynamics studied by (15)N and (1)H R1rho relaxation dispersion: Application to wild-type and G53A ubiquitin crystals. *Solid State Nucl Magn Reson* **2017**, *87*, 86-95.
103. Lewandowski, J. R.; Sass, H. J.; Grzesiek, S.; Blackledge, M.; Emsley, L., Site-Specific Measurement of Slow Motions in Proteins. *Journal of the American Chemical Society* **2011**, *133* (42), 16762-16765.
104. Schanda, P.; Huber, M.; Boisdouvier, J.; Meier, B. H.; Ernst, M., Solid-State NMR Measurements of Asymmetric Dipolar Couplings Provide Insight into Protein Side-Chain Motion (vol 50, pg 11005, 2011). *Angew Chem Int Edit* **2012**, *51* (40), 9959-9959.
105. Busi, B.; Yarava, J. R.; Hofstetter, A.; Salvi, N.; Cala-De Paepe, D.; Lewandowski, J. R.; Blackledge, M.; Emsley, L., Probing Protein Dynamics Using Multifield Variable Temperature NMR Relaxation and Molecular Dynamics Simulation. *J Phys Chem B* **2018**, *122* (42), 9697-9702.
106. Mack, J. W.; Usha, M. G.; Long, J.; Griffin, R. G.; Wittebort, R. J., Backbone motions in a crystalline protein from field-dependent H-2-NMR relaxation and line-shape analysis. *Biopolymers* **2000**, *53* (1), 9-18.
107. Smith, A. A.; Testori, E.; Cadalbert, R.; Meier, B. H.; Ernst, M., Characterization of fibril dynamics on three timescales by solid-state NMR. *Journal of Biomolecular Nmr* **2016**, *65* (3-4), 171-191.
108. Good, D.; Pham, C.; Jagas, J.; Lewandowski, J. R.; Ladizhansky, V., Solid-State NMR Provides Evidence for Small-Amplitude Slow Domain Motions in a Multispanning Transmembrane alpha-Helical Protein. *J Am Chem Soc* **2017**, *139* (27), 9246-9258.
109. Oster, C.; Kosol, S.; Hartlmuller, C.; Lamley, J. M.; Iuga, D.; Oss, A.; Org, M. L.; Vanatalu, K.; Samoson, A.; Madl, T.; Lewandowski, J. R., Characterization of Protein-Protein Interfaces in Large Complexes by Solid-State NMR Solvent Paramagnetic Relaxation Enhancements. *Journal of the American Chemical Society* **2017**, *139* (35), 12165-12174.
110. Shi, X. Y.; Prasanna, C.; Nagashima, T.; Yamazaki, T.; Pervushin, K.; Nordenskiold, L., Structure and Dynamics in the Nucleosome Revealed by Solid-State NMR. *Angew Chem Int Edit* **2018**, *57* (31), 9734-9738.

111. Linser, R., Solid-state NMR spectroscopic trends for supramolecular assemblies and protein aggregates. *Solid State Nucl Mag* **2017**, *87*, 45-53.
112. Bockmann, A.; Gardiennet, C.; Verel, R.; Hunkeler, A.; Loquet, A.; Pintacuda, G.; Emsley, L.; Meier, B. H.; Lesage, A., Characterization of different water pools in solid-state NMR protein samples. *Journal of Biomolecular Nmr* **2009**, *45* (3), 319-327.
113. O'Brien, E. S.; Wand, A. J.; Sharp, K. A., On the ability of molecular dynamics force fields to recapitulate NMR derived protein side chain order parameters. *Protein Sci* **2016**, *25* (6), 1156-60.
114. Ackermann, B. E.; Debelouchina, G. T., Heterochromatin Protein HP1alpha Gelation Dynamics Revealed by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **2019**, *58* (19), 6300-6305.

Chapter 2 Analysis of the motional model and application to the model protein GB1

As discussed in the introduction, protein dynamics studied by MAS NMR relaxometry or other techniques rely on descriptions of specific motions. Building on the method introduced previously using variable temperature solid state nuclear magnetic resonance relaxation measurements,¹ here, we extend this approach to measure relaxation at multiple magnetic field strengths, allowing us to better constrain the motional models, and to simultaneously evaluate the robustness and physical basis of the method. The data confirm backbone and sidechain motions exhibiting low and high-energy modes with temperature coefficients around $5 \text{ kJ}\cdot\text{mol}^{-1}$ and $25 \text{ kJ}\cdot\text{mol}^{-1}$.

The results are compared to variable temperature molecular dynamics simulation of the crystal lattice, providing further support for the interpretation of the experimental data in terms of molecular motion. The initial investigation measured longitudinal and transverse spin relaxation at a single magnetic field strength, limiting the ability to unambiguously identify the nature of the observed relaxation phenomena and to cross-validate our interpretation in terms of dynamic behaviour. Here, again using the protein GB1, we therefore extend the approach to measure longitudinal relaxation over a range of different magnetic field strengths and temperatures, thereby exploiting the dependence of relaxation rates on the Larmor frequency expected from Redfield theory. This allows us to test the dynamic origin of the measurements, to better constrain the motional models used to interpret the data and to evaluate the reproducibility of the method. Specifically, we perform the same set of experiments at four different magnetic field strengths (9.4 T, 11.75 T, 14.1 T and 18.8 T), on different sample batches and for different temperature variation protocols. The field-dependent measurements allow a more precise description of the timescale and amplitudes of the backbone motions in GB1, substantiating and refining our findings from the previous single-field analysis.

2.1 Relaxation and dynamics

2.1.1 Theoretical framework

Note that the goal of this section is only to present the tools and theory used. The full derivations present in the cited references have not been reproduced here.

Nuclear spin dynamics can be summarized in the semiclassical vector model formulated by Bloch. The Bloch equation can be written in the rotating frame as

$$\frac{d\mathbf{M}(t)}{dt} = \begin{bmatrix} -R_2 & -\Omega & \omega_1 \sin\Phi \\ \Omega & -R_2 & -\omega_1 \cos\Phi \\ -\omega_1 \sin\Phi & \omega_1 \cos\Phi & -R_1 \end{bmatrix} \mathbf{M}(t) + R_1 M_0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{Eq II1}$$

With the matrix $\mathbf{M}(t) = \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix}$

where R_1 is the spin-lattice or longitudinal relaxation rate and R_2 is the spin-spin or transverse relaxation rate and where the matrix $\mathbf{M}(t)$ representing the bulk magnetic moment, Ω is $\omega_0 - \omega_{rf}$ the offset, which is close to 0 for on resonance pulses, ω_1 is $-\gamma B_1$ the strength of the applied pulse and Φ the phase of pulse. Here we see that the R_1 relaxation rates tend to drive the longitudinal magnetization towards equilibrium, while the R_2 rates drive the transverse magnetization towards zero. R_1 and R_2 rates can be measured for example by inversion recovery or CPMG pulse sequence respectively. To measure R_1 , examples are shown using the pulse sequences in Figure 10.

To relate the measured rates to microscopic parameters, a quantum mechanical description using the density matrix formalism has been used for decades and is at the basis of any modern NMR description.² Nuclear spin relaxation is not based on spontaneous emission but rather due to couplings and interaction between different spins, or with spin-lattice interaction.

Here we will briefly introduce the Bloch-Redfield-Wangness theory (or Redfield relaxation theory) that can treat weak modulation of the spin interactions happening on short timescale. It relies on a semiclassical model, in which the spin systems are treated quantum mechanically, while the lattice is treated classically.

Redfield theory yields a so-called master equation for relaxation³

$$\frac{d\tilde{\sigma}(t)}{dt} = - \int_0^\infty \overline{\left[\tilde{\mathcal{H}}_1(t), [\tilde{\mathcal{H}}_1(t-\tau), \tilde{\sigma}(t) - \sigma_{eq}] \right]} d\tau \quad \text{Eq II2}$$

If the change of density operator is not much slower than the motions, the Eq II2 cannot be simplified from the previous form

$$\frac{d\tilde{\sigma}(t)}{dt} = - \int_0^\infty \overline{\left[\tilde{\mathcal{H}}_1(t), [\tilde{\mathcal{H}}_1(t-\tau), \tilde{\sigma}(t-\tau) - \sigma_{eq}] \right]} d\tau \quad \text{Eq II2bis}$$

The problem is generally solved by using Stochastic-Liouville equation as described in³⁸ since the exact calculation are not feasible.

With the overbar representing an ensemble average, σ the density operator (in the laboratory frame), and where the tilde symbol indicates that the operator is in the interaction frame. σ_{eq} is the value of the density operator at the equilibrium. $\mathcal{H}_1(t)$ the Hamiltonian which contains the time dependent spin interactions that are due to stochastic processes. By considering the nature of the Hamiltonian, and making a series of approximations, we can obtain:

$$\frac{d\tilde{\sigma}(t)}{dt} = - \sum_\alpha \left[V_\alpha, [V_\alpha^\dagger, (\tilde{\sigma}(t) - \sigma_{eq})] \right] J_\alpha(\omega_\alpha) \quad \text{Eq II3}$$

Where V the eigenoperators and the spectral density J_α is the Fourier transform of the autocorrelation function

$$G_\alpha(\tau) = \overline{F_{-\alpha}(t) F_\alpha(t - \tau)} \quad \text{Eq II4}$$

With F the random function of time.

The spectral density thus depends on the nature of the random fluctuations in the Hamiltonian H_1 . In most cases in NMR the fluctuations in H_1 are induced by molecular motions that modulate the orientation dependent interactions, in particular as far as we will be concerned with the dipolar coupling

and the chemical shift anisotropies. A complete description of these formalisms can be found in the texts by Goldman⁴ and Abragam⁵.

Importantly, the approximations made mean that this approach is valid in the ‘fast motion limit’, which means that the dynamics should be faster than the timescale of the interaction that is modulated by the dynamics itself. We assume that the random fluctuations of the magnetic field (which drive relaxation) are much shorter than the evolution of the density operator. This is referred to as the Redfield limit (or weak collision limit). This limitation should be kept in mind.^{6, 7}

With the master equation, we can now use models for the spectral densities in order to predict relaxation times quantitatively. The so-called “model-free” model is popular since it assumes no specific geometry for the motion, only that different orientations are possible and that the system will jump from one to the other at a specific rate and that the amplitude of the motion is restricted with an order parameter.⁸

In this model, the spectral density function, which is the Fourier transform of the correlation function describing the fluctuation of the magnetic field will be

$$J(\omega) = \frac{5}{2} \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} \quad \text{Eq II5}$$

With τ_c being the rate of the motion, and S^2 the order parameter. The order parameter describes the spatial restriction of the motion. For fully isotropic motion it is equal to 0, and for a fully restricted motion its value is 1.

We then have the final expression to describe R_1 for a heteronuclear two spin system.

$$R_1 = \left(\frac{\delta_D}{4} \right)^2 (J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)) + \delta_{CSA}^2 \frac{3}{4} J(\omega_S) \quad \text{Eq II6}$$

With the corresponding dipolar and chemical shift terms being

$$\delta_D = -2 \left(\frac{\mu_0}{4\pi} \right) \frac{\hbar \gamma_I \gamma_S}{r_{IS}^3} \quad \text{Eq II7}$$

And

$$\delta_{CSA} = \Delta\sigma\gamma_I B_0 \quad \text{Eq II8}$$

where $\Delta\sigma$ is the difference between the two axes of the axially symmetric (as an approximation) shift tensor, B_0 is the magnetic field strength, ω_I and ω_S the Larmor frequency of the two spins, γ_I (respectively γ_S) represents the gyromagnetic ratio of the I (respectively. S) atoms, μ_0 denotes the permeability of free space, \hbar the reduced Plank's constant and r_{IS} is the distance between the two nuclei.

2.1.2 Model free approach for temperature dependent measurements

2.1.2.1 Theory of the model

As discussed above NMR relaxation is mainly caused by fluctuations of the anisotropic interactions, like the CSA tensor or the dipolar couplings due to relative motions of the nuclei. Thus, the relaxation rates of individual resonances can be used to report on the dynamics of different regions of the protein.^{9, 10}

Here we assume that each motional mode has a temperature dependent rate of motion according to an Arrhenius relation¹¹ as follows

$$\tau_k = \tau_{\infty,k} e^{(E_k/RT)} \quad \text{Eq II9}$$

Where τ_k is the correlation time of motion at a given temperature, E_k denotes the activation energy for the mode, and $\tau_{\infty,k}$ is analogous to a timescale and in abstract terms corresponds to the correlation time of the motion at infinite temperature. Note that comparisons should be made between the times τ_k at given temperatures, since $\tau_{\infty,k}$ is not particularly informative.

This equation is not universal to describe motions (or more generally exchange processes). For example the deformed Arrhenius equation with : $\beta = 1/k_B T.$, and the parameter d. This equation simplify to the Arrhenius equation for d=1

$$k(\beta) = A(1 - dE\beta)^{1/d} \quad \text{Eq II10}$$

The correlation times are related to the relaxation times through the spectral densities, $J(\omega)$, presented below.) The following expressions are used for the different relaxation rates:¹²

-spin-lattice relaxation for protons

$$R_1 = C_{interaction}(J(\omega_H) + 4J(2\omega_H)) \quad \text{Eq II11}$$

-dipolar spin-lattice relaxation for heteroatoms

$$R_1^{DD} = C_{interaction}(J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)) \quad \text{Eq II12}$$

-CSA spin-lattice relaxation for heteroatoms (not for methyl R_1 because the CSA is negligible compared to the dipolar coupling to drive relaxation in those conditions.)

$$R_1^{CSA} = C_{interaction}J(\omega_X) \quad \text{Eq II13}$$

where ω_H is the proton Larmor frequency, ω_X the the heteroatom Larmor frequency (^{13}C or ^{15}N) and ω_1 the nutation frequency of the radio frequency (rf) field. $C_{interaction}$ is a constant depending on the type of interaction, as discussed below.

We assume a Gaussian-Markoff process (this approximation assumes that any motion is independent from each other) for the individual motional modes. This results in a spectral density $J(\omega)$ at frequency ω , represented as a sum of the different modes.

$$J(\omega) = \sum_{k=1}^N C_{k,amplitude} \frac{\tau_k}{1 + \omega^2 \tau_k^2} \quad \text{Eq II14}$$

If the Gaussian-Markoff approximation are not true, any further treatment would need to identify the positive (or negative) contribution of a motion to the next one.

With N being the number of different motional modes (in our case from 1 to 3). $C_{k,amplitude}$ is a dimensionless factor associated with the amplitude of the motion, which can also be described

through the order parameter of the motion. It is equivalent to $\frac{5}{2} * (1 - S^2)$ described in the previous section. Note that in the following we assume that the amplitude of motion is constant as a function of temperature. This assumption is further verified as discussed later. This simplification is necessary to avoid overfitting the relaxation data.

C_k is the product of factor associated with the amplitude, and the interaction of the motion ($C_k = C_{k,amplitude} \times C_{interaction}$). $C_{interaction}$ depends on the exact geometry of the interaction and since we are working with an ensemble of resonances (the signal is typically integrated over a large region of the spectrum, where all the observable resonances of a given type contribute) we can not have a precise value. Moreover, coherent effects may arise from not totally averaging of the dipolar interaction, which will perturb our measurement of *stricto sensu* $C_{interaction}$. In order to perform our fits, we first estimated $C_{interaction}$ and added a multiplicator factor as a fitting parameter.

The $C_{interaction}$ constants are given as:¹⁰

¹⁵N dipolar R_1

$$\frac{1}{10} \left(\frac{\mu_0 \gamma_H \gamma_N \hbar}{4\pi r_{NH}^3} \right)^2 \quad \text{Eq II15}$$

¹⁵N CSA R_1

$$\frac{2}{15} \omega_N^2 (\sigma_{\parallel} - \sigma_{\perp})^2 \quad \text{Eq II16}$$

¹³C dipolar R_1 C'

$$\frac{1}{10} \left(\frac{\mu_0 \gamma_H \gamma_C \hbar}{4\pi r_{CH}^3} \right)^2 \quad \text{Eq II17}$$

¹³C dipolar R_1 methyl

$$\frac{3}{10} \left(\frac{\mu_0 \gamma_H \gamma_C \hbar}{4\pi r_{CH}^3} \right)^2 \quad \text{Eq II18}$$

^{13}C CSA R_1 C'

$$\frac{2}{15} \omega_C^2 (\sigma_{\parallel} - \sigma_{\perp})^2 \quad \text{Eq II19}$$

γ_X represents the gyromagnetic ratio of the X atom, μ_0 denotes the permeability of free space, \hbar the Plank constant, $\sigma_{\parallel} - \sigma_{\perp}$ the chemical shift tensor (assumed axially symmetric for a first approximation to decrease the complexity) with the value of -160 ppm for the N-H coupling in the amide bond, -15 ppm for the N-H coupling of the lysine side chain, -172 ppm for the carbonyl-proton coupling of the carbonyl and negligible for the methyl. All those approximations have been widely used. $^{13}r_{CH}$ and r_{NH} denote the average inter-nuclear distance between carbon and proton, and nitrogen and proton atoms respectively, given as 1.115 Å for the C-H coupling of the methyl, 2.04 Å for the C-H coupling of the carbonyl and 1.02 Å for the N-H coupling.¹⁰

2.1.2.2 *Type specific variable temperature relaxation measurements.*

In the light of the above, in order to access activation energies, we carry out relaxation measurements as a function of temperature on the solid samples. Experimentally we can access a range from 100 K up to room temperature (or above). The different spy nuclei can be selected by using different pulse sequences or the difference in the respective chemical shifts of the nuclei. At low temperatures resolution in the NMR spectrum degrades¹⁴⁻¹⁷ such that site-specific information is no longer accessible, but type specific information is still resolvable as presented in figure below.

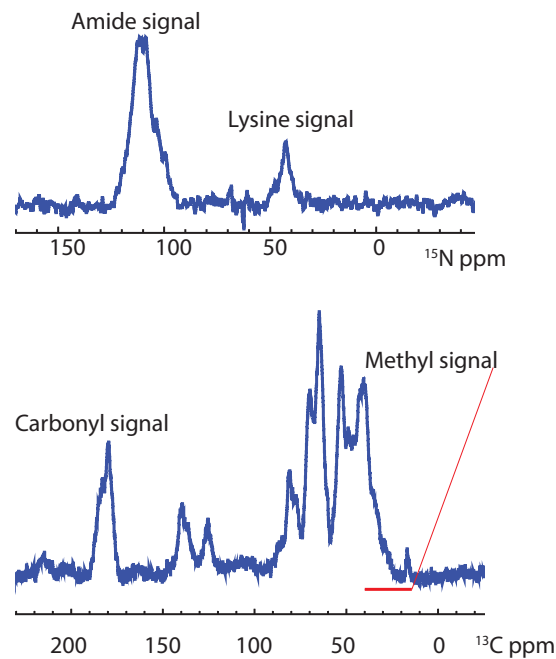


Figure 8 : Site specific resolution and signal used for our experiment, here in the GB1 spectrum at 400MHz and 100K (lowest resolution of our dataset)

Here we use ^{13}C relaxation from the backbone carbonyl or from the methyl group, ^1H relaxation from the hydration shell or bulk solvent, and ^{15}N relaxation from the backbone or the lysine side chain. Thus, the rates we measure here are the sum of the contributions from all of the observable nuclei of a given type in the protein. Here we measure the relaxation parameters T_1 . Fast motions (ps-ns) will be efficient in inducing spin-lattice relaxation (T_1), while slower motion (ns-ms) will have more impact on the relaxation in the rotating frame ($T_{1\rho}$) or on spin-spin relaxation (T_2').¹⁸

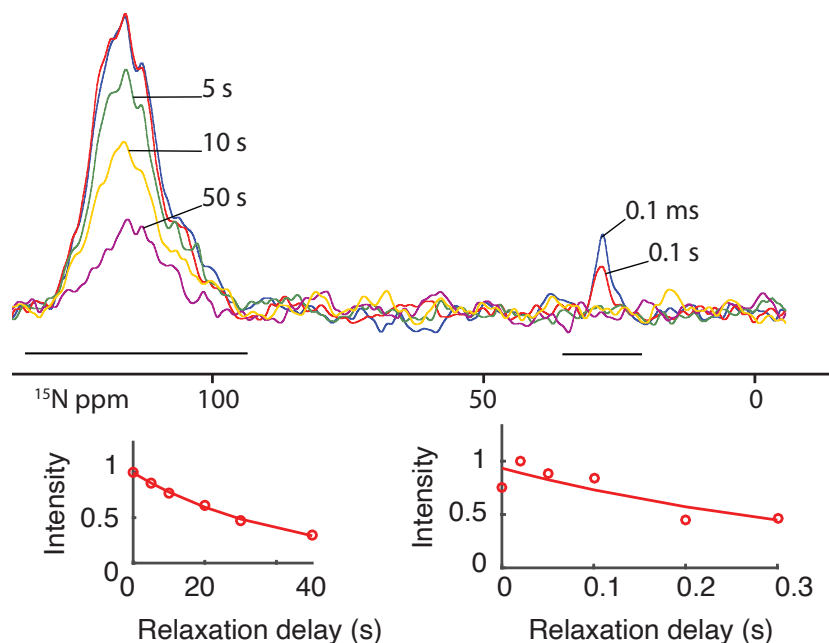


Figure 9 : Spectra of ^{15}N R_1 measurements on GB1 at 600MHz and 270K, the different curved represent the different delays. The pulse can be found in Figure 10. The top panel represents few spectra, with the integrated region underlined. The bottom graphs are their respective fits to a single exponential, with R_1 of 2.5 s^{-1} for the lysine resonance and 0.03 s^{-1} for the amide.

For ^{13}C and ^{15}N we determined the relaxation rates by fitting the slope of the data to a single exponential Ae^{-t/T_1} , with A and T_1 as fitting parameters, as presented in Figure 9. The ^1H measurements were performed using saturation recovery experiments, the pulses sequences are presented Figure 10.

We fitted the curves to $A(1 - e^{-(t/T_1)^\beta})$ with A , T_1 and β as fitting parameters. The stretching parameter becomes higher at low temperature, and reports on the heterogeneity of the relaxation curve. We believed that it is more important in the case of proton relaxation because of the unique signal integrated taking in account all protons. The stretching parameter represents a situation where there is a distribution of relaxation rates centered around a mean. It allows to obtain a good approximation of T_1 without having to rely on too many different parameters. This is not always useful but will give the same results (for T_1) as a fit without stretching parameter if the disorder is low, so the risk of overfitting is negligible.

All our experiments were performed at 10 kHz MAS rate. In this spinning rate regime, and in presence of a large network of coupled protons, spin diffusion, and especially proton driven spin diffusion (PDSD) can become a confounding factor for some rates. Indeed, magnetization will diffuse between the different nuclei, and our measurement will probe an average property of the protein rather than

the ones of a specific sub-part. We detail in section 2.2 how we were able to spot when PDSD was occurring.

2.1.2.3 Pulse sequences

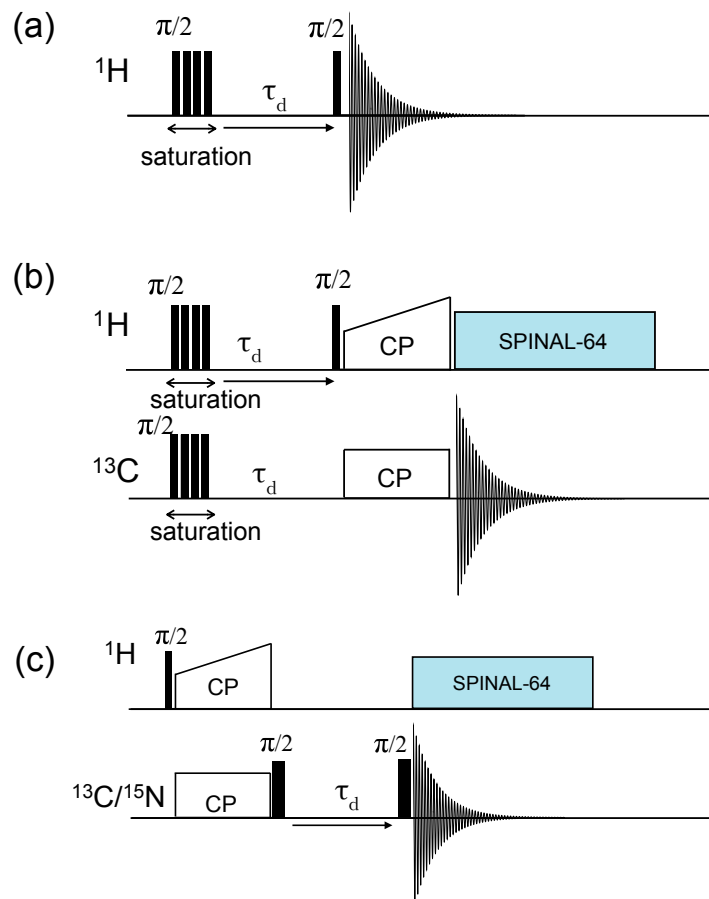


Figure 10: Schematic representation of the pulse sequences. a) R_1 measurement for ^1H b) ^{13}C detected R_1 measurement for ^1H c) R_1 measurement for heteroatoms (^{13}C , ^{15}N). The CP parameters and the $\pi/2$ pulses were optimized for each temperature if needed, note that the tuning and matching change with temperature. In particular an abrupt change of the tuning, that is also reflected in the change of linewidth for the proton signals, is observed at around 250K, where the bulk solvent freezes.

Figure 10 represents the pulses sequences used for measuring the R_1 . The pulses were adjusted at different temperatures when the condition for a 90° pulse changes. This was correlated to change in the tuning and matching condition of our sample.

2.1.2.4 Model assumptions

Note that the MD calculation in this subsection were performed by Nicola Salvi and Martin Blackledge, not directly by myself. I was involved in the discussion of the results.

As already mentioned, our model comports certain assumptions, notably that the amplitude of the motions (timescale) does not change significantly with temperature, and that the activation energy of the motions is following an Arrhenius law. In order to test these assumptions, we performed molecular dynamics simulations of the crystalline lattice of the hydrated GB1 protein at four temperatures from 248 to 278 K. The results demonstrate almost negligible change in the amplitude of relaxation-active backbone motions and suggest that the steep change in longitudinal relaxation rates within this range is indeed due to significant slowing of intermediate timescale (nanosecond) backbone motions with decreasing temperature.

Three independent 200 ns trajectories of a periodic box containing 32 explicit copies of the protein comprising 8 asymmetric units of 4 proteins,¹⁹ were run at four temperatures from 248 to 278 K. Figure 11 shows the order parameters derived from the angular correlation functions describing the motion of inter-nuclear backbone NH bond vectors which were modeled by fitting to a three-exponential fit, in order to extract motional amplitudes and characteristic timescales. The full procedure of the MD simulations is described below. The MD simulations were not performed with a wider range of temperature since force fields are valid only for a specific temperature range.

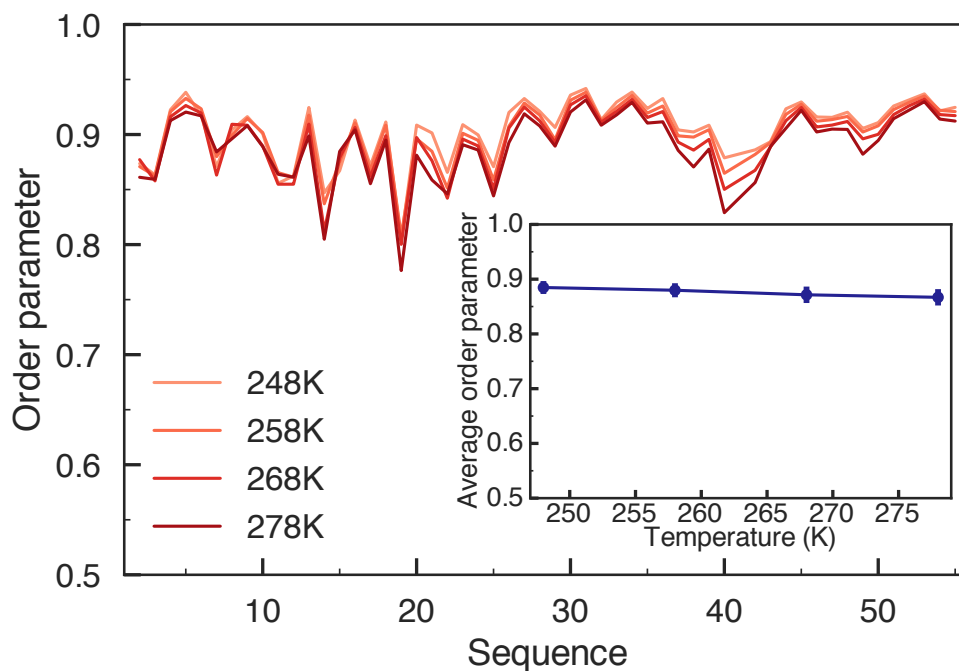


Figure 11: Temperature dependence of dynamic parameters for backbone amide NH bonds derived from MD simulations. Inset: The average order parameter calculated by averaging the values of each of the four curves in the main panel

The marginal systematic decrease observed in Figure 11 when increasing the temperature is because of the increase of the amplitude of very fast (<10 ps) fluctuations that do not contribute significantly to relaxation. The inset shows that the average order parameter varies only slightly as a function of the temperature, and thus justifies our approximation that the order parameters do not vary significantly with temperature.

More specifically, as described previously,¹⁹ a “superlattice” of 32 copies of protein GB1 (equivalent to 8 unit cells of 4 proteins each), 4227 water molecules and 128 Na⁺ ions was constructed according to the crystallographic symmetry. The dimensions of the superlattice allow for a periodic box. Simulations were carried out using GROMACS 5.1.2²⁰ with the AMBER99SB force field.²¹ After minimization using the steepest descent algorithm three independent 200 nanosecond trajectories were calculated at each of the four temperatures - 248, 258, 268 and 278 K. Angular correlation functions representing the motion of the relaxation-active interactions were analyzed as previously described, in terms of amplitudes and timescales of distinct motional modes:

$$C(t) = A_0 + \sum_{k=1}^n A_k e^{-t/\tau_k} \quad \text{Eq II20}$$

with $n=3$ for all cases discussed here. Most notably the amplitude of motions derived from the fit of the correlation functions are seen to vary with respect to temperature, but within a very restricted way described in the Figure 11.

The characterization of the effective timescales extracted from these fits is less easy to compare with experiment, firstly because characteristic timescales are less accurately predicted from MD simulation than motional amplitudes and, secondly, because the number of relaxation active contributions at higher temperatures are not equivalent to the experimental study (where one main contribution is assumed to dominate). Nevertheless, the trend seems to qualitatively reproduce experimental observation.

As an example, correlation functions from amino acids in each of the 32 copies of the protein in the simulation box were averaged and fitted to a three-exponential fit Figure 12. Note, that most of the temperature-dependent variation of the curve is due to the amplitude of very fast motions (i.e. the value of the correlation function at its first plotted point, 10 ps). Average timescales derived from the weighted averaged of the results of the fit show systematic changes in timescale (from 0.76 to 1.3 ns over the range of 278 to 248 K). This would correspond to an effective activation energy of 12 kJmol⁻¹.

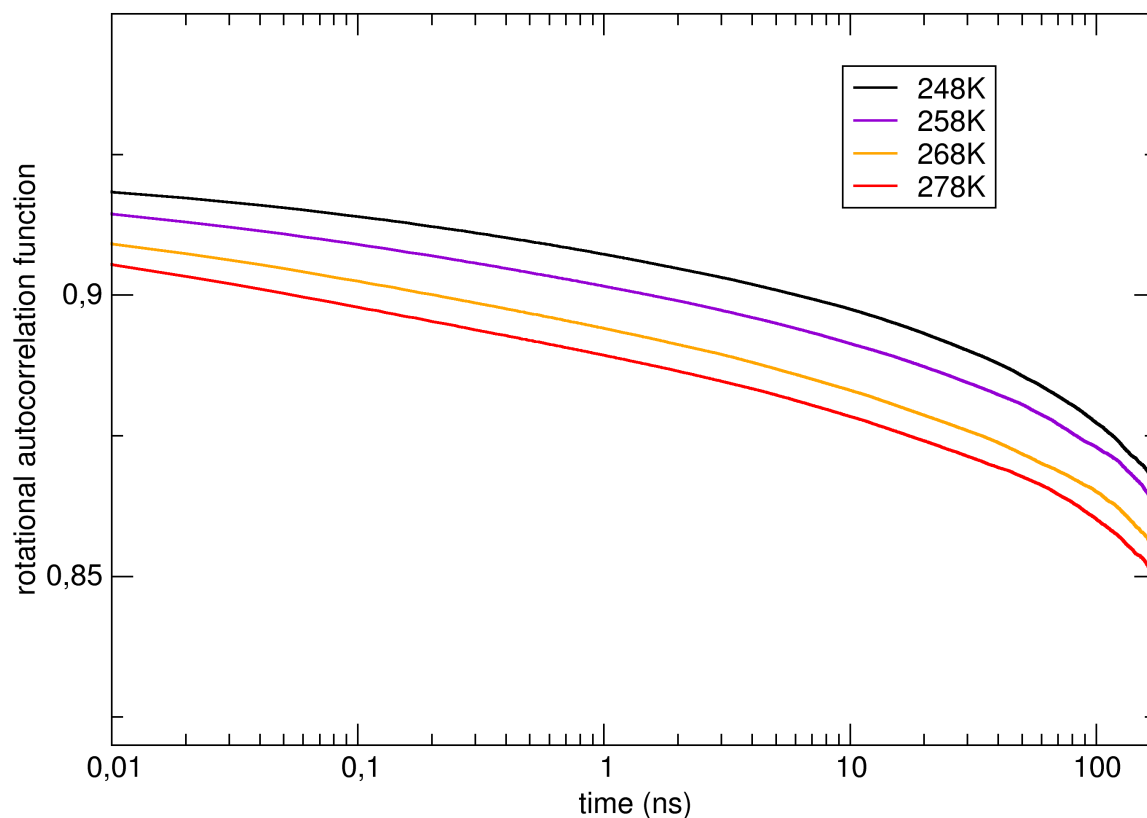


Figure 12: Average rotational correlation function of backbone N-H bonds calculated as described above.

Note the most evident differences between the curves of Figure 12 for different temperatures results from the increasing amplitude of the fast (<10 ps) initial decay of the correlation function.

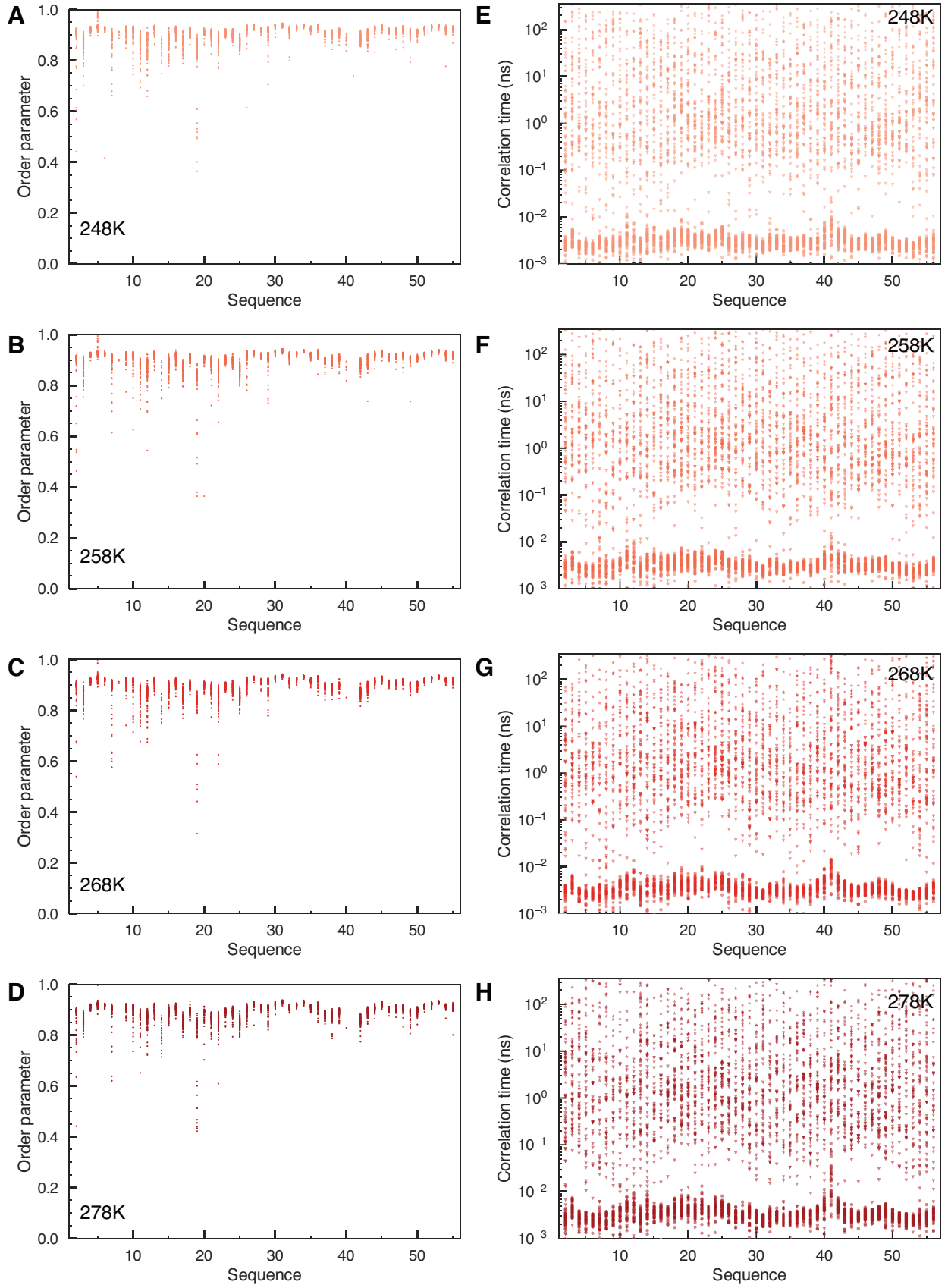


Figure 13: Order parameters calculated for each of the 32 copies of the protein in the superlattice at 248 (panel A), 258 (B), 268 (C) and 278 K (D). Motional timescales panels E-H from the corresponding temperature.

Order parameters in the superlattice in the Figure 13 are clustered around a well-defined average value, providing a useful comparison with experimentally-determined parameters. On the contrary, motional timescales with the exception of the fast component around 5 ps, are much more scattered, thereby hindering any quantitative comparison with the experimental results.

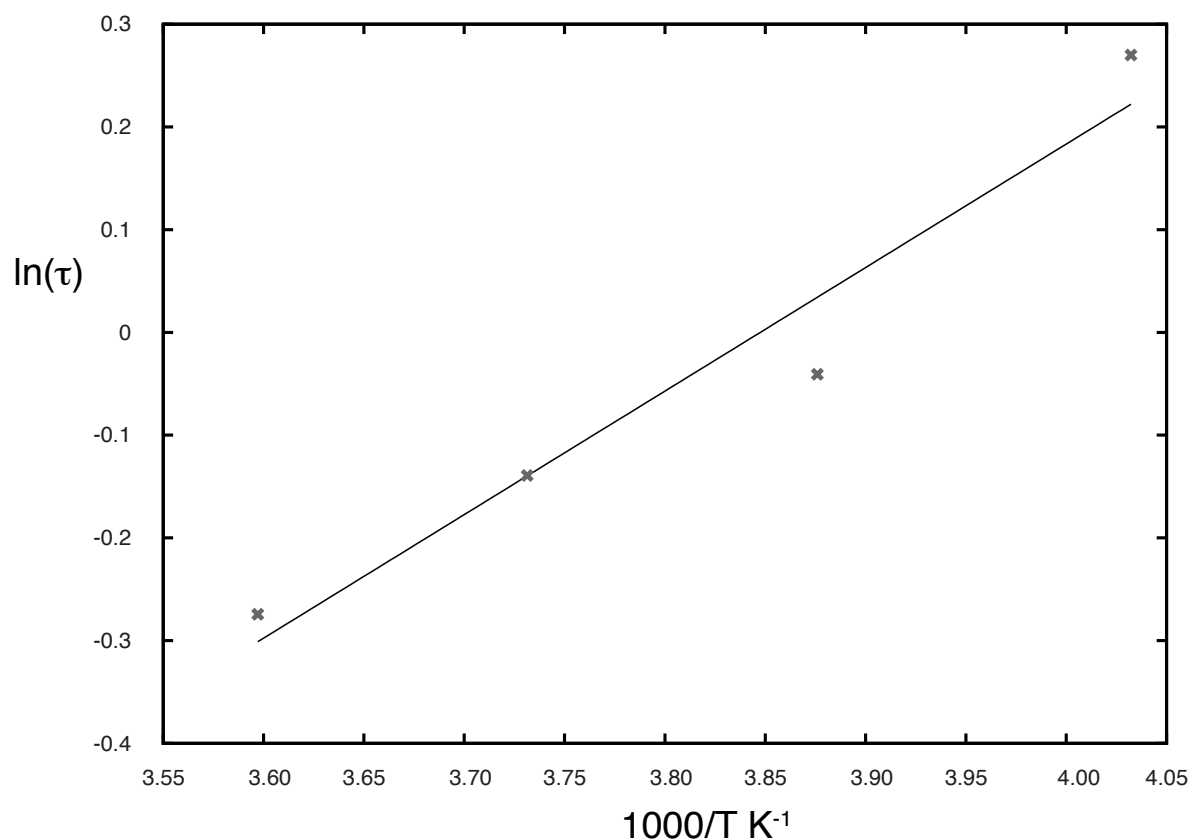


Figure 14: Timescales resulting from the fit of individual correlation functions (see Figure 18) at each temperature – points show averages over 32 copies of 56 N-H^N pairs. Calculated values (points) are fit to an Arrhenius relationship with effective activation energy of 12 kJmol⁻¹ (line). Averages are weighted by their contribution to the associated correlation function.

In conclusion, it seems that the temperature does not significantly impact the amplitude of motions that are driving relaxation.

2.2 Reproducibility and self consistency of the model

2.2.1 Method

In this section we study a microcrystalline preparation a small globular protein, GB1. We measure five ^1H , ^{13}C and ^{15}N relaxation rates at multiple fields (from 400 to 900 MHz) and at temperatures in the range 100 K to 300 K.

2.2.1.1 Sample preparation

The B-1 Immunoglobuline G Binding Domain (GB1 in the following), is a 6.23 kDa, 56 aminoacids water-soluble protein, studied here in a microcrystalline form. Its structure and sequence are shown in Figure 15.

The uniformly ^1H , ^{13}C , ^{15}N labelled, microcrystalline GB1 protein was purchased from Cortecnet (10 mg for each patch), and has been packed without any further treatment in a 3.2 mm rotor by ultracentrifugation using Giotto Biotech device (100 000 g, 4°C) in order to sediment most of the microcrystals and remove as much as possible of the supernatant.²² The rotor is stored at 4°C between experiment sessions. In the crystallization procedure (performed by Cortecnet), the buffer was firstly replaced by extensive dialysis against 50 mM sodium phosphate buffer (pH 5.5). The protein was concentrated to 25 mg.ml⁻¹. A mixture of 1:2 isopropanol:methyl-2-4-pentane-diol was used for precipitation of the protein by adding 3 volumes of the mixture for 1 volume of protein solution which gave the sample the appearance of a milky solution.²³

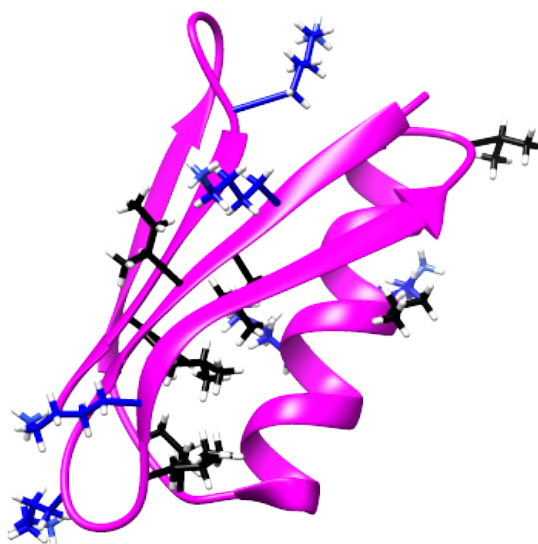


Figure 15 : Structure of GB1 with the highlighted lysine side chains (blue) and methyl groups (black). Full sequence: MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWYDDATKTFTVTE

2.2.1.2 Solid-state NMR spectroscopy

Solid state NMR measurements were carried out on Bruker AV-I 9.4 T, AV-IIIHD 11.75 T, AV-III 14.1 T, AV-III 18.8 T and 21.1 T spectrometers. For all the measurements 3.2 mm low temperature MAS probes were used in triple resonance (^1H , ^{13}C , ^{15}N) mode. All the experiments were performed at the same spinning speed (10 kHz) to reduce overall variability. The spinning speed set to 10 kHz in order to remove the overlap with the sidebands manifold with the signals of interest.

The measurements were performed on two batches of GB1 protein. The two 11.75 T dataset were acquired on the first batch of GB1. 9.4 T, 14.1 T, 18.8 T measurements were done on a second batch of GB1. Each of these 5 measurements was made as a function of the temperature, going from 105 K to 285 K in steps of 5 or 10 K. VT, bearing and drive display temperatures were held identical in the cooling cabinet, in order to minimise temperature gradients across the sample. All the noted temperatures had a precision of ± 10 K. At 11.75 T, measurements were carried out twice (same batch), first with an increasing order of temperatures and a second time with a random order of temperatures. The data are reproduced from.²⁴

The temperature was measured using the internal sensors of the probe. We verified the agreement between the actual and measured temperature by using lead nitrate, as described below. We are aware that using a different sample/pulse sequence could change our results, but since we are using a wide range of temperatures, such small differences are not expected to be important.

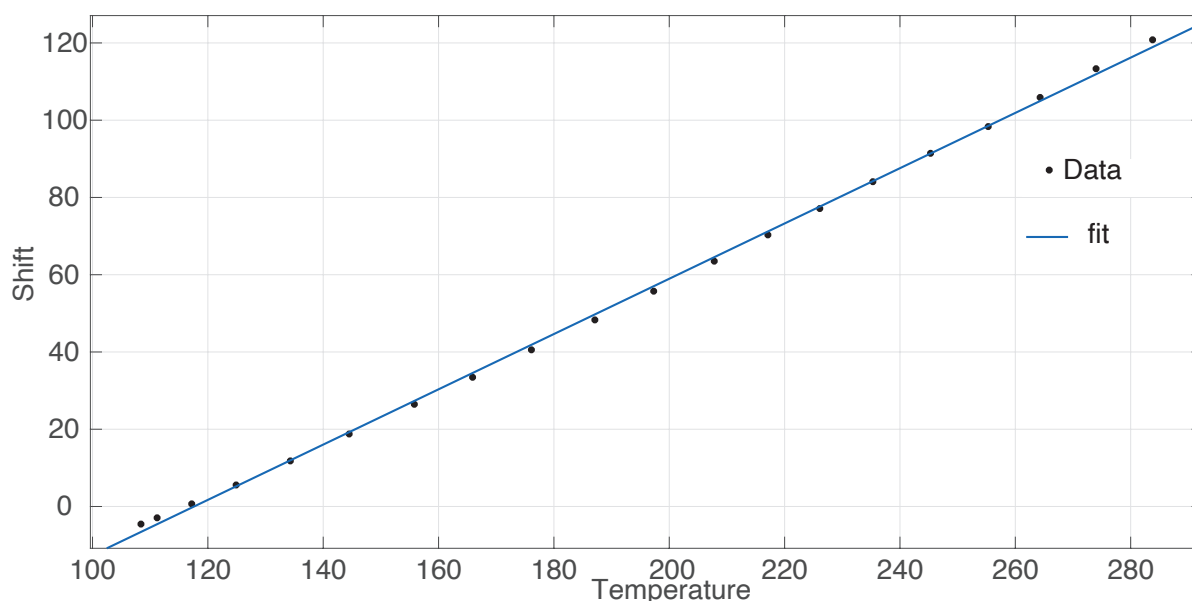


Figure 16: Sensor temperature as a function of chemical shift in ppm for lead nitrate.

To control the temperature, we measured the lead nitrate shift as a function of temperature as described in.²⁵ From a linear fitting of the measured shift as a function of the measured temperature, we obtained a value for the slope of 0.72 ± 2 ppm/K, as compared with the reported value of 0.753ppm. We can therefore conclude that indeed a small shift between real temperature and read temperature is present in our measurements, nonetheless due to the large range of our measurements, this will not significantly impact our results. We will therefore use the “sample temperature” sensor of the spectrometer as a good estimation of the real temperature.

2.2.2 Confidence intervals

The error bar for any specific relaxation rate was obtained using the Student law.²⁶ This gives us the 95% confidence interval for the R_1 value. Equations describing the return to equilibrium of a spin system are well known, and are illustrated in the previous chapter.

In order to verify the reproducibility of the method, we checked if it was possible to accurately replicate the results.

Figure 17a shows ^{15}N R_1 values measured as a function of temperature for GB1 in three different experiments. The red symbols correspond to the data from Lewandowski *et al.*¹ recorded in Lyon in 2012 (here called set I), and the blue symbols are the R_1 rates extracted from data recorded in Lausanne with a different sample of GB1 and progressively increasing temperatures (here called set II). Green symbols show relaxation rates from experiments on the same sample, performed two months later on the same spectrometer using a random order of temperatures (here called set III). The experiments were performed on different batches of the protein (from the same source), with different spectrometers, by different people and in different order, to maximize the variability. This is done to ensure that, when comparing different proteins, small experimental variations would not have a strong influence on the extracted motional parameters.

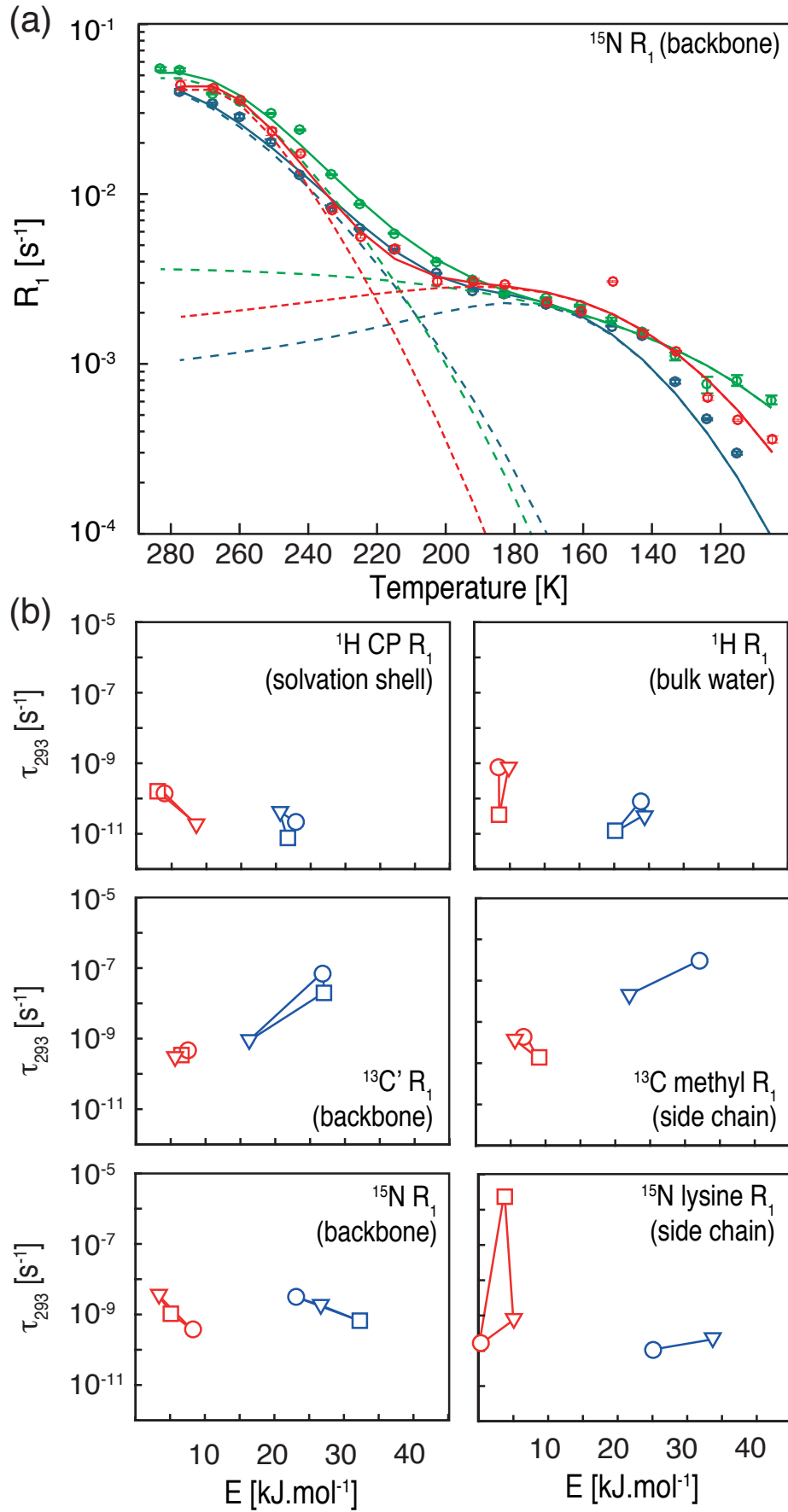


Figure 17: a) Fitting of repeated measurements of ^{15}N R_1 (amide signals, mostly representative of the backbone) for GB1. Red circles show data set I, the blue circles data set II, and the green circles show data set

III. Dashed lines show the contribution due to two motional modes, and the sum is shown as the solid line, where data were fit to a two-component relaxation model. The number of components was statistically justified with an F-test ($p < 0.05$). b) Fitted activation energy plotted against the correlation time at 293 K, from $\tau_k = \tau_{\infty,k} e^{(E_k/RT)}$, for the indicated relaxation rate for the three independent measurements. In each panel the red symbols show the low energy motional mode and the blue symbols show the higher energy mode. Squares show data set I,¹ triangle show data set II, and circles show data set III. The lines connecting the symbols for the three data sets are intended as a guide to the eye.

The same temperature behavior is observed in all three datasets that can all be fit to two motional modes, one with low activation energy dominant at low temperatures, and one with high activation energy appearing at higher temperatures. The activation energies and correlation times match well between the three samples and there is no evidence for hysteresis in the temperature dependence. Similar results are found for activation energies and correlation times extracted for the other 6 relaxation rates (Figure 17). Note that while the energy barriers are accurately defined, the motional time-scales are less accurate, and indeed relatively large variation of this parameter will have a moderate impact on the fit obtained.²⁷

For the cross-validation procedure²⁸ (10-fold-split) the experimental data from the three sets (I, II and III) were combined and 90% of the combined data (R_1) were randomly sampled (training set) and used to predict the remaining 10% (validation set) of the experimental data (R_1). This procedure was repeated 100 times. The random sampling was restricted to avoid double sampling and to contain at least two data points per temperature range (± 5 K). For ^{15}N lysine we only used the data sets II and III in the cross-validation procedure. Also for ^{15}N lysine only set III contains R_1 values above 220 K. Thus, the random sampling was only restricted to avoid double sampling. Here we plot the experimental against the predicted R_1 values for both the training and the validation sets.

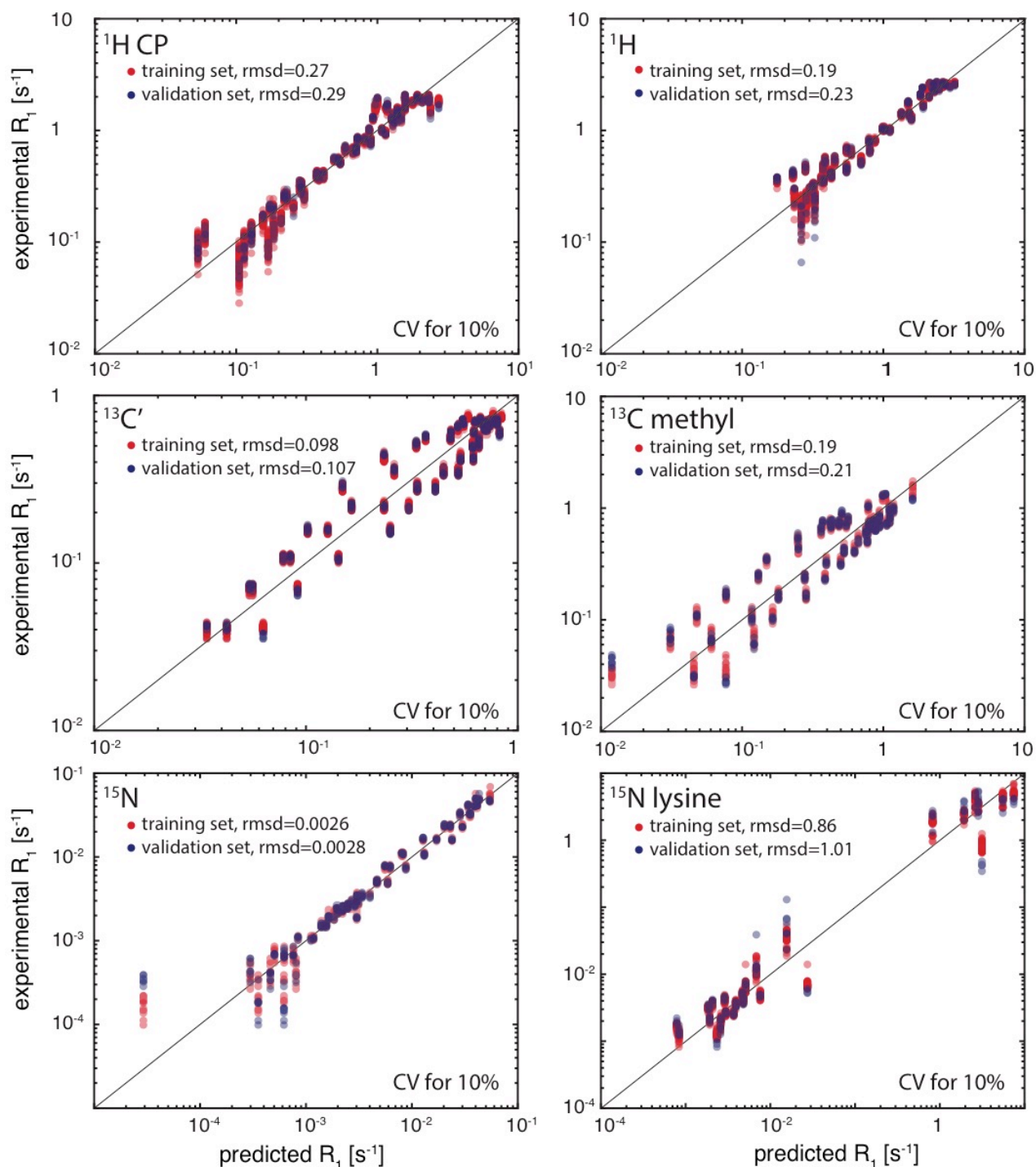


Figure 18: Scatterplots showing the experimental against the predicted R_1 values for all the used probes. The predicted R_1 values are calculated using a model fit to 90% of the combined data sets (I, II and III). The red points show the training set and the blue points show the validation set. The black line corresponds to a one-to-one mapping between the experimental and predicted values.

As we can see from Figure 18, the training dataset allows to generalise our finding to an ‘independent’ dataset, furthermore indicating the high replicability of our results.

In order to estimate the error on our result, we used a Monte Carlo approach. To do so, a set of data was first fitted with the model previously described. To estimate the errors on the extracted constants, 500 new artificial data sets containing the same number of points, with similar displacements from the fitted curve, were generated using the fitted constants. These new data sets were then fitted again. The 500 sets of extracted values were then compared. Note that this method only measures the precision and not the accuracy. Therefore we are testing the precision of the extracted parameters, assuming our model is correct. So, the error bars regarding a specific energy will be representing the 95% scatter interval of our artificial fits.

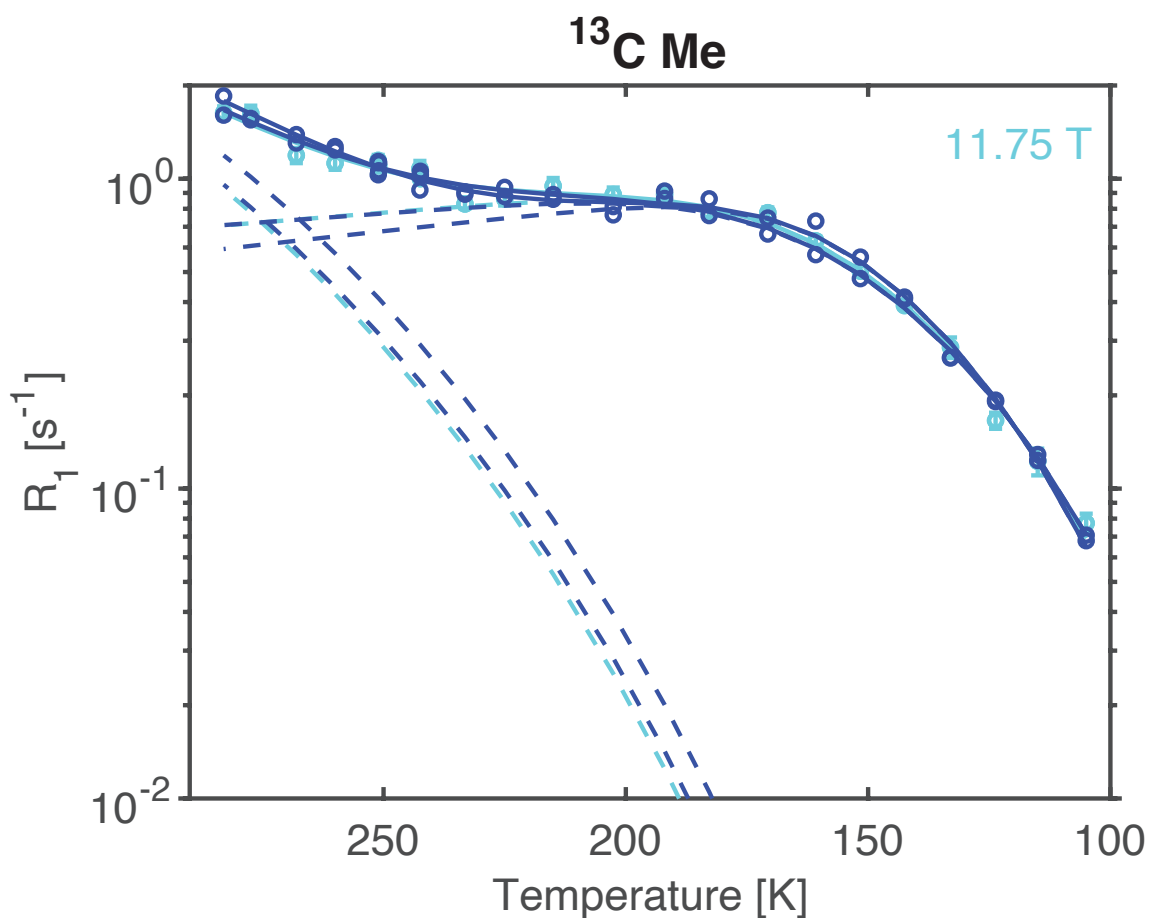


Figure 19: An example with in light blue the fit, and in dark blue two generated sets of data with the associated fits, while below the associated code for one example is reported in the Appendix.

2.2.3 Self-consistency of the model

In this section we will show how using multiple fields allows us to confirm the accuracy of our relaxation model, by focusing on the specific case of carbonyl detected R_1 relaxation.

First, we note that intrinsic $^{13}\text{C}'$ R_1 rates, which would be primarily due to ^{13}C CSA, are expected to be longer than those we observe. Indeed, Lewandowski et al.²⁹ already noted that in a fully ^{13}C enriched protonated protein, spin diffusion of longitudinal magnetization between nearby carbon atoms is likely to be faster than the R_1 relaxation rates, and it is well known that this will lead to averaged apparent R_1 relaxation rates that are dominated by the fastest relaxing local species.³⁰ This will be much less significant for transverse relaxation rates or rates between distant spins such as ^{15}N .³¹ This means in particular that the apparent relaxation of the carbonyl nuclear spins are likely to be dominated by the faster relaxation of the nearby methyl groups within the same amino acid.²⁹ The multi-field data here provide very compelling evidence to support this hypothesis. Figure 20 shows that whereas both pure CSA relaxation or dipole-dipole relaxation can explain the behaviour observed at a single field as a function of temperature, the field dependence of the two mechanisms is different, and only the dipole-dipole mechanism can correctly explain the field dependence of the $^{13}\text{C}'$ R_1 , confirming relayed relaxation of the carbonyls by the methyl sinks. As expected in this analysis, the $^{13}\text{C}'$ R_1 are now finally also explained by two modes with this model, as are the $^{13}\text{C}_{\text{methyl}}$ R_1 .

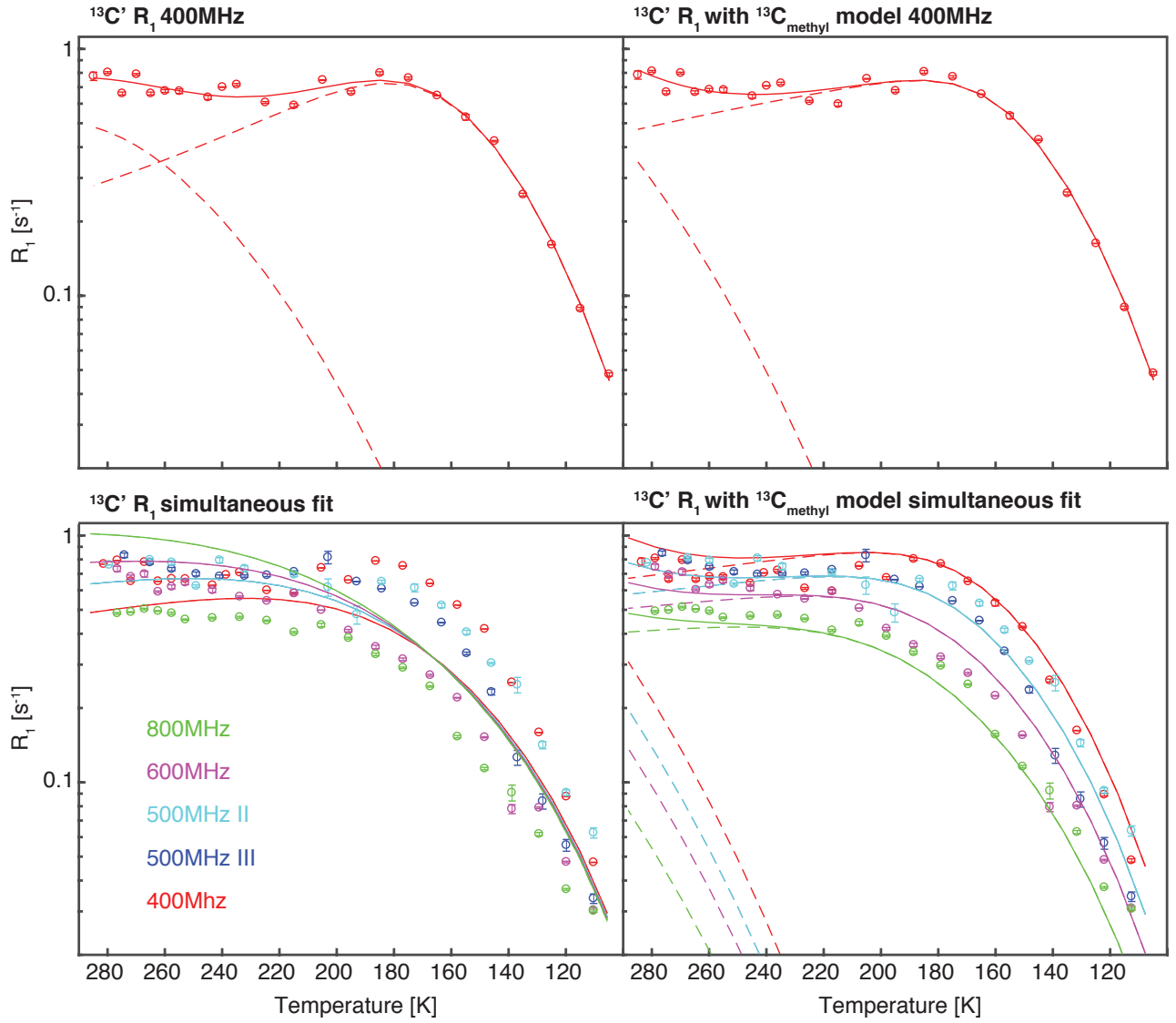


Figure 20: $^{13}\text{C}'$ R_1 rates fitted with two different models. Either with the model for CO relaxation on the left, with major relaxation contribution coming from CSA, or on the right using the model for $^{13}\text{C}_{\text{methyl}}$ R_1 , with the major contribution to relaxation coming from dipole-dipole interactions. The upper panels show the single field fit while the lower panels shows the conjoint fit for the complete dataset.

This is why in the following interpretation we will not consider the carbonyl detected resonances since they actually inform us on methyl dynamics

The full developed equation describing methyl relaxation is here (only CSA and no dipole-dipole relaxation):

$$R_1 = C_{\text{interaction}(1)} \Delta\sigma\gamma_I B_0 \frac{3}{4} \left(\frac{\tau_{k(1)}}{1 + \omega^2 \tau_{k(1)}^2} \right) + C_{\text{interaction}(2)} \Delta\sigma\gamma_I B_0 \frac{3}{4} \left(\frac{\tau_{k(2)}}{1 + \omega^2 \tau_{k(2)}^2} \right) \quad \text{Eq II21}$$

With the (1) and (2) representing the two motional modes.

The possibility to perform self-consistently the fit at multiple fields simultaneously validates the relaxation model used, and moreover allows to increase the precision, due to the better constriction of the model, as discussed in the following section.

2.3 Findings for the GB1 protein

2.3.1 Results

In order to fit our data as explained in the previous chapter, we introduce the code that takes as an entry the relaxation rates at different temperature, and that fits them with the corresponding model. This specific code is shown in the appendix.

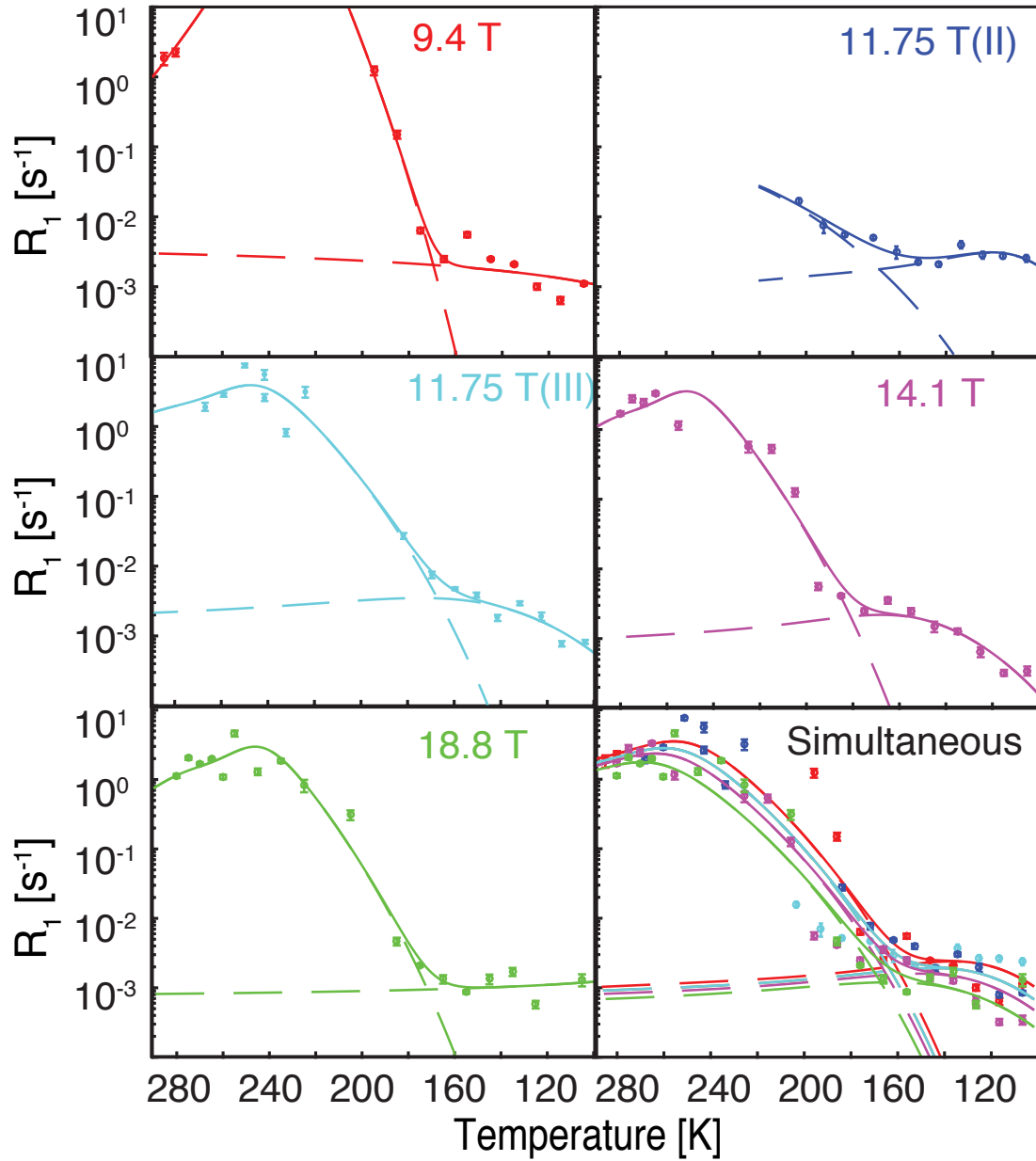


Figure 21 : ^{15}N Lys R_1 rates measured as function of temperature and magnetic field strength. Data were recorded from 105 to 285 K, and at 9.4 T, 11.75 T (set II), 11.75 T (set III), 14.1 T and 18.8 T. The error bars reflect the error on the R_1 fits. Dashed lines show contributions of a given motional mode to relaxation, and the contribution from the sum of the modes is shown as the solid line, which is fit to the relaxation data (points). The number of modes needed (here 2) for the fits was statistically justified with a F-test ($p < 0.05$), except for the 11.75 T set III data, where two modes were also used for comparison purposes. The first five panels show the temperature dependent relaxation fit to the relaxation rates obtained at a single magnetic field. The sixth panel shows the temperature dependent relaxation fit simultaneously to all magnetic fields.

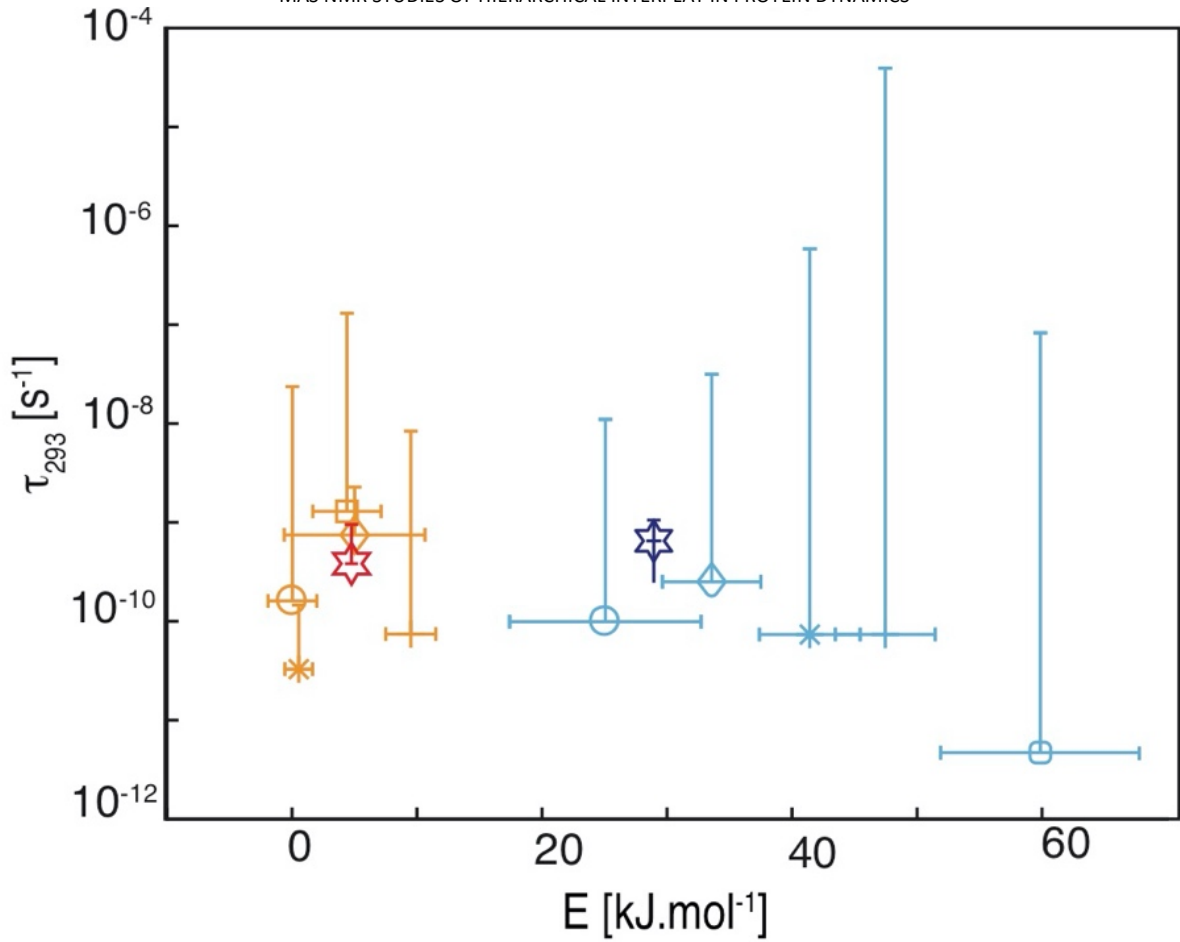


Figure 22 : The activation energies and correlation times at 293K, $\tau_k = \tau_{\infty,k} e^{(E_k/RT)}$, for $^{15}\text{N}_{\text{Lys}}$ extracted from the data in Figure 2. The red and blue symbols refer to the low energy and the high energy mode respectively for 9.4 T (square), 11.75 T (set II) (diamond), 11.75 T (set III) (circle), 14.1 T (cross), and 18.8 T (asterisk), and denote the values extracted from the single field fits. The bold stars show the values extracted from the simultaneous fit to all the data. The error bars are set to \pm one standard deviation, calculated with a Monte Carlo error estimation over 500 runs. Because of the logarithmic scales negative error bars are not displayed.

Figure 21 shows relaxation rates from ^{15}N lysine sidechain moieties, indicating that relaxation shows similar trends as a function of temperature at all the different fields. Analysis of all data in terms of two contributions to the spectral density function reveals clear clustering, both in the correlation times and activation energies of the two extracted motional modes (Figure 22). For $^{15}\text{N}_{\text{Lys}}$ solvent rotations become active at around 160 K, which allows side chains to access different rotameric states on time scales that interfere with the pulse sequence and which leads to the disappearance of the $^{15}\text{N}_{\text{Lys}}$ resonance signals in the region between 160 K to 240 K.^{32, 33} This results in large errors and significant differences in the values extracted from the fits to the motional modes at each field. By combining the information obtained at different fields and fitting them simultaneously we are able to determine the motional parameters with much greater accuracy.

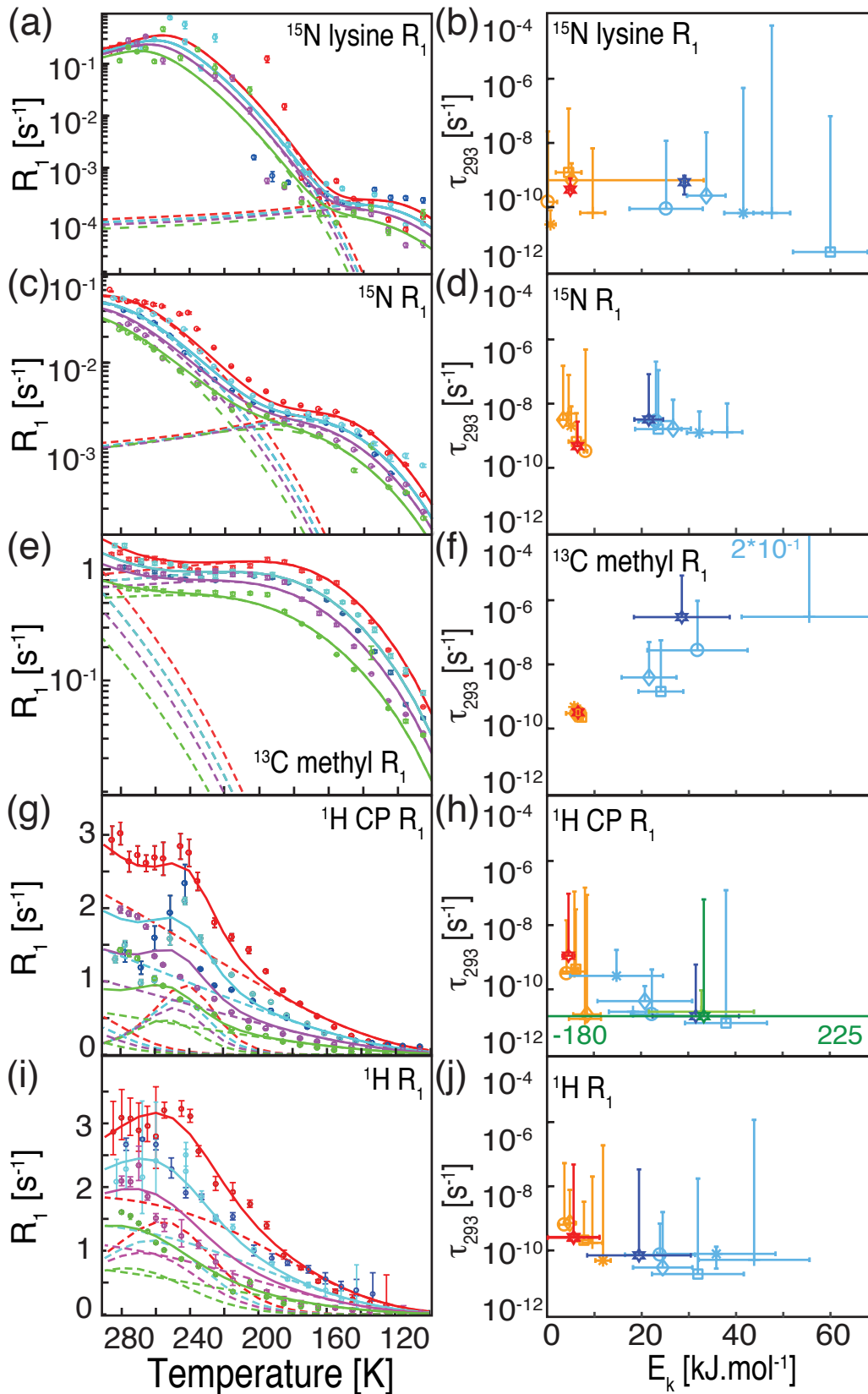


Figure 23. The left panels show the longitudinal relaxation rates as function of temperature for all investigated nuclei. Data were recorded from 105 to 285 K. The error bars reflect the error on the R_1 fits. Dashed lines show contributions of a given motional mode to relaxation, and the contribution from the sum of the modes is shown as the solid line, which is fit to the relaxation data (points) for all fields simultaneously. The colour code is given as: red 9.4 T, dark blue 11.75 T (set II), cyan 11.75 T (set III), purple 14.1 T and green 18.8 T. The number of modes was statistically justified with a F-test ($p < 0.05$). The right panels show the activation

energy plotted against the motional rates τ_{293} for all the investigated relaxation rates. Dark symbols (blue, red and green) show the simultaneous fit over all magnetic fields while the light symbols (cyan, orange and green) show the fits at a single magnetic field. Error bars are set to \pm one standard deviation, calculated with a Monte Carlo estimation over 500 runs. In the right panels the symbols represent: 9.4 T (square), 11.75 T(set II) (diamond), 11.75 T(set III) (circle), 14.1 T (cross), and 18.8 T (asterisk), and denote the values extracted from the single field fits. Bold stars show values extracted from a simultaneous fit to all data. The first modes are represented in orange, the second in blue and the third in green.

Figure 23 summarizes the field-dependent variable-temperature data for all 6 measured relaxation rates. ^{15}N R_1 rates, which report on motion in the protein backbone,³⁴ or the ^{13}C methyl R_1 measurements, reporting on hydrophobic side chains, show very clear field dependencies that are well reproduced in the joint fits. The figure also shows the modes that are found as a result of the joint fits to the data. Again, using the information from all fields simultaneously leads to a much more precise determination of the motional parameters from all the measured rates. The exact details are presented in Table 1. For example, the R_1 maximum is better defined at 9.4 T, leading to a more precise estimate of both timescales and amplitudes, as is evident in Figure 23. Both the decreasing R_1 and the observed shift of R_1 maximum with respect to magnetic field strength closely follow theoretically expected behavior, substantiating the dynamic origin of the relaxation phenomenon.

MAS NMR STUDIES OF HIERARCHICAL INTERPLAY IN PROTEIN DYNAMICS

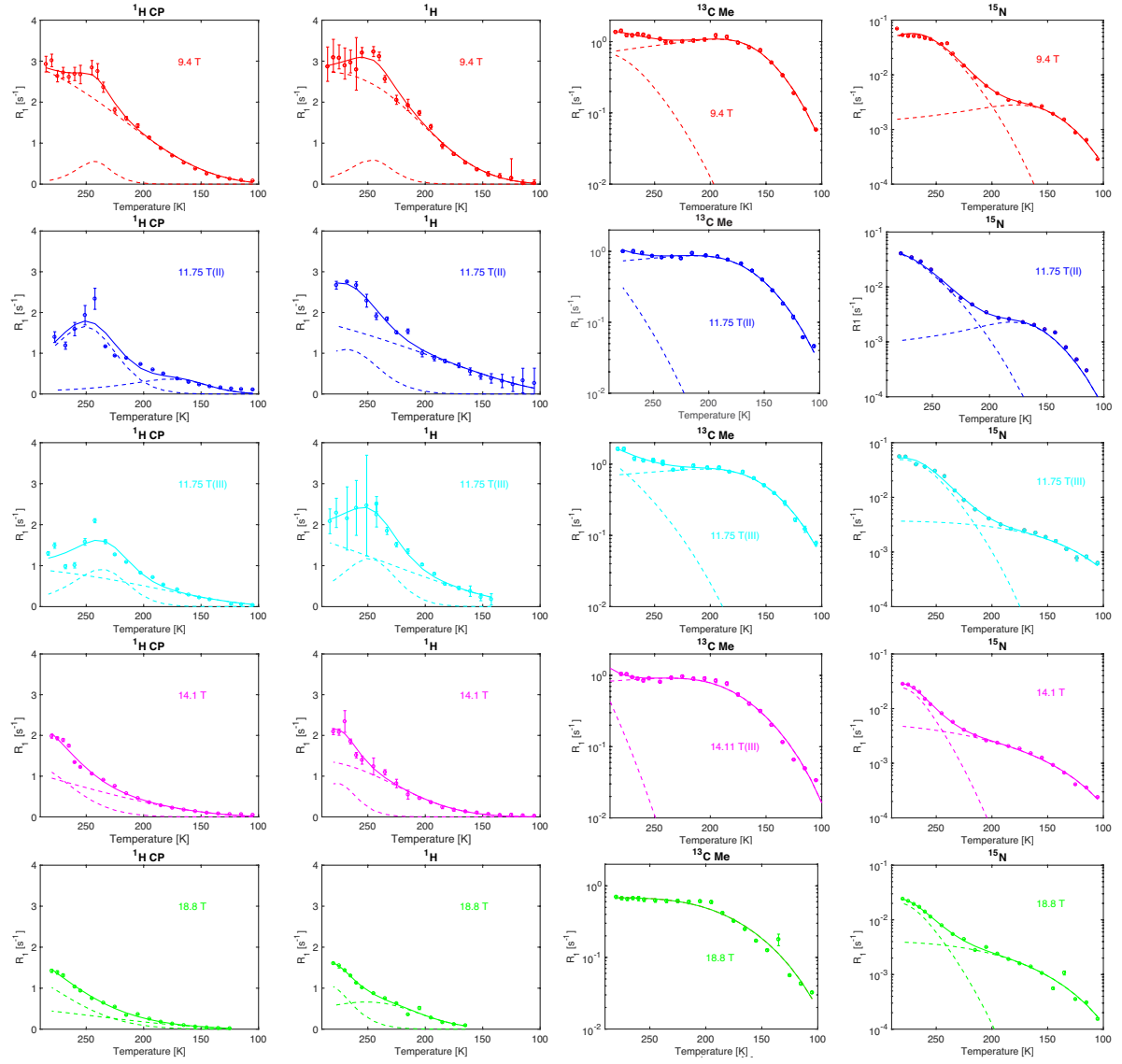


Figure 24: Individual fits of the different observables for GB1

Table 1: Numerical values for the fits of the relaxation parameters. The activation energy and the correlation time refer to the Arrhenius fit, and the C represent the $C_{k,amplitude}$ used in fit as described above. Sd stand for one standard deviation as a measure of error. The single fit are represented in Figure 24.

Mode	Activation Energy (J/mol)	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit	Sd	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit
1	¹ HCP E 1	6.2E+03	4.2E+03	8.5E+03	8.1E+03	5.9E+03	4.6E+03		5.0E+02	1.3E+03	3.4E+03	1.5E+03	1.4E+03	4.7E+02
	¹ Hdirect E 1	7.8E+03	4.8E+03	3.6E+03	9.5E+03	1.2E+04	5.6E+03		1.5E+03	1.3E+03	1.1E+03	2.2E+03	1.5E+03	5.0E+03
	¹³ Cmet E 1	7.3E+03	6.0E+03	6.9E+03	7.0E+03	5.7E+03	6.5E+03		1.9E+02	2.0E+03	1.5E+02	2.0E+02	2.0E+02	1.7E+02
	¹⁵ N R1 E 1	6.2E+03	3.4E+03	8.1E+03	4.5E+03	5.1E+03	6.4E+03		1.5E+03	7.2E+02	3.9E+02	1.5E+02	1.2E+02	6.1E+02
	¹⁵ N Lys E 1	4.4E+03	5.0E+03	3.2E+01	9.5E+03	5.4E+02	4.8E+03		2.8E+03	2.9E+04	1.9E+03	3.0E+03	1.1E+03	3.0E+02
2	¹ HCP E 2	3.8E+04	2.2E+04	2.1E+04	1.8E+04	1.5E+04	3.2E+04		8.5E+03	4.4E+03	1.0E+04	4.2E+03	3.7E+03	1.9E+03
	¹ Hdirect E 2	3.2E+04	2.4E+04	2.4E+04	4.4E+04	3.6E+04	1.9E+04		1.1E+04	6.2E+03	7.3E+03	1.1E+04	1.4E+04	1.2E+04
	¹³ Cmet E 2	2.4E+04	2.2E+04	3.2E+04	5.6E+04	3.4E+05	2.9E+04		4.7E+03	5.3E+03	1.1E+04	1.5E+04	3.7E+04	9.6E+03
	¹⁵ N R1 E 2	2.4E+04	2.7E+04	2.3E+04	3.8E+04	3.2E+04	2.2E+04		5.0E+03	3.8E+03	3.7E+03	3.3E+03	2.6E+03	3.0E+03
	¹⁵ N Lys E 2	6.0E+04	3.4E+04	2.5E+04	4.7E+04	4.1E+04	2.9E+04		7.9E+03	4.0E+03	7.1E+03	4.1E+03	4.1E+03	1.5E+02
3	¹ HCP E 3	0.0E+00	0.0E+00	0.0E+00	3.3E+04	0.0E+00	3.3E+04		0.0E+00	0.0E+00	0.0E+00	5.4E+03	0.0E+00	9.9E+04
Mode	Correlation Time (s)	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit	Sd	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit
1	¹ HCP τ 1	3.6E-11	5.9E-11	5.7E-13	4.9E-13	3.5E-11	1.8E-10		6.1E-10	2.5E-09	3.1E-09	7.7E-12	9.3E-11	3.1E-10
	¹ Hdirect τ 1	9.4E-12	1.1E-10	1.5E-10	3.8E-12	4.3E-13	2.9E-11		1.8E-09	1.1E-09	1.2E-08	1.4E-09	1.5E-09	5.1E-09
	¹³ Cmet τ 1	1.3E-11	3.0E-11	2.1E-11	2.4E-11	5.2E-11	2.5E-11		2.5E-12	8.3E-12	3.0E-12	4.8E-12	1.1E-11	4.6E-12
	¹⁵ N τ 1	5.7E-11	8.5E-10	1.3E-11	8.2E-10	2.8E-10	3.8E-11		4.0E-10	3.8E-08	2.0E-08	1.1E-08	9.1E-10	1.7E-10
	¹⁵ N Lys τ 1	2.1E-10	9.5E-11	1.6E-10	1.5E-12	2.6E-11	5.4E-11		2.0E-08	2.1E-10	2.4E-08	1.0E-10	5.0E-11	6.6E-11
2	¹ HCP τ 2	1.7E-18	2.0E-15	9.4E-15	1.2E-14	6.6E-13	3.9E-17		1.5E-14	5.1E-14	1.5E-14	1.1E-13	1.9E-12	7.4E-18
	¹ Hdirect τ 2	4.3E-17	1.5E-15	4.5E-15	8.8E-19	3.6E-17	2.7E-14		3.5E-14	8.3E-14	3.5E-14	1.9E-14	2.3E-17	9.0E-14
	¹³ Cmet τ 2	8.2E-14	6.2E-13	6.2E-14	4.0E-17	8.1E-12	2.5E-12		2.9E-12	6.6E-12	3.4E-12	4.0E-12	3.7E-12	5.3E-11
	¹⁵ N τ 2	1.1E-13	3.2E-14	2.4E-13	2.2E-16	2.4E-15	5.0E-13		6.1E-12	2.4E-13	2.0E-11	3.8E-14	1.9E-14	1.2E-11
	¹⁵ N Lys τ 2	1.0E-22	2.6E-16	3.4E-15	2.5E-19	3.0E-18	4.5E-15		1.6E-18	2.6E-14	3.8E-13	2.3E-14	1.9E-14	2.6E-15
3	¹ HCP τ 3	0.0E+00	0.0E+00	0.0E+00	1.5E-16	0.0E+00	4.4E-16		0.0E+00	0.0E+00	0.0E+00	1.2E-15	0.0E+00	2.1E-15
Mode	C	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit	Sd	400 MHz	500 MHz set III	500 MHz set II	600 MHz	800 MHz	Simultaneous fit
1	¹ HCP Ck 1	5.9E+09	2.9E+09	8.1E+08	4.0E+08	2.7E+09	9.0E+09		3.9E+10	1.2E+11	6.7E+10	7.6E+08	6.8E+09	1.2E+10
	¹ Hdirect Ck 1	4.8E+09	6.9E+09	6.8E+09	3.7E+09	2.3E+09	3.3E+09		1.9E+11	8.4E+10	2.8E+11	5.6E+11	1.5E+12	1.8E+11
	¹³ Cmet Ck 1	6.5E-02	3.1E-01	6.5E-02	8.2E-02	8.2E-02	7.0E-02		2.0E-03	5.7E-02	2.3E-03	5.6E-03	3.6E-03	3.5E-03
	¹⁵ N Ck 1	7.7E-04	1.2E-03	7.3E-04	2.3E-03	1.6E-03	6.8E-04		3.6E-01	4.0E-02	2.9E-01	2.8E-02	4.2E-03	7.0E-04
	¹⁵ N Lys Ck 1	9.3E-04	1.4E-03	3.5E-03	1.1E-03	4.8E-03	7.4E-04		3.4E-02	1.7E-03	9.0E-03	3.2E-03	3.4E-03	2.1E-04
2	¹ HCP Ck 2	9.9E+08	2.4E+09	3.7E+09	1.6E+09	5.7E+09	1.7E+09		5.0E+09	8.8E+08	3.6E+10	1.2E+09	1.4E+10	2.4E+08
	¹ Hdirect Ck 2	1.1E+09	2.6E+09	2.4E+09	2.2E+09	3.8E+09	2.6E+09		2.0E+10	9.0E+08	9.7E+08	9.7E+10	2.1E+11	1.4E+11
	¹³ Cmet Ck 2	4.3E-02	9.2E-01	6.3E-01	8.3E+00	2.0E-01	7.0E+00		1.7E-01	1.8E-01	8.4E+01	8.4E+01	1.8E-01	1.5E+02
	¹⁵ N Ck 2	1.5E-02	1.6E-02	1.5E-02	8.7E-03	8.4E-03	1.6E-02		2.3E-03	2.9E-03	2.6E+01	1.1E-01	1.9E-02	1.3E-01
	¹⁵ N Lys Ck 2	3.4E+01	1.5E+00	8.2E-03	1.8E+00	1.8E+00	1.1E+00		1.9E+01	4.3E-01	3.2E+00	1.9E-01	2.2E-01	1.2E-01
3	¹ HCP Ck 3	0.0E+00	0.0E+00	0.0E+00	4.7E+09	0.0E+00	1.0E+09		0.0E+00	0.0E+00	0.0E+00	1.4E+10	0.0E+00	2.5E+09

2.3.2 Conclusion

This analysis illustrates the problems associated with describing the dynamics by a so-called “dynamic transition temperature”. ¹⁵N R_1 and ¹³C methyl R_1 relaxation measurements clearly show that the observed “crossover” temperature, where the higher activation energy mode becomes the dominant source of relaxation, changes as a function of the magnetic field, as expected from theory. This

highlights that the apparent “transition temperatures” are properties that depend both on the physical probe and the intrinsic properties of the protein that are better described in terms of the energy barriers to the motion.³⁵ We note that all but three of the curves can be explained with two motional modes.

The $^1\text{H}_{\text{CP}}$ R_1 data justifies the use of three modes. This was also observed in the previous study at a single field and is validated here by the multi-field data that fits to three modes with energy barriers of 5, 34 and 35 kJmol^{-1} . These rates are influenced by both the protein, and the crystal water, which implies that they may indeed be different from, for example, the $^{13}\text{C}_{\text{methyl}}$ R_1 which report mainly on the hydrophobic sidechain dynamics.³⁶

Figure 25 compares the two activation energies for the different modes extracted from the multi-field analysis (numerical values are given in Table 1). Note that in the analysis we use here we obtain a single activation energy for each mode per residue type. This description is obviously an approximation, since the relaxation rates measured are almost certainly due to a finite distribution of motions over a range of timescales and amplitudes. The activation energies and amplitudes obtained should thus be considered as weighted averages. The energies clearly cluster, with a low energy barrier between 5 and 10 kJmol^{-1} and a high-energy barrier between 20 and 40 kJmol^{-1} . These results are in good agreement with our previous work,³⁶ substantiating, in accordance with previous publications,³⁵⁻³⁷ the identification of two to three main motional modes per component: a uniform low energy mode ($<10 \text{ kJmol}^{-1}$) and one to two high energy ($>15 \text{ kJmol}^{-1}$) modes per component. We note that the activation energies of the second mode are not as similar among the different probes as for the first mode. This apparent inhomogeneity can be tentatively assigned to more variability in the high-energy motions as compared to the low energy motions which appear more homogenous across the protein. Overall these observations are consistent with our previous assignment of the high activation energy components to anisotropic motions of peptide planes or sidechain moieties that dominate solution state spin relaxation in proteins, while the low activation energy components appear to report on smaller amplitude librational motions.

In summary, we have presented a variable temperature, multi-field NMR relaxation approach to disentangle motional modes governing protein dynamics, and used temperature-dependent MD simulation to support the interpretation. The combination of multi-field relaxation rates and MD simulation substantiates the interpretation of ^{15}N R_1 relaxation in terms of backbone dynamics and more accurately defines the activation energies and correlation times of the associated motional modes.

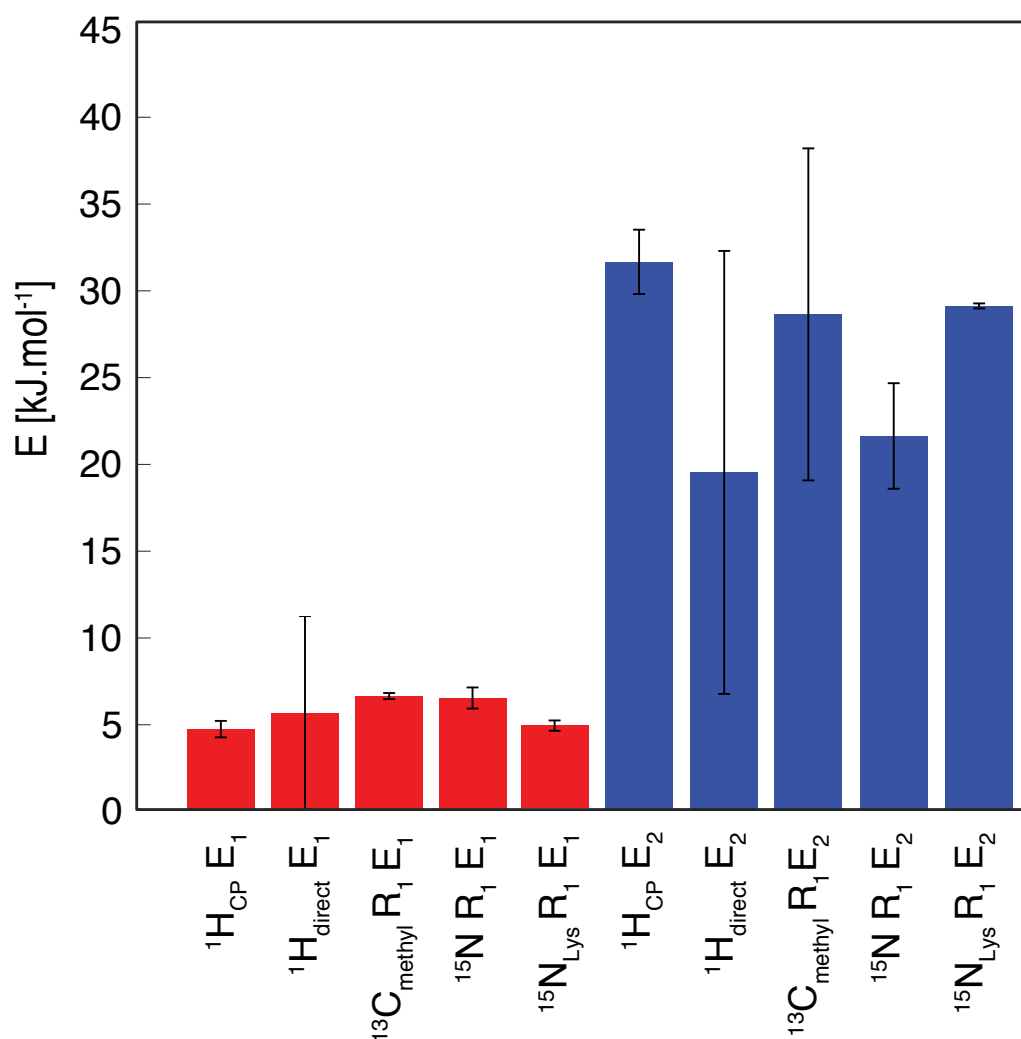


Figure 25 : Activation energies from the simultaneous fits, plotted for each nuclear spin probe. The error bars are set to one standard deviation, calculated with a Monte Carlo error estimation over 500 runs. The red bars represent the energies of the first mode, the blue bars the energies of the second mode. The third mode of the ¹H_{CP} R₁ is not plotted. The energies for ¹³C' R₁ are not reported since as described in the text spin diffusion renders them equivalent to the ¹³C_{methyl} R₁.

2.4 References

1. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct observation of hierarchical protein dynamics. *Science* **2015**, 348 (6234), 578-581.
2. Wangsness, R. K.; Bloch, F., The Dynamical Theory of Nuclear Induction. *Phys Rev* **1953**, 89 (4), 728-739.
3. Schanda, P.; Ernst, M., Studying Dynamics by Magic-Angle Spinning Solid-State NMR Spectroscopy: Principles and Applications to Biomolecules. *Prog Nucl Magn Reson Spectrosc* **2016**, 96, 1-46.
4. Goldman, M., *Quantum description of high-resolution NMR in liquids*.
5. Abragam, A., *The principles of nuclear magnetism*. Oxford university press: 1961.
6. Redfield, A. G., Relaxation Theory: Density Matrix Formulation. In *Emagres*, John Wiley & Sons, Ltd: 2007.
7. Redfield, A. G., On the Theory of Relaxation Processes. *Ibm J Res Dev* **1957**, 1 (1), 19-31.
8. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *Journal of the American Chemical Society* **1982**, 104 (17), 4546-4559.
9. Keeler, J., *Understanding NMR Spectroscopy*. Wiley: 2011.

10. J. Cavanagh, W. J. F., A. G. I. Palmer, N. J. Skelton, M. Rance, *Protein NMR Spectroscopy*. 2006.
11. Beckmann, P. A.; Buser, C. A.; Mallory, C. W.; Mallory, F. B.; Mosher, J., Methyl reorientation in solid 3-ethylchrysene and 3-isopropylchrysene. *Solid State Nucl Mag* **1998**, *12* (4), 251-256.
12. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Protein dynamics. Direct observation of hierarchical protein dynamics. *Science* **2015**, *348* (6234), 578-81.
13. Frederick, K. K.; Kranz, J. K.; Wand, A. J., Characterization of the backbone and side chain dynamics of the CaM-CaMKIIp complex reveals microscopic contributions to protein conformational entropy. *Biochemistry-Us* **2006**, *45* (32), 9841-8.
14. Tycko, R., NMR at low and ultralow temperatures. *Acc Chem Res* **2013**, *46* (9), 1923-32.
15. Jacso, T.; Franks, W. T.; Rose, H.; Fink, U.; Broecker, J.; Keller, S.; Oschkinat, H.; Reif, B., Characterization of membrane proteins in isolated native cellular membranes by dynamic nuclear polarization solid-state NMR spectroscopy without purification and reconstitution. *Angew Chem Int Edit* **2012**, *51* (2), 432-5.
16. Ni, Q. Z.; Markhasin, E.; Can, T. V.; Corzilius, B.; Tan, K. O.; Barnes, A. B.; Daviso, E.; Su, Y.; Herzfeld, J.; Griffin, R. G., Peptide and Protein Dynamics and Low-Temperature/DNP Magic Angle Spinning NMR. *The Journal of Physical Chemistry B* **2017**, *121* (19), 4997-5006.
17. Debelouchina, G. T.; Bayro, M. J.; Fitzpatrick, A. W.; Ladizhansky, V.; Colvin, M. T.; Caporini, M. A.; Jaroniec, C. P.; Bajaj, V. S.; Rosay, M.; Macphee, C. E.; Vendruscolo, M.; Maas, W. E.; Dobson, C. M.; Griffin, R. G., Higher order amyloid fibril structure by MAS NMR and DNP spectroscopy. *J Am Chem Soc* **2013**, *135* (51), 19237-47.
18. Lamley, J. M.; Lewandowski, J. R., Relaxation-Based Magic-Angle Spinning NMR Approaches for Studying Protein Dynamics. *Emagres* **2016**, *5* (3), 1423-1433.
19. Mollica, L.; Baia, M.; Lewandowski, J. R.; Wylie, B. J.; Sperling, L. J.; Rienstra, C. M.; Emsley, L.; Blackledge, M., Atomic-resolution structural dynamics in crystalline proteins from NMR and molecular simulation. *J. Phys. Chem. Lett.* **2012**, *3* (23), 3657-62.
20. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
21. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65* (3), 712-25.
22. Bockmann, A.; Gardienet, C.; Verel, R.; Hunkeler, A.; Loquet, A.; Pintacuda, G.; Emsley, L.; Meier, B. H.; Lesage, A., Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. Nmr.* **2009**, *45* (3), 319-27.
23. Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M., Magic-angle spinning solid-state NMR spectroscopy of the beta1 immunoglobulin binding domain of protein G (GB1): ¹⁵N and ¹³C chemical shift assignments and conformational analysis. *J. Am. Chem. Soc.* **2005**, *127* (35), 12291-305.
24. Busi, B.; Yarava, J. R.; Hofstetter, A.; Salvi, N.; Cala-De Paepe, D.; Lewandowski, J. R.; Blackledge, M.; Emsley, L., Probing Protein Dynamics Using Multifield Variable Temperature NMR Relaxation and Molecular Dynamics Simulation. *J Phys Chem B* **2018**, *122* (42), 9697-9702.
25. Neue, G.; Dybowski, C., Determining temperature in a magic-angle spinning probe using the temperature dependence of the isotropic chemical shift of lead nitrate. *Solid State Nucl Mag* **1997**, *7* (4), 333-336.
26. Encyclopedia of Mathematics.
27. Beckmann, P. A.; Buser, C. A.; Mallory, C. W.; Mallory, F. B.; Mosher, J., Methyl reorientation in solid 3-ethylchrysene and 3-isopropylchrysene. *Solid State Nucl. Magn. Reson.* **1998**, *12* (4), 251-6.
28. Stone, M., Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc B* **1974**, *36* (2), 111-147.
29. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct observation of hierarchical protein dynamics. *Science* **2015**, *348* (6234), 578.
30. Bloembergen, N., On the Interaction of Nuclear Spins in a Crystalline Lattice. *Physica* **1949**, *15* (3-4), 386-426.
31. Giraud, N.; Blackledge, M.; Bockmann, A.; Emsley, L., The influence of nitrogen-15 proton-driven spin diffusion on the measurement of nitrogen-15 longitudinal relaxation times. *J Magn Reson* **2007**, *184* (1), 51-61.
32. Lee, A. L.; Wand, A. J., Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* **2001**, *411* (6836), 501-4.
33. Ni, Q. Z.; Markhasin, E.; Can, T. V.; Corzilius, B.; Tan, K. O.; Barnes, A. B.; Daviso, E.; Su, Y.; Herzfeld, J.; Griffin, R. G., Peptide and protein dynamics and low-temperature/DNP magic angle spinning NMR. *J. Phys. Chem. B* **2017**, *121* (19), 4997-5006.
34. Lipari, G.; Szabo, A., A model-free approach to the interpretation of NMR relaxation in macromolecules. *Biophys. J.* **1981**, *33* (2), 307-307.
35. Fenimore, P. W.; Frauenfelder, H.; Magazu, S.; McMahon, B. H.; Mezei, F.; Migliardo, F.; Young, R. D.; Strope, I., Concepts and problems in protein dynamics. *Chem Phys* **2013**, *424*, 2-6.
36. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct observation of hierarchical protein dynamics. *Science* **2015**, *348* (6234), 578-81.

37. Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Strope, I. R.; Swenson, J.; Young, R. D., A unified model of protein dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (13), 5129-5134.
- 38 Moro, G.; Freed, J. H., Efficient Computation of Magnetic-Resonance Spectra and Related Correlation-Functions from Stochastic Liouville Equations. *J Phys Chem-Us* **1980**, *84* (22), 2837-2840.

Chapter 3 Similarities and differences among protein dynamics

In the previous chapter we described and developed on a method using variable temperature relaxation measurements to show how it is possible to gain insights about the energy landscape of proteins, with a discrimination within the different types of protein probes (backbones, solvent and side chain of specific hydrophobicity).

In this chapter we will analyse to which extent the previous findings can be generalized to other types of proteins, and if differences in the classes of proteins can be associated with changes in the energy landscape. This chapter is adapted from reference.²

3.1 Introduction

Large differences in both structure and function among proteins of different families are observed, and despite the inevitable presence of dynamic heterogeneity at the local level due to the presence of specific bindings sites or loops, overall protein dynamics appears to exhibit many common general features. In particular the coupling between protein regions and the solvent, which causes a second order dynamic transition, has been postulated to lead to ‘universal’ properties.³

Here we explore this concept by a comparative study of the similarities and differences in activation energies of different processes and the associated dynamic transitions of four different proteins in three principally different kinds of sample conditions, using the variable-temperature variable-field relaxation approach presented in the previous chapter. Specifically, we study microcrystalline preparations of two small globular proteins, GB1 and the SH3 domain of α -spectrin, which appear as a solid at all temperatures. The soluble N-terminal tail of the nucleoprotein of the Sendai virus (Sendai Ntail), which is an intrinsically disordered protein, is dissolved in buffer and behaves as a solid upon sample flash freezing in liquid N₂. The membrane protein OmpG is incorporated into lipid bilayers composed of *E.coli* lipids and thus also accessible to solid-state NMR methods at all temperatures. We measure five ¹H, ¹³C and ¹⁵N NMR longitudinal relaxation rates (R_1) at multiple magnetic fields (from 400 to 900 MHz) and at temperatures in the range 100 K to 300 K. We used the approach

described in the previous chapter and in the literature.^{1, 2} The correlation between the R_1 , and the temperature and field allows to qualitatively map the thermodynamical landscape describing protein dynamics.

For this comparison we determine the hierarchical activation energies of motions of the four different proteins described above. The activation energies map to motions occurring in the sidechains, in the backbone, and in the hydration shells of the proteins. We were able to identify similarities and differences in the average motions of the proteins. We find that the proteins do share similar features. The average behaviour of backbone motions is found to be very similar across the series of proteins, as are methyl group rotations, suggesting that proteins exhibit universal dynamic features. For these motions, similar activation energies are measured for all four proteins. The main observed difference occurs for the intrinsically disordered Sendai Ntail protein, where we observe much lower energy of activation for protons associated with the protein-solvent interface as compared to the others. We also observe variability between the proteins in the hydrophilic sidechains, with the onset of slow motions in OmpG occurring with lower energies than in the other three proteins.

3.2 Experimental

3.2.1 Samples

GB1 (56 amino acids), an antibody-binding protein was previously used to introduce the methodology used here and to assess the reproducibility of the experiments and the robustness of the model². The protein was studied as a micro-crystalline precipitate. The water to protein to MPD/IPA ratio in the GB1 sample was 1 : 0.03 : 0.59. The data shown here are reproduced from ref ². Those data have also been presented alone in the previous chapter, no new data were recorded for the comparison between the different proteins.

The Src-homology 3 (SH3) domain of α -spectrin (62 amino acids) is studied as microcrystals grown from H₂O, without cryoprotectant, and served as test case for a MAS structure determination methodology. Its structure and sequence are shown in Figure 26.

Both GB1 and SH3 samples have the characteristics that the crystal lattice is expected to impose some rigidity and reduce solvent accessibility in addition to restriction in the hydrophobic core, particularly in crystal contact areas.

The Ntail of the nucleoprotein of the Sendai virus, referred as Sendai Ntail, is an intrinsically disordered protein (IDP) with 126 amino acids. The sample was studied as a frozen solution with a concentration of 750 μ M. The protein has no hydrophobic core, and the whole protein may be considered as solvent-exposed.

OmpG is a membrane protein, with 301 amino acids, here inserted into an *E.coli* lipid bilayer. Two samples of OmpG were used for the measurements.

The ^1H , ^{13}C , ^{15}N labelled, microcrystalline SH3 protein, generously provided by our collaborators, was overexpressed and purified as described by Castellani et al.⁴ Briefly, crystals were grown in a 296 large scale sitting drop (500 μ l) with the hanging drop vapour diffusion method using 100 mM ammonium sulphate titrated to a pH value from 3.5 to 7.5. The sample is then centrifugated three times on a bench Eppendorf centrifuge to remove all the excess water, and packed, estimates of 5 mg in the rotor.

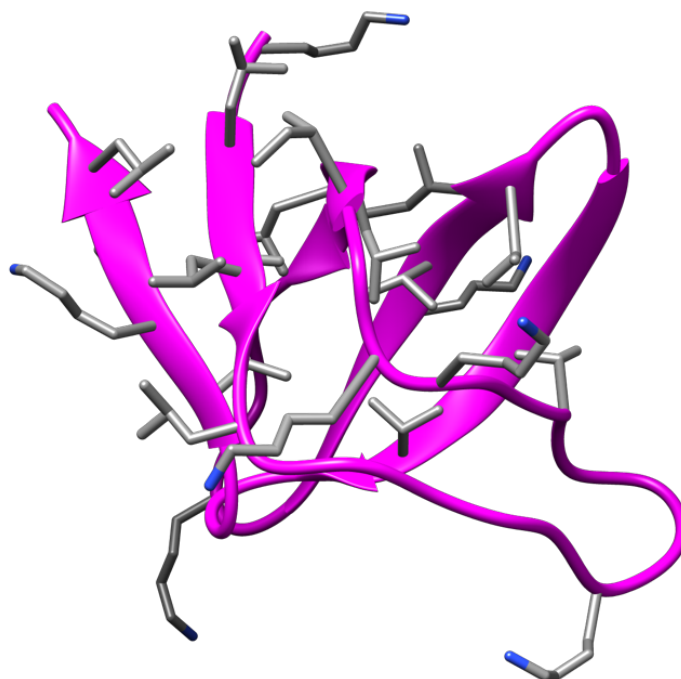


Figure 26: Structure of SH3. The aliphatic chains are in grey and the Lysine are in blue. The full sequence is MDETGKELVLALYDYQEKSPREVTMKGDI LLLNSTNKDWWKVEVND RQGFVPAAYVKKLD

We may anticipate problems with ice crystal formation since there is no specific cryoprotectant, but we did not see any changes in the sample after one full temperature cycle, as shown in Figure 27. This resistance to freeze cycling has been previously observed for SH3.⁵

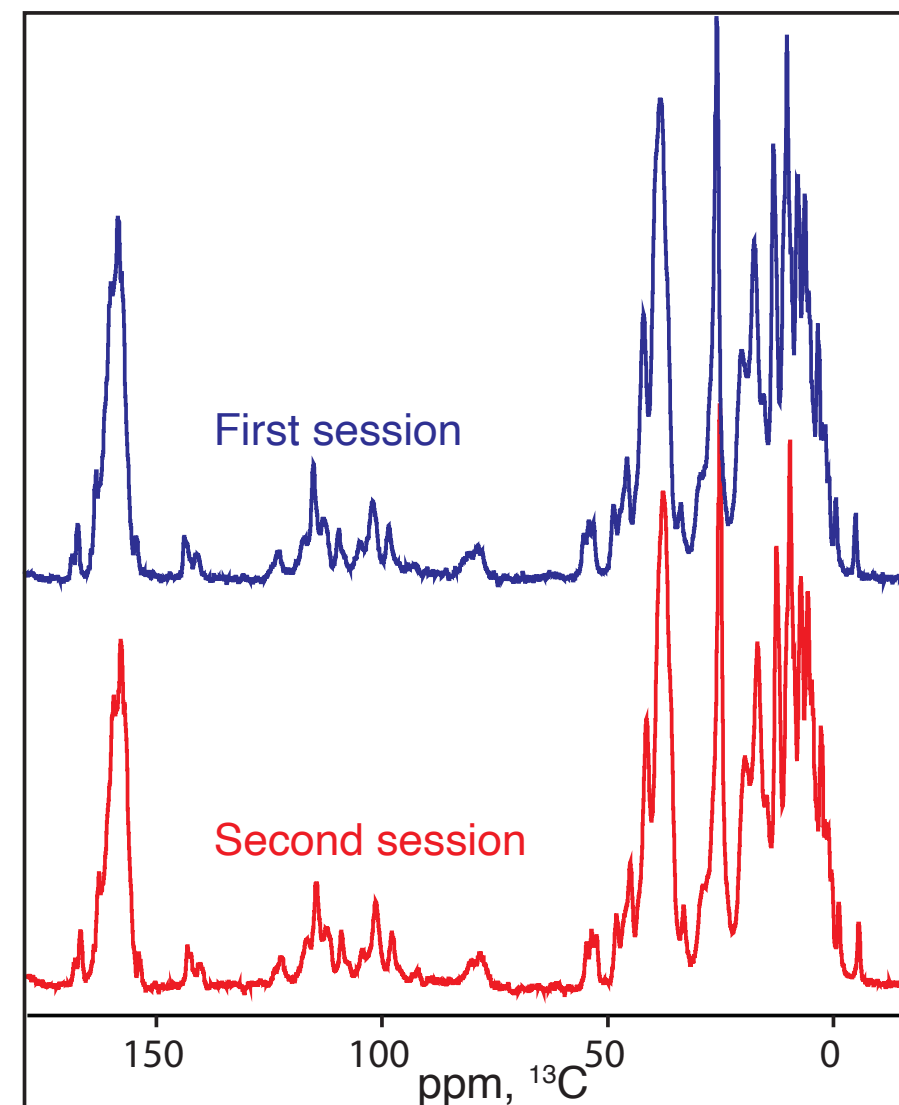


Figure 27: Comparison of 1D ^{13}C spectra of the same batch of SH3, at room temperature. First session a fresh sample, and second session after a 100 K-280 K cycle.

The intrinsically disordered Ntail of the large protein of the Sendai virus (Sendai Ntail in the following), is a 13.39 kDa, 126 amino-acids intrinsically disordered protein (IDP). It acquires a specific structure when in contact with specific ligands, which were not present in our sample in order to study the intrinsically disordered state.⁶ The structure is presented in Figure 26.

Two different samples with specific preparation were used. A ^1H , ^{13}C uniformly labeled and a ^1H , ^{13}C , ^{15}N uniformly labelled one. The Sendai Ntail fragments were expressed as described previously by Blanchard et al.⁶ The solution contained $\sim 750\ \mu\text{M}$ of protein in a pH6, 50 mM sodium phosphate and 500 mM NaCl buffer. The solution was flash frozen while inside the rotor using liquid N_2 before measurement.

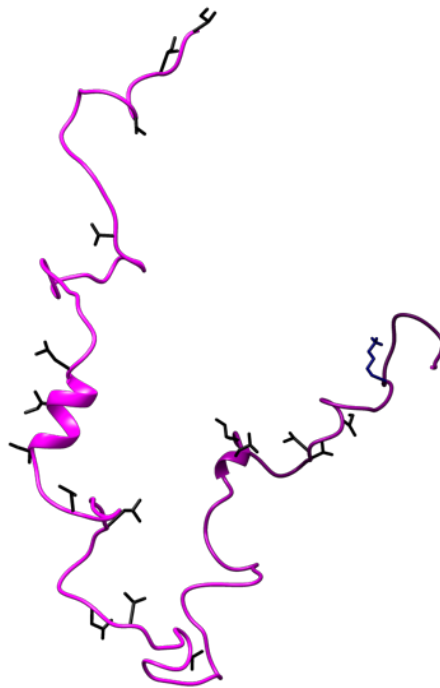


Figure 28: One of the many structures of the Sendai Ntail (all of them have this typical lack of secondary structure). The Lysines are in blue and the aliphatic chains are in grey. The full sequence is GHMSGG-DGAYHKPTGGGAIEVALDNADIDLE-TEAHADQDARGWGGESGERWARQVSGGHFVTLHGAERLEEETNDEDVSDIERRIAMRLAERRQEDSATHGDEGRNNGVDHDEDDDDAAAVAGIGGI

The OmpG protein is a selective channel naturally found in bacteria. This membrane protein is 286 amino acids and its molecular weight is 32.91 kDa. Its properties are dependent on its environment and will be studied with the protein inserted in a lipid bilayer. The structure and sequences are presented in Figure 29.

The ^1H , ^{13}C , ^{15}N labelled OMPG was kindly produced by our collaborators. The expression is as described in Hiller et al.⁷ For refolding, purified OmpG was diluted into a buffer containing n-dodecyl- β -D-maltoside DDM (1 mM) and l-arginine (0.6 M) at pH 8.8. The refolded protein was washed and concentrated with buffer 10 mM Tris-HCl, pH 8.0 and DDM (1 mM) to a final concentration of 1–2 mg/mL. The detergent concentration was reduced by binding OmpG to a Resource-Q column and washing with 3 column volumes of buffer 10 mM Tris-HCl, pH 8.0 with DDM (0.4 mM). OmpG was eluted with NaCl (0.3 M) and concentrated by a centrifugal filter device final concentration 3 mg/mL).

Refolded OmpG was reconstituted into lipid bilayers. For this purpose, an *E. coli* total lipid chloroform extract (20 mg) was dried in a nitrogen stream. The resulting lipid film was dissolved in buffer 10 mM Tris-HCl, pH 8.0 containing n-octyl- β -D-glucopyranoside 34 mM. Aliquots of this lipid solution and refolded OmpG (2 mg/mL) were mixed to yield a lipid-to-protein ratio (LPR) of 1:2 (w/w). For 2D crystallization, the detergent was removed by dialysis (dialysis-tube cut-off 25 kDa, Roth, Karlsruhe, Germany) at 20 °C against buffer B (5 L, 20 mM Tris-HCl, pH 7.0, 25 mM MgCl₂, 3 mM NaN₃, 150 mM NaCl) for 6 to 7 weeks. The dialysis buffer was changed every 5 days.

No specific cryoprotectant was used. This was not an issue since the sample was only subjected to one series of measurements with decreasing temperature.

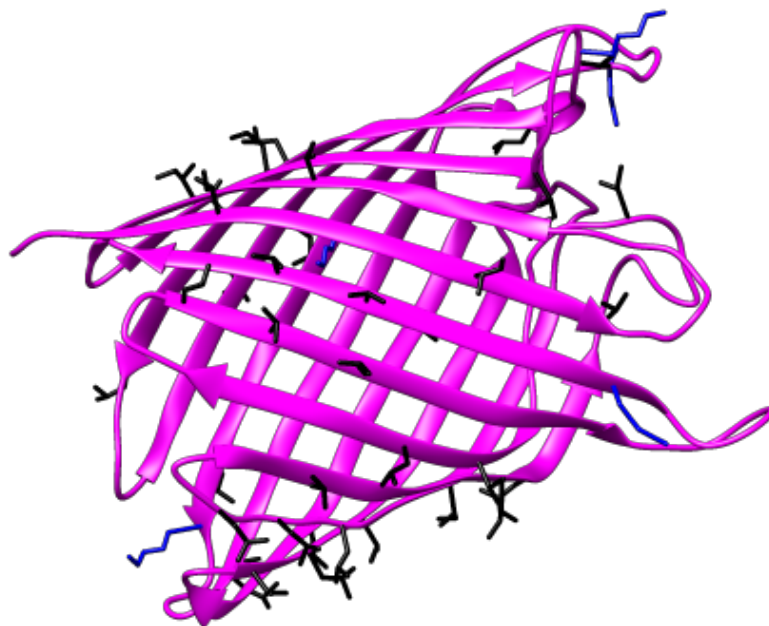


Figure 29: Structure of OmpG. The aliphatic chains are in grey and the Lysine are in blue. The full sequence is MEERNDWHFNIGAMYEIENVEGYGEDMDGLAEPVYFNAANGPWRIALAYYQEGPVDYS-AGKRGTWFD RPELEVHYQFLENDDFSGLTGGFRNYGYHYVDEPGKDTANMQRWKIAPDWVDVKLTDDLRFNGWLSMYKFANDLNTTGYADTRVETETGLQYTFNETVALRVNYYLARGFNMDDSRNNGEFSTQEI-RAYLPLTLGNHSVTPYTRIGLDRWSNWDWQDDIEREGHDFNRVGLFYGYDFQNGLSVSLEYAFEWQDHDEGSDSKFHYA GVGVNYSF

Measurements on SH3 were done on a single batch, first at 11.75 T and then at 21.1 T. The Sendai Ntail measurements were done on two different batches, one for ^{15}N and ^1H relaxation and one for ^{13}C . for both SH3 and Sendai Ntail, the rotors were frozen to 100 K and the experiments were made with increasing temperature. For OmpG the experiments were performed at 280 K first and then in decreasing temperature order.

All proteins were uniformly ^{13}C , ^{15}N labelled, except one batch of Sendai Ntail samples which was only ^{13}C - labelled.

The water and lipid content of the samples as packed into the rotors was evaluated qualitatively through signal integrals. The water to protein to MPD/IPA molar ratio in GB1 was 1 : 0.03 : 0.59. The water : protein ratio in SH3 was 1 : 0.04. The water : protein ratio in Sendai was not measurable, and below 1 : 0.01. The original sample in this case is unchanged by sample packing, and was 750 $\mu\text{mol/L}$. Two samples of OmpG were used for the measurements. The water : lipid : protein ratio in the first sample was 1 : 0.06 : 0.03 and in the second sample was 1 : 1.17 : 0.02.

3.2.2 NMR experiments, measurements and activation energies:

The approach has been described in the previous chapter as well as in the literature.^{1, 2} The main points are highlighted below.

We measured spin lattice relaxation estimated from integration of defined regions of ^1H , ^{15}N and ^{13}C spectra from 100 to 290K. This allows us to have five observables: bulk solvent, crystal water/protein protons, backbones, hydrophilic side chain and hydrophobic side chains.

Using the well-established equation⁸ and a previously described protocol,¹ as discussed in the previous chapter, it is possible to explain changes in relaxation as a function of temperature with changes of dynamics as a function of temperature.

This allows to map the thermodynamical landscape of the motions separately for each individual probe.

3.3 Results

3.3.1 Relaxation Rates description

For each sample we measured: ^1H R_1 from ^1H spectra, ^1H R_1 on ^{13}C through cross polarization (CP), methyl ^{13}C R_1 , amide ^{15}N R_1 and lysine $^{15}\text{N}_\epsilon$ R_1 .

Data were taken at 11.7 T (500 MHz ^1H frequency) and 21.1 T (900 MHz ^1H frequency) for SH3, Sendai Ntail and only 11.7 T for OmpG, while the data for GB1 were taken at 9.4, 11.7, 14.1 and, 18.8 T and are here reproduced from ref. ^{1, 2} For GB1, SH3 and OmpG data were taken over a temperature range from 100 to 280 K. The Ntail sample melts at around 250 K and therefore transits to rapid overall isotropic motion, and consequently the signals in CP-excited spectra disappear, since CP is not efficient in this motional regime. This sets an upper limit on temperature for Sendai Ntail for the measurements with our approach.

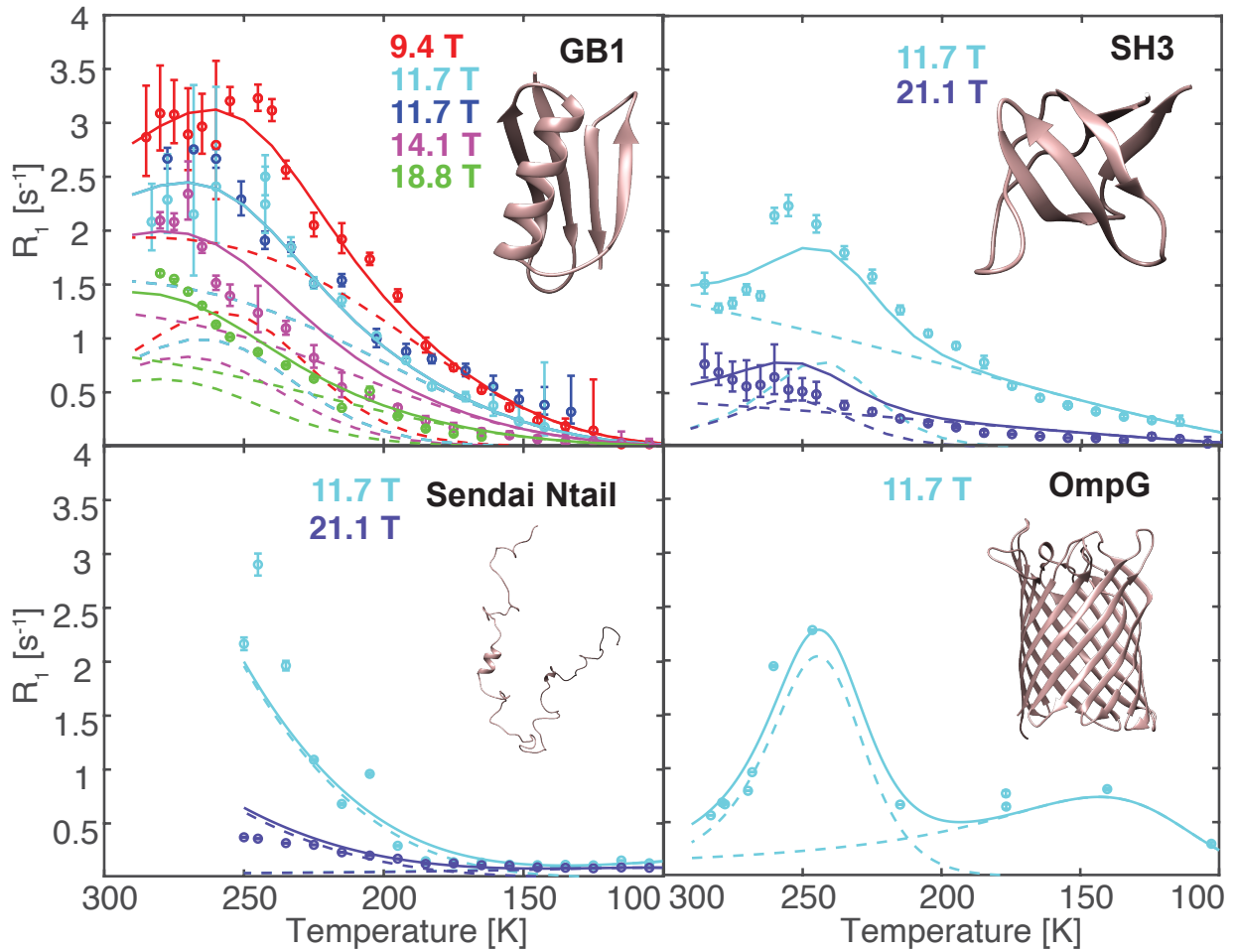


Figure 30: Measured R_1 of the directly observed proton signals (solid circles), R_{1H} , which is mainly associated with the solvent, together with the associated error bar, as a function of temperature. Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The solid line is the result of the joint fit for all the data simultaneously to a model with $N = 2$ modes for GB1, SH3 and OmpG, and to $N = 1$ for Sendai Ntail. The dotted lines show the contribution of each mode to the full relaxation process. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$).

As expected in Figure 30 the relaxation rates roughly increase with temperature, as motions become more efficient to drive relaxation. The relaxations rates occur at several orders of magnitudes of differences, as presented in Figure 31.

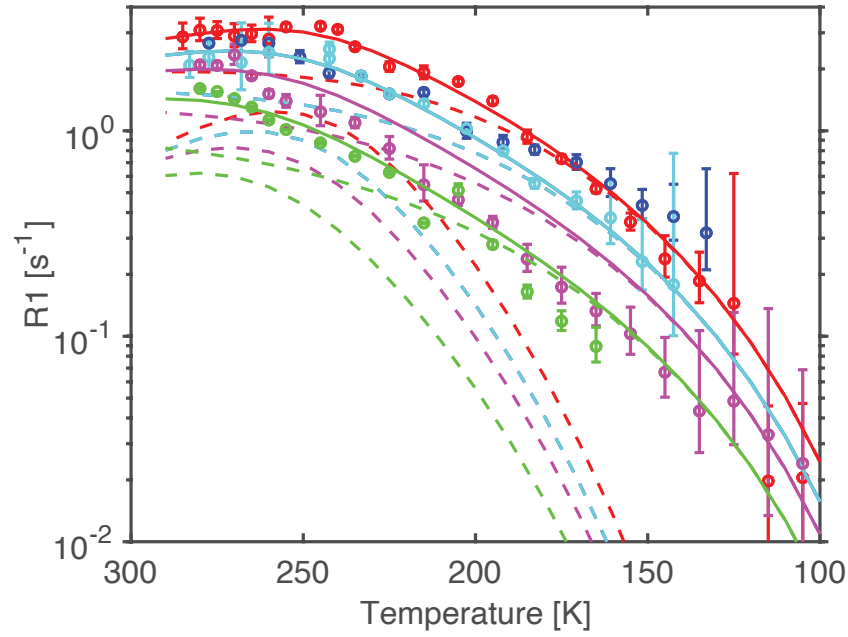


Figure 31 : relaxation rates of GB1 with a logarithmic scale, compared to Figure 30.

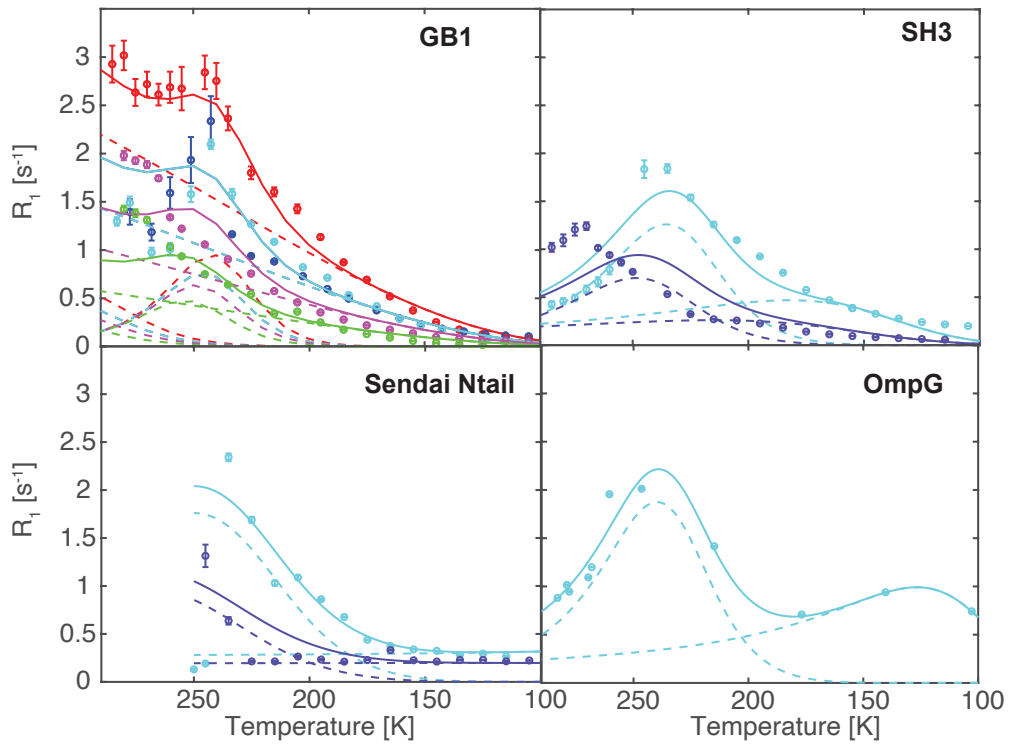


Figure 32 : Measured R_1 of the ^1H signals measured through CP to ^{13}C are represented by solid circles together with the associated estimated errors, as a function of temperature. Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The solid line is the result of the joint fit for all the data simultaneously to a model with $N = 3$ modes for GB1, and to $N = 2$ for, SH3, OmpG and Sendai Ntail. The dotted lines show the contribution of each mode to the full relaxation process. The signals arise mostly from the protons bound to the protein, but since the relaxation times are in the order of magnitude of seconds, proton spin diffusion takes place. This homogenises the relaxation rates with the 'crystal water'. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$).

The relaxation rates presented in Figure 30 and Figure 32 are highly similar due to the solvent being the main contributor to both. For a more precise description, the relaxation rates alone are not enough and we will refer to the Figure 40 and Table 2 for a more detailed analysis.

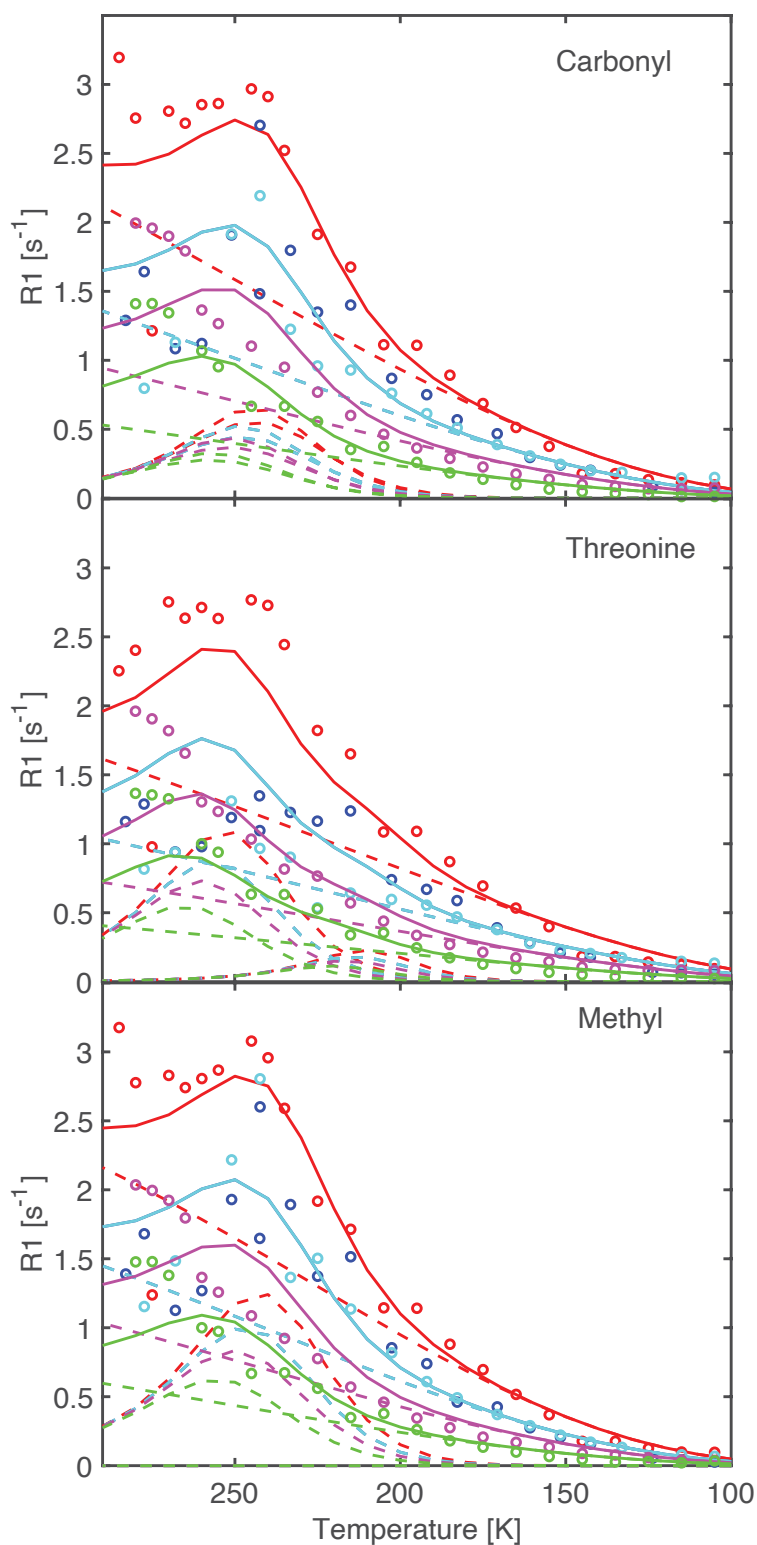


Figure 33 : Multi field fit of the of the ^1H R_1 rates measured through CP to ^{13}C obtained from the integration of different regions of the ^{13}C spectra analysed in Figure 2. The pattern and the fit constants are very similar.

Since the CP mechanism is only efficient at short distances, it might be expected that the ^{13}C spectra detected ^1H relaxation will give different relaxation time for different part of the spectra. But, since our samples are fully protonated, and since the T_1 are long compared to spin diffusion timescales, relaxation rates are efficiently homogenized by spin (Figure 33), and as a consequence we only observe a single relaxation time being the weighted average of the protein protons and “crystal water”. We therefore conducted our experiments by integrating the entirety of the ^{13}C spectra.

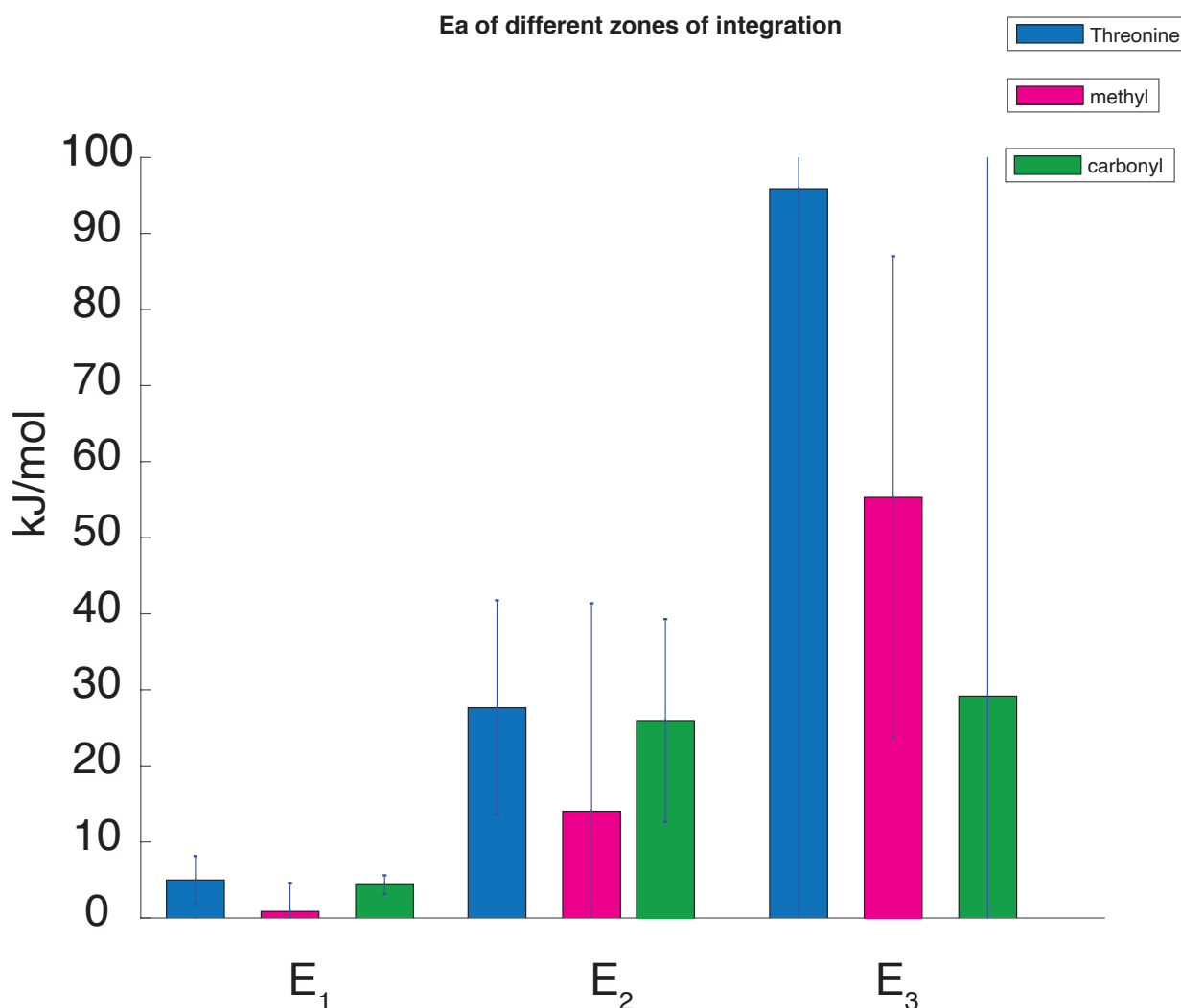


Figure 34 : Energies of activation for motions for the different probes obtained from the data shown on Figure 33. A all the differences are well in the error bars. The full extent of the error is not shown in the mode 3.

We first briefly recall what each of the measured rates reports on. ^1H R_1 are evaluated by integrating the whole ^1H spectrum. As the spectrum is largely dominated by the solvent (and lipid for OmpG) signal, the measured R_1 mainly reports on the solvent dynamics (Figure 30).

The ^1H ^{13}C CP signals arise mostly from the protons bound to the protein, but since the relaxation rates are on the seconds timescale, ^1H - ^1H spin diffusion will average relaxation rates between hydration water, channel water in the microcrystal, and any bulk water by diffusion length scales up to around hundreds of nanometers from the protein surface even at intermediate temperatures. Therefore the ^1H relaxation observed through the ^{13}C spectra probes a weighted average of near and far protons. If this is the case, the measured relaxation rates are seen to be uniform at the protein scale, i.e. across all the ^{13}C chemical shifts (Figure 34).

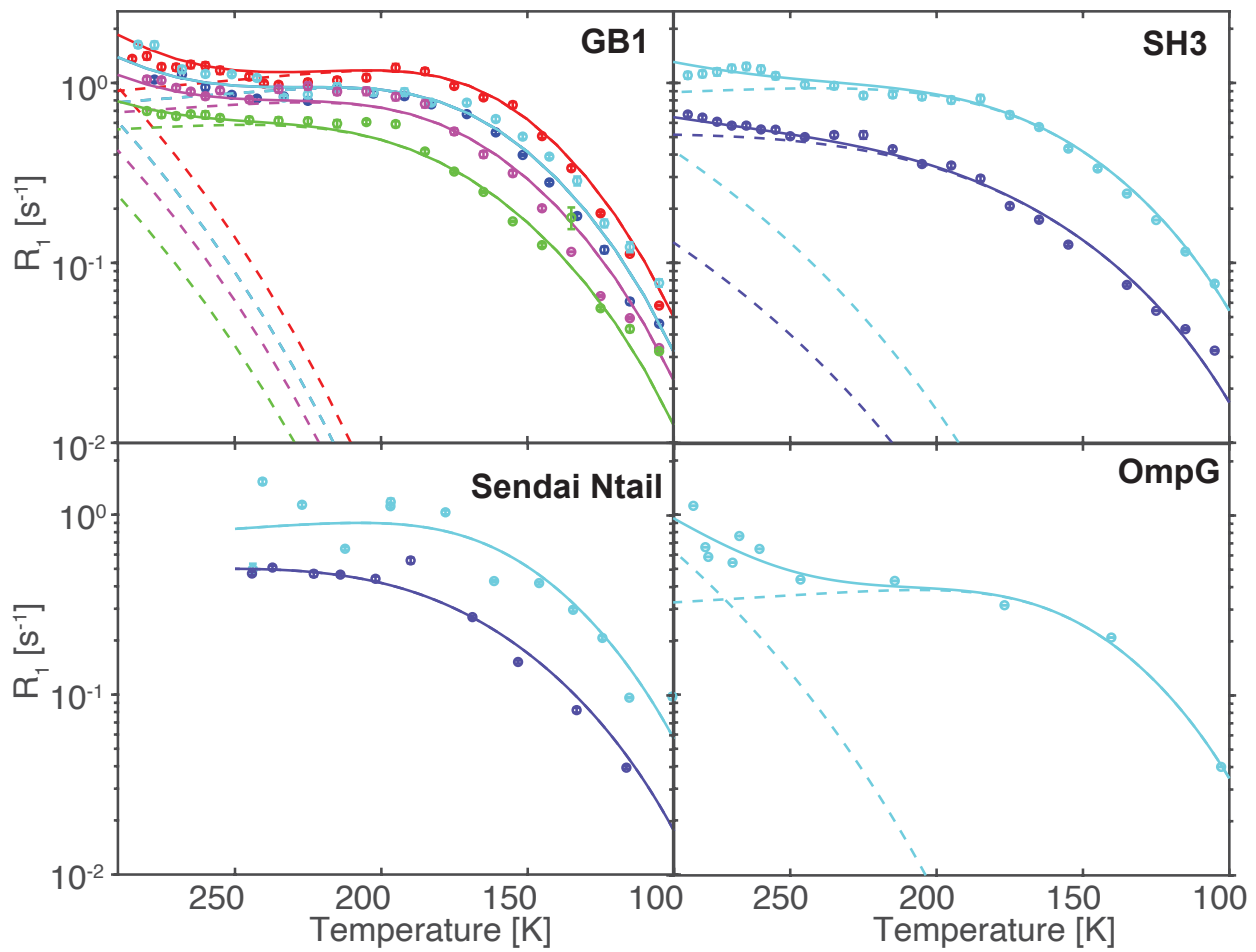


Figure 35 : Measured ^{13}C R_1 of the methyl signals (solid circles), which are mainly associated with the hydrophobic side chains, together with the associated error estimated bars, as a function of temperature. . Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The solid line is the result of the joint fit for all the data simultaneously to a model with $N = 2$ modes for GB1, SH3 and OmpG, and to $N = 1$ for Sendai Ntail. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$).

The methyl signal relaxation represented in Figure 35 has been used in numerous studies⁹⁻¹³, and the reconstructed data at room temperature using this model fit direct observation through other methods.¹

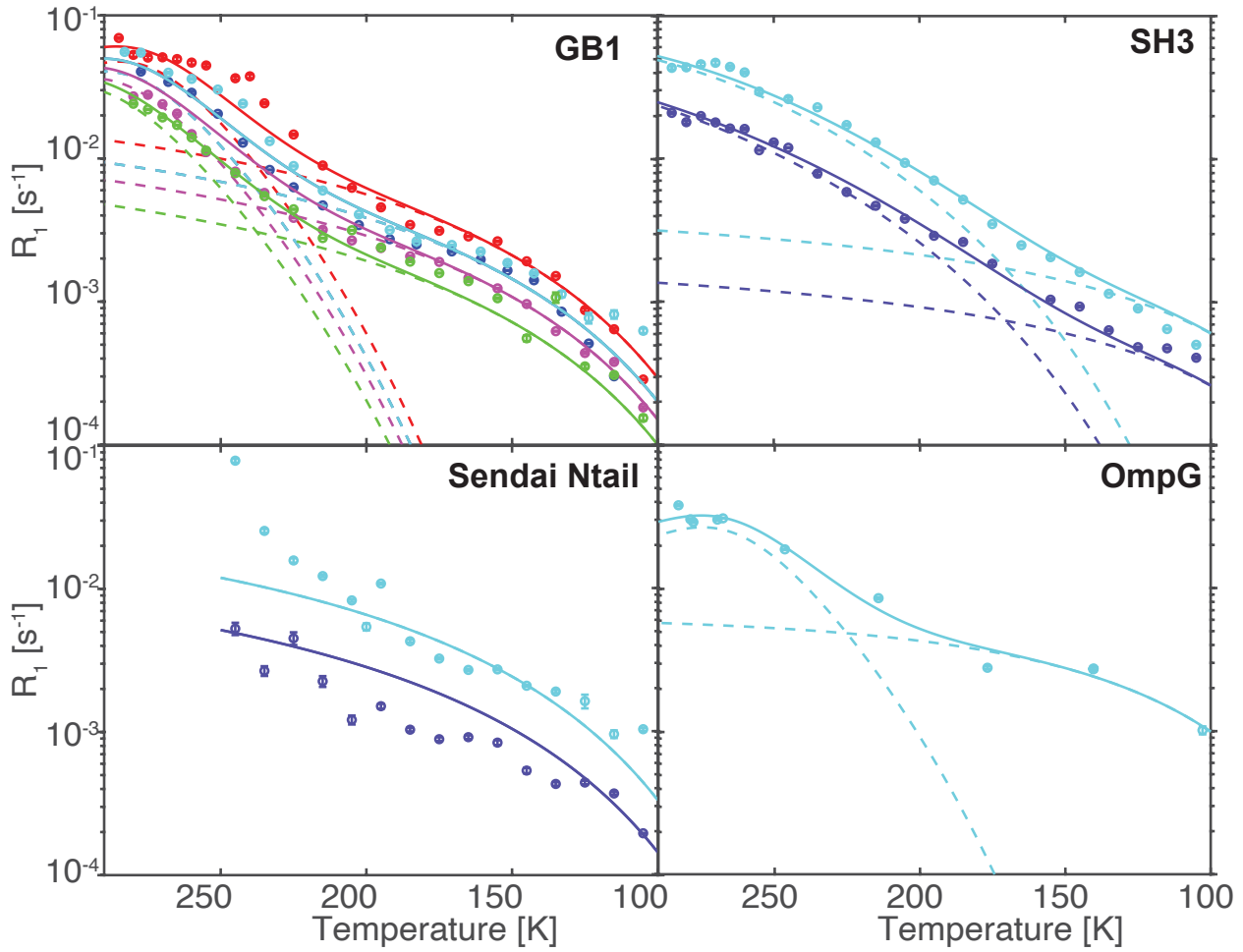


Figure 36 : Measured ^{15}N R_1 of the signals near 120 ppm, represented by solid circles, which are mainly associated with the backbone, and hydration water, together with the associated estimated error bars, as a function of temperature. Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The solid line is the result of the joint fit for all the data simultaneously to a model with $N = 2$ modes for GB1, SH3 and OmpG, and to $N = 1$ for Sendai Ntail. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$).

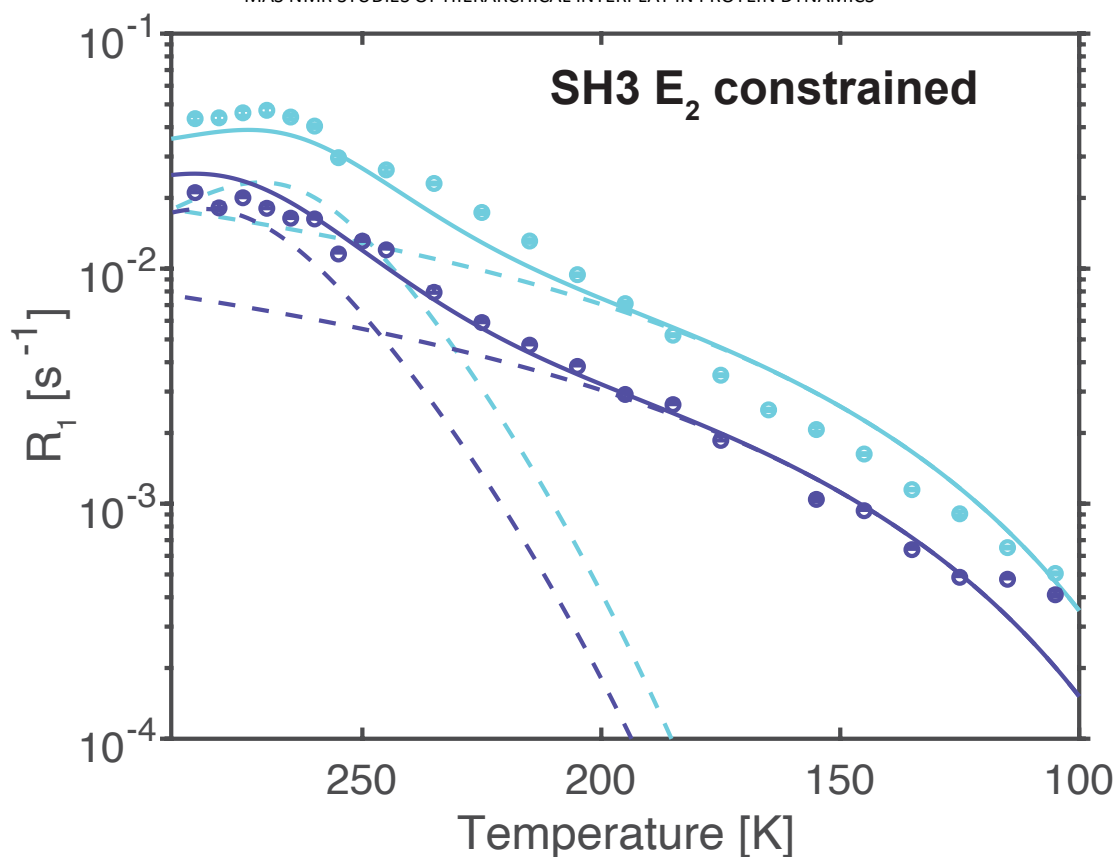


Figure 37 : Measured ^{15}N R_1 of the signals near 120ppm of SH3, previously shown in Figure 4b, represented by solid circles, which are mainly associated with the backbone, and hydration water, together with the associated estimated error bars, as a function of temperature. Each colour corresponds to a different magnetic field, with blue 500MHz and violet 900MHz. The solid line is the result of the joint fit for all the data simultaneously to a model with $N = 2$ modes. The dotted lines show the contribution of each mode to the full relaxation process. The energy of the second mode was arbitrary fixed to $30 \text{ kJ} \cdot \text{mol}^{-1}$ in order to better illustrate the uncertainty on the energy of the second mode.

Methyl ^{13}C R_1 are mostly associated with aliphatic side chains (Figure 3). Amide ^{15}N R_1 , are associated with the protein backbone and hydration water (Figure 32). Lysine ^{15}N R_1 probes hydrophilic side chains (Figure 38). The relaxation pattern is qualitatively similar for all the proteins studied, as their composition is also similar.

At first sight the most obvious difference in the behaviour between samples in the raw relaxation data shown in Figure 30 to Figure 39 for the Sendai Ntail, which, since it is a frozen solution, melts to an isotropic liquid at around 250 K. This means that relaxation above this temperature will now include contributions from overall isotropic molecular tumbling, and the formalism used here is no longer directly applicable. A side effect is also that the isotropic motion will average dipolar couplings to zero, and the cross polarization experiments will no longer work. Since the other proteins are precipitated in lipid bilayers (OmpG) or are microcrystalline (SH3 and GB1) they remain solid even when

the bulk solvent unfreezes. As a result, in those samples overall tumbling is still highly restricted,¹⁴ and cross polarization is still effective.

The characteristic relaxation behaviour that is observed for the other proteins above 250 K is therefore not observable in Sendai Ntail. This has the consequence that for the heteronuclear relaxation rates, only the low energy mode is measurable in Sendai Ntail. However, this does not in itself mean that the underlying dynamics up to the melting point are different.

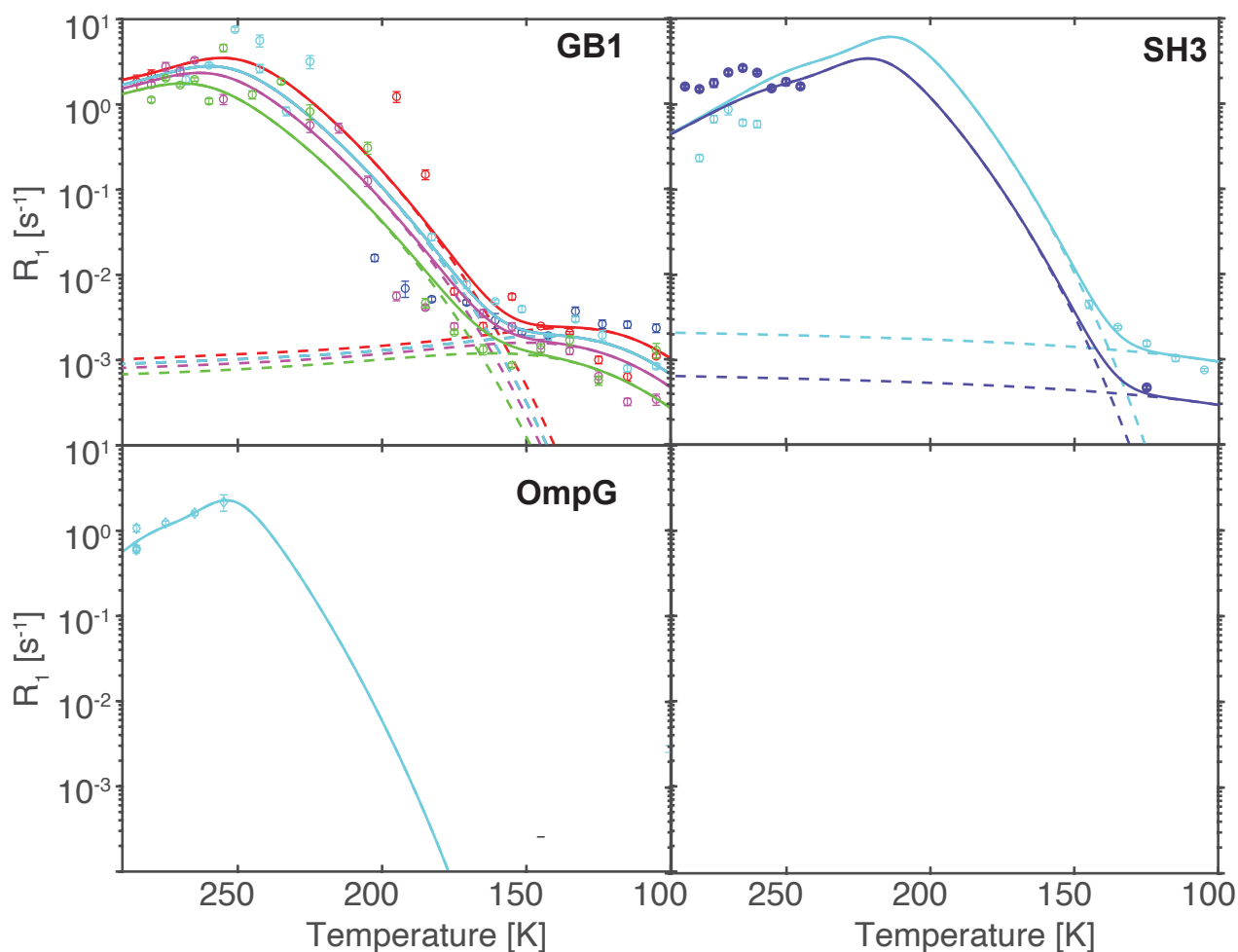


Figure 38: Measured ^{15}N R_1 of the signals near 40 ppm, represented by solid circles, which are associated with the hydrophilic lysine side chains, together with the associated estimated error bar, as a function of temperature. Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The solid line is the result of the joint fit for all the data simultaneously. $N = 2$ modes for GB1, SH3 and OmpG, and to. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$). The signal of Sendai Ntail is too weak to be measured by our methods, due to the low concentration.

The comparison between the Figure 37 and the SH3 panel of Figure 36 highlights the low precision of the measures of the highest energy modes. This is due to not sampling a wide region of the temperature range where this mode drives relaxation.

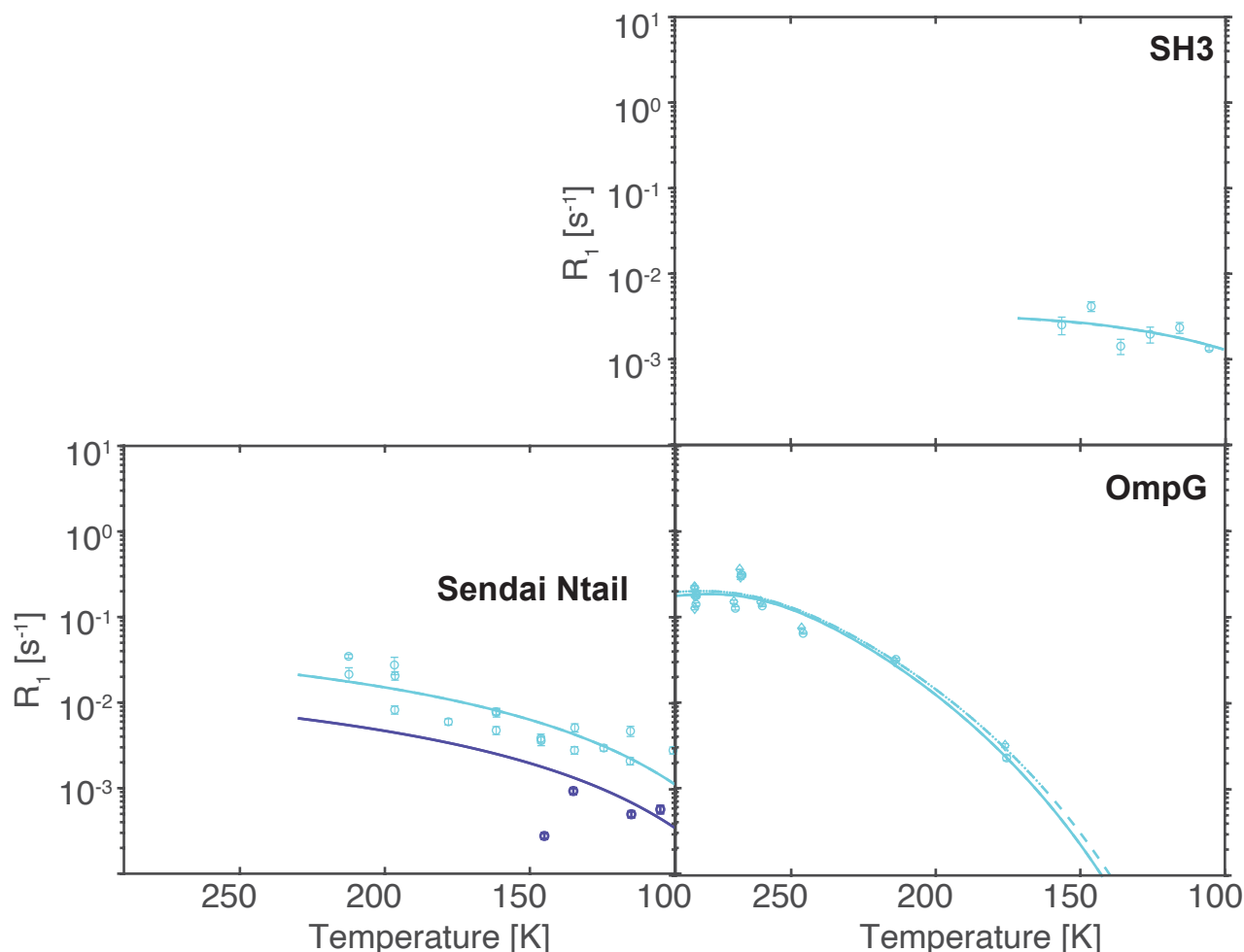


Figure 39: Measured ^{15}N R_1 of the signals near 70 ppm, represented by solid circles, which are associated with the hydrophilic arginine side chains, together with the associated estimated error bar, as a function of temperature. Each colour corresponds to a different magnetic field, with red being 400 MHz, blue 500 MHz, purple 600 MHz, green 800 MHz, and violet 900 MHz. The number of modes needed for the fits was statistically justified with an F-test ($p < 0.05$).

The other obvious difference is in the data for ^{15}N lysine R_1 , shown in Figure 38. Here we observe a loss of signal for GB1 and SH3 between the two measurement points 150 K and 250 K. This has been associated with intermediate exchange motions interfering with decoupling or MAS (motions which are much slower than the motions impacting R_1).^{1,2} In the case of Sendai Ntail it was not possible to properly integrate the signal of interest due to the low signal to noise, and the analysis is thus omitted. The case of OmpG is particularly interesting, since the data obtained here can potentially provide mechanistic insight into the function of positively charged residues in the context of the ‘positive-

inside rule' determining the orientation of helical membrane proteins. Although we are investigating a β -barrel protein, the interaction of the positively charged head groups of arginine and lysine with negatively charged groups of the lipids, mainly phosphates, may lead to an altered relaxation behavior in comparison to the corresponding residues in the three soluble proteins. . OmpG contains six lysine residues, of which five are situated indeed in the loops, turns or just at the ends of the β -strands that may readily interact with lipid molecules, and fifteen arginine residues, with the majority of them located in loops or turns, but also several pointing into the pore. The lysine side chain ^{15}N signals are observable only in the high temperature range (in our measurements above 250 K), and they do not reappear in the low temperature range. This seems already to illustrate a different behavior for these slow motions in the water-soluble proteins as compared to the membrane protein. The data here suggest that the slow lysine side chain motions are significantly modified in OmpG, with an earlier onset in the membrane protein as compared to the water-soluble proteins. Example ^{15}N spectra illustrating this intermediate exchange effect are shown in Figure 41, including spectra with high signal to noise ratios to confirm that the OmpG signals do not reappear at low temperatures. Similarly, the arginine side chain ^{15}N signals are too small at low temperatures to determine R_1 reliably, and a similar, yet slightly shifted curve is obtained as displayed in Figure 39.

If we look in detail at the ^{13}C filtered ^1H relaxation data shown in Figure 32, which is representative of the protons of the protein, the hydration shell, and a contribution from solvent outside the hydration shell,¹⁵ we find that three modes can be distinguished in case of the richer GB1 dataset, but this was not the case for the other proteins that only required two modes to explain the data. However, the two high energy modes found for GB1 are very similar in energy, and we consider them as one mode in the following comparison.

Figure 37 is a fit of the SH3 ^{15}N backbone signal (like Figure 4B) with the high energy mode fixed at 30 $\text{kJ}\cdot\text{mol}^{-1}$. This illustrates the relatively low accuracy of the previous fit, and illustrates that in this case the difference between GB1 and SH3 shown in Figure 36 might not be significant, as forcing this energy to be similar does still give us a reasonable answer. We note that this is not the case for other constrained fits.

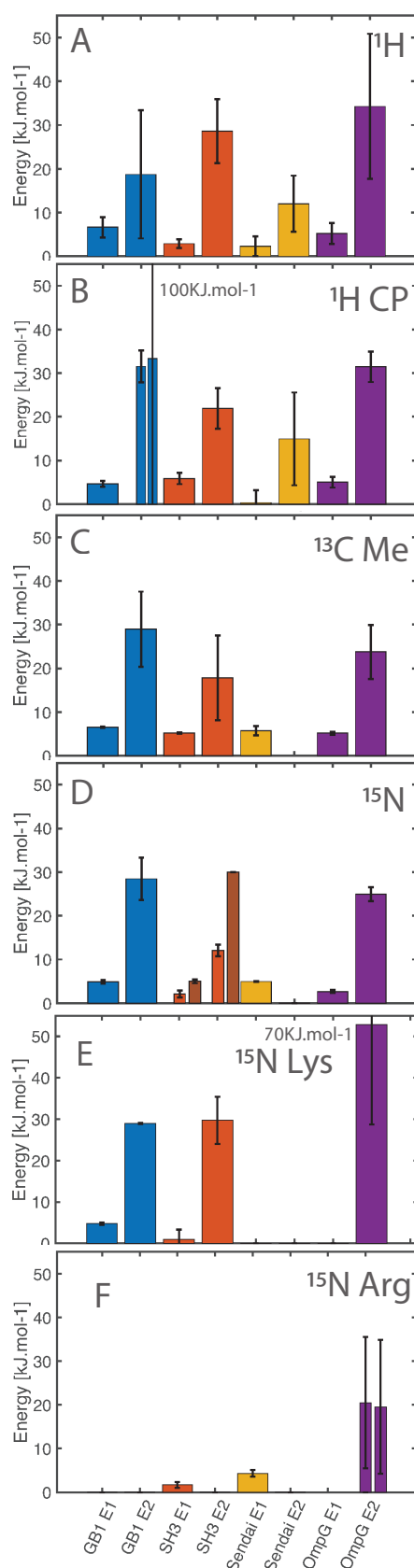


Figure 40: Energies extracted from the fits to the different relaxation rates for the dynamics associated with the 5 probes. Error bars are estimated based on a Monte Carlo analysis as described in the motional mode and activation energies subsections. The double bars in the ¹H CP panel represent the modes 2 and 3. The darker bars in the ¹⁵N panel for the SH3 resonances represent the data for the forced fit shown in figure 6. The two bars in the ¹⁵Nargs panel represent the two separate lysine signals from OmpG.

3.3.2 Activation Energies.

We are now in a position to compare the energetics of the fast motion modes exhibited by the different types of protein, as shown in Figure 40.

As it was discussed more extensively in the previous chapter, it is important to note that we do not obtain a single activation energy for each mode per residue type. Our description is obviously an approximation, because the relaxation rates measured are averaged over many sites and thus due to a distribution of motions over a range of timescales and amplitudes. The activation energies and amplitudes obtained are thus weighted averages. Nonetheless, this implies both that if we can identify two modes with statistical significance, then these modes are measurably different, and that any differences in these averages between proteins corresponds to significant changes in dynamics between proteins.

Overall, we see that qualitatively all the proteins have low and high energy modes in broadly similar ranges, and that this tends to confirm the idea that protein dynamics is to some extent universal. The low energy modes range from 2 to 6 kJ.mol⁻¹ (excluding the Sendai Ntail mode for the protons near the protein), and the high energy modes range from 10 to 35 kJ/mol.

When we take a deeper look at the activation energies, we can identify differences in behaviour that can be rationalised in terms of the atomic-level structures.

The directly measured ¹H rates, which probe activation energies of solvent motion both within the crystal lattices and in bulk excess water or lipid, are reported in the first panel of Figure 40. The low energy mode is mainly due to libration of the solvent since at this temperature water rotation as well as translation is almost entirely blocked.¹⁶ The high energy mode has been assigned to solvent rotations. We observe quite consistent behaviour across the series, with the most striking difference being the relatively low value for the high energy process in Sendai Ntail. We attribute this to the less stringent requirement for solvent ordering around the intrinsically disordered protein (which is only present at very low concentration yet widely distributed), as compared to the in the other three, somewhat compartmentalized samples. We assume that the latter is particularly the case for the samples containing both the crystalline GB1 and SH3 structures due to the presence of spaces in the crystal

lattice, and also for the one with the membrane inserted OmpG where the extended lipid bilayer-water interface is expected to contribute dominantly to this effect rather than the protein. Indeed, the relaxation behavior of the overall ^1H signal of the OmpG sample indicates the highest activation energy of the set for these motions, whereby it should be noted that part of the observed R_1 rate is obtained on the signals of the lipid bilayer.

The measured activation energies obtained from the ^{13}C -filtered ^1H relaxation rates, which are sensitive to dynamics of protons in the solvent close to the protein are shown in Figure 26. The low energy mode has been attributed to fluctuation of side chains within local energy wells, as well as to the above-mentioned restricted solvent motions. The high energy mode is assigned to solvent rotation and larger amplitude side chain motions, with the first being concomitant with the second.¹⁶ The most obvious difference is the energy for the first mode of the very dilute and the most highly water accessible Sendai Ntail being one order of magnitude lower than its counterparts. We can also notice that the more complete data set of GB1 allows us to distinguish three modes here. The energy of the second mode for Sendai Ntail also seems to be lower than its counterparts. We would explain this by the IDP being less structured, and therefore being less constrained, leading to a lower energy of activation. We previously noted² that the ^1H and ^{13}C filtered ^1H signals appear to be reporting on the same processes as the backbone relaxation, and it is not possible to distinguish between the solvent inducing (slaving) protein motions or vice versa. The solvation water moves in concert with the protein. Therefore in this case the ^{13}C filtered ^1H relaxation reports on the slow component of the backbone motion of Sendai Ntail. We note that this is invisible in the ^{15}N relaxation, but measurable here, and is, not surprisingly, lower than for the other proteins. This yields values that are very similar to those measured in solution.¹⁷ This is coherent with previous descriptions and values that have been measured in solution,¹⁸ although they have been made at higher temperature.

Measures of ^{13}C relaxation in methyl groups report on the activation energies of the methyl group motions in the protein and are reported in the third panel of Figure 40. The low energy mode was assigned to methyl rotation, while the higher mode was associated with the jump between different side-chain rotameric states.¹ The first observation is that the low energy mode (i.e. rotation) seems to be very consistent between the four proteins. This might be unexpected, since methyl rotations are known to vary significantly.^{10, 11} However, it is not surprising that there is overall very high similarity here, due to the fact that we measure the average activation energy for all the methyl groups in the protein. The average environment of methyl groups in the proteins is expected to be similar, and therefore the average activation energies should be similar among proteins. It is also known that

methyl group rotation does not depend strongly on solvent, such that the contacts with lipids for OmpG is not expected to significantly affect the dynamics.¹⁹

Measures of ^{15}N amide relaxation rates which are mostly influenced by backbone motions are shown in the corresponding panel in Figure 40. We note that amide bonds also interact with water, depending on the fold. The low energy mode was assigned to small scale and small amplitude motions, while the high energy was assigned to collective small amplitude backbone motions – for example of the peptide plane.¹ The main concern here, as mentioned above, is with the SH3 fit, that is not highly accurate. The forced fit shown in Figure 6 is an indication that the differences between SH3 and GB1 might not be significant, especially if we notice that the energies are similar among the other proteins as shown in Figure 40. If we discount this difference, then these activated processes seem to be very similar between the four proteins.

The arginine group and the lysine NH_3^+ are sensitive to motions related to the hydrophilic side chains enabling interesting comparisons, although the data set is somewhat incomplete. Arginines are absent from GB1, and data for SH3 are missing at high temperature due to low signal-to-noise ratios and chemical exchange, and therefore only the low energy motions are observable. The energies extracted from the arginine ^{15}N R_1 for SH3 and Sendai NTail are shown in Figure 39.

The analysis of the observable lysine ^{15}N R_1 s for GB1, SH3 and OmpG yields the energies shown in Figure 40. The data again indicate at least a bimodal distribution of energies for side-chain motions in GB1 and SH3. It appears that the difference is the most significant in the comparison between OmpG and the other proteins, where we observe a higher activation energy associated with the second mode in OmpG. This might be associated with a higher rigidity of the lysine sidechains in the loops and turns. In OmpG, they are all available for interactions with negatively charged lipids in the membrane, in particular phosphatidylglycerol. The corresponding activation energies associated with the arginine guanidinium group signals cannot be compared to those of SH3 and Sendai NTail, but they are closer to the values observed for backbone amide nitrogen. This implies the mixture of rigidity and flexibility that is represented by the loop/ β -sheet composition of OmpG: the larger part of the arginine residues in OmpG may interact with the membrane, but a large fraction also not. As discussed above, due to the disappearance/low intensity of side chain ^{15}N signals below 250 K, it was not possible to extract the motional parameters associated with the low energy mode of lysine residues in OmpG. This difference also corroborates an overall difference in the lysine side chain dynamics. We conclude therefore that the data reflect interactions of the lysine side chains with the lipids in the bilayer.

The difference of behavior of the lysine sidechains of the membrane protein for both ^{15}N R_1 and the slow exchange behavior, point both to an overall difference in lysine motions between the membrane protein and the microcrystalline globular proteins. This corroborates interactions of the positively charged lysine head group with negatively charged lipids in the bilayer, and provides a possible mechanism that explains the role of positively charged amino acids in membrane protein biology. The comparison of the lysine data obtained on SH3 and GB1 with the activation energy E_2 obtained for OmpG also suggest a stronger interaction with the membrane than with the aqueous environment within the crystallites present in the SH3 and GB1 samples.

In Table 2 to decide if the mode was a the ‘low energy’ or the ‘high energy’ mode, we focus on where it drove relaxation. Motion that were driving relaxation at higher temperature were assigned to high energy, and motions that drive at low temperature where assigned to the low energy.

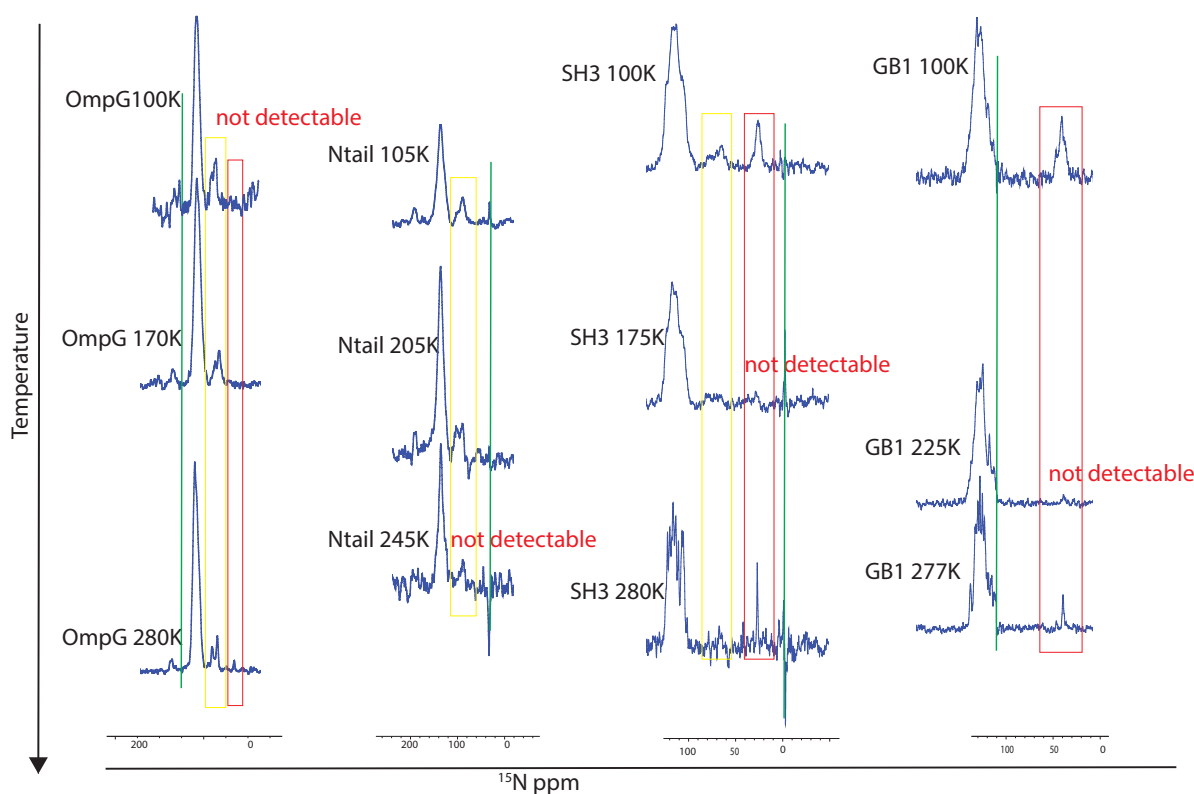


Figure 41 : broadening and disappearance of the lysine signal as a function of temperature for ^{15}N spectra. The scales are not similar between the experiments but they have been made at 500MHz. The red box indicates the lysine signal, the yellow box the arginine signal and the green line the pulse transmitter location.

Table 2 : Numerical values for the fits of the relaxation parameters. The activation energy and the correlation time refer to the Arrhenius fit, and the C represent the $C_{k,amplitude}$ used in fit as described above. Sd stand for one standard deviation as a measure of error.

1H CP												
results	GB1	SH3	Sendai Ntail	OmpG	error	GB1	SH3	Sendai Ntail	OmpG			
ck1	9.00E+09	1.03E+09	7.83E+08	2.92E+09		2.80E+10	2.46E+08	3.87E+11	3.81E+08			
ck2	1.69E+09	2.78E+09	3.88E+09	2.24E+09		4.52E+08	3.83E+08	3.70E+11	6.76E+08			
ck3	1.04E+09					3.70E+09						
t1 (s)	1.78E-10	4.04E-12	7.74E-11	2.87E-11		4.94E-10	5.34E-12	1.50E-09	1.57E-11			
t2 (s)	3.87E-17	2.68E-15	1.49E-13	2.00E-17		1.34E-16	1.50E-14	2.43E-10	1.02E-14			
t3 (s)	4.39E-16					1.55E-15						
e1 (J/mol)	4.63E+03	5.88E+03	3.33E+02	5.02E+03		6.88E+02	1.30E+03	2.81E+03	1.24E+03			
e2 (J/mol)	3.16E+04	2.19E+04	1.49E+04	3.14E+04		3.67E+03	4.66E+03	10620.6203	3.48E+03			
e3 (J/mol)	3.33E+04				7.40E+04							
1H direct												
	GB1	SH3	Sendai Ntail	OmpG	error	GB1	SH3	Sendai Ntail	OmpG			
ck1	3.42E+09	2.29E+10	3.57E+08	1.94E+09		1.25E+09	2.74E+10	9.39E+10	9.83E+08			
ck2	2.20E+09	1.76E+09	3.72E+10	5.39E+09		1.40E+09	1.37E+09	3.06E+13	1.22E+09			
t1 (s)	1.49E-11	1.04E-09	7.57E-12	1.97E-12		1.44E-10	2.85E-24	1.46E-08	1.37E-10			
t2 (s)	3.99E-14	1.42E-16	1.18E-11	7.84E-18		4.58E-16	2.85E-09	7.68E-09	3.16E-16			
e1 (J/mol)	6.65E+03	2911.12088	2.32E+03	5.26E+03		2.32E+03	1.02E+03	2.26E+03	2.40E+03			
e2 (J/mol)	1.87E+04	28603.5815	1.20E+04	3.43E+04		1.46E+04	7.28E+03	6.42E+03	1.66E+04			
13C Me												
	GB1	SH3	Sendai Ntail	OmpG	error	GB1	SH3	Sendai Ntail	OmpG			
ck1	7.04E-02	3.40E-01	3.27E-01	2.90E-02		2.87E-03	3.47E-02	5.83E-01	2.58E-03			
ck2	7.09E+00	3.24E+01		1.89E-01		1.84E+02	2.20E+03		3.11E+00			
t1 (s)	2.56E-11	7.31E-11	3.55E-11	4.78E-11		4.36E-12	1.30E-11	2.81E-09	2.04E-11			
t2 (s)	2.11E-12	2.99E-10		4.59E-13		6.06E-11	2.28E-08		1.39E-11			
e1 (J/mol)	6.54E+03	5.22E+03	5.76E+03	5.21E+03		1.69E+02	1.41E+02	1.05E+03	3.44E+02			
e2 (J/mol)	2.90E+04	1.78E+04		2.38E+04		8.57E+03	9.70E+03		6.17E+03			
15N backbone												
	GB1	SH3	Sendai Ntail	OmpG	error	GB1	SH3	Sendai Ntail	OmpG			
ck1	8.81E-03	7.91E+01	6.36E-01	2.00E-03		2.50E-01	1.75E+03	8.92E-03	6.16E-03			
ck2	1.29E-02	2.25E-02		8.48E-03		9.55E-02	5.37E-01		4.59E-08			
t1 (s)	2.27E-09	2.00E-04	9.42E-08	1.57E-09		7.18E-08	5.00E-03	1.48E-07	6.72E-09			
t2 (s)	2.24E-14	5.00E-11		5.40E-14		6.31E-13	1.48E-09		1.34E-13			
e1 (J/mol)	4.92E+03	2.10E+03	4.95E+03	2.66E+03		3.98E+02	7.90E+02	1.27E+02	3.61E+02			
e2 (J/mol)	2.85E+04	1.21E+04		2.50E+04		4.82E+03	1.29E+03		1.62E+03			
15N Lys												
	GB1	SH3	Sendai Ntail	OmpG	error	GB1	SH3	Sendai Ntail	OmpG			
ck1	7.37E-04	1.52E-02				0.00020888	1.61E+00					
ck2	1.08E+00	2.34E+00		7.96E-01		0.121699	6.66E+00		2.26E+01			
t1 (s)	5.42E-11	7.78E-08				6.6176E-11	5.37E-06					
t2 (s)	4.50E-15	1.63E-16		4.06E-20		2.6234E-15	1.12E-14		4.26E-16			
e1 (J/mol)	4.76E+03	9.95E+02				297.6975	2.34E+03					
e2 (J/mol)	2.89E+04	2.97E+04		5.29E+04		145.1861	5.73E+03		2.41E+04			
15N Arg												
	GB1	SH3	Sendai Ntail	OmpG 1	OmpG 2	error	GB1	SH3	Sendai Ntail	OmpG 1	OmpG 2	
ck1		3.90E-02	1.49E-01					2.72E-02	1.45E+00			
ck2				7.43E-02	0.07788171						4.39E-01	0.71028054
t1 (s)		6.53E-08	1.16E-08					4.24E-08	9.34E-08			
t2 (s)				4.25E-13	6.84E-13						4.89E-10	1.18E-09
e1 (J/mol)		1.72E+03	4.34E+03					6.64E+02	8.02E+02			
e2 (J/mol)				2.05E+04	19551.2057						7.93E+03	6319.67228

3.4 Conclusions

In conclusion, we have shown that by measuring a series of 6 relaxation rates at variable temperatures and variable magnetic fields, we can obtain averaged activation energies for motions in different parts of a series of four proteins representing three of the broad classes of proteins: soluble globular proteins, intrinsically disordered proteins and membrane proteins.

We observe many similarities, notably in backbone motions and methyl group rotations, suggesting that proteins do share universal dynamic features. For these motions, similar activation energies are measured for all four proteins.

Besides this, some differences are evident, with the main observed difference between the various types of protein occurring at the interface between solvent and protein. This is unsurprising since the influence of the solvent on the dynamics of proteins has been well documented²⁰⁻²⁵.

One difference is related to the reduced activation energy of the Sendai Ntail (reported on by ^1H and ^{13}C mediated ^1H relaxation). This suggests that the well defined energy minimum associated with a folded protein impacts local dynamic modes by constraining the degrees of conformational freedom, while this does not happen in IDPs because the energy landscape is flatter. This lack of restraint imposed by the rigid scaffold appears to lead to lower activation energies. In the case of the membrane protein, the activation energy is similar to the water-soluble proteins, suggesting that the fold is more important than the nature of the solvent.

Another difference is observed in the lysine sidechains of the membrane protein for both ^{15}N R_1 and the slow exchange behavior, pointing to an overall difference in lysine motions between the membrane protein and the microcrystalline globular proteins. This is consistent with interactions of the positively charged lysine head group with negatively charged lipids in the bilayer, and suggests a possible mechanism for the role of positively charged amino acids in membrane protein biology. The comparison of the lysine data obtained on SH3 and GB1 with the activation energy E_2 obtained for OmpG also suggest a stronger interaction with the membrane than with the aqueous environment within the crystallites present in the SH3 and GB1 samples.

Future work will be directed to extending this analysis to more proteins, in order to validate the preliminary observations here, and further refine the understanding of the similarities and differences between dynamics in proteins.

3.5 References

1. Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L., Direct observation of hierarchical protein dynamics. *Science* **2015**, *348* (6234), 578-581.
2. Busi, B.; Yarava, J. R.; Hofstetter, A.; Salvi, N.; Cala-De Paepe, D.; Lewandowski, J. R.; Blackledge, M.; Emsley, L., Probing Protein Dynamics Using Multifield Variable Temperature NMR Relaxation and Molecular Dynamics Simulation. *J Phys Chem B* **2018**, *122* (42), 9697-9702.
3. Frauenfelder, H.; Chen, G.; Berendzen, J.; Fenimore, P. W.; Jansson, H.; McMahon, B. H.; Stroer, I. R.; Swenson, J.; Young, R. D., A unified model of protein dynamics. *Proceedings of the National Academy of Sciences* **2009**, *106* (13), 5129-5134.
4. Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H., Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **2002**, *420* (6911), 98-102.
5. Linden, A. H.; Franks, W. T.; Akbey, U.; Lange, S.; van Rossum, B. J.; Oschkinat, H., Cryogenic temperature effects and resolution upon slow cooling of protein preparations in solid state NMR. *Journal of Biomolecular Nmr* **2011**, *51* (3), 283-292.
6. Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L., Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J Virol* **2007**, *81* (13), 6807-16.
7. Hiller, M.; Krabben, L.; Vinothkumar, K. R.; Castellani, F.; van Rossum, B. J.; Kuhlbrandt, W.; Oschkinat, H., Solid-state magic-angle spinning NMR of outer-membrane protein G from Escherichia coli. *Chembiochem* **2005**, *6* (9), 1679-84.
8. J. Cavanagh, W. J. F., A. G. I. Palmer, N. J. Skelton, M. Rance, *Protein NMR Spectroscopy*. 2006.
9. Beckmann, P. A.; Buser, C. A.; Mallory, C. W.; Mallory, F. B.; Mosher, J., Methyl reorientation in solid 3-ethylchrysene and 3-isopropylchrysene. *Solid State Nucl Mag* **1998**, *12* (4), 251-256.
10. Carbonell, P.; del Sol, A., Methyl side-chain dynamics prediction based on protein structure. *Bioinformatics* **2009**, *25* (19), 2552-8.
11. Chatfield, D. C.; Augsten, A.; D'Cunha, C., Correlation times and adiabatic barriers for methyl rotation in SNase. *J Biomol NMR* **2004**, *29* (3), 377-85.
12. Copie, V.; McDermott, A. E.; Beshah, K.; Williams, J. C.; Spijker-Assink, M.; Gebhard, R.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G., Deuterium solid-state nuclear magnetic resonance studies of methyl group dynamics in bacteriorhodopsin and retinal model compounds: evidence for a 6-s-trans chromophore in the protein. *Biochemistry-U S* **1994**, *33* (11), 3280-6.
13. Olejniczak, E. T.; Weiss, M. A., Are Methyl-Groups Relaxation Sinks in Small Proteins. *J Magn Reson* **1990**, *86* (1), 148-155.
14. Lamley, J. M.; Lewandowski, J. R., Relaxation-Based Magic-Angle Spinning NMR Approaches for Studying Protein Dynamics. *Emagres* **2016**, *5* (3), 1423-1433.
15. Bockmann, A.; Gardiennet, C.; Verel, R.; Hunkeler, A.; Loquet, A.; Pintacuda, G.; Emsley, L.; Meier, B. H.; Lesage, A., Characterization of different water pools in solid-state NMR protein samples. *Journal of Biomolecular Nmr* **2009**, *45* (3), 319-327.
16. Lee, A. L.; Wand, A. J., Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature* **2001**, *411* (6836), 501-504.
17. Taricska, N.; Bokor, M.; Menyhard, D. K.; Tompa, K.; Perczel, A., Hydration shell differentiates folded and disordered states of a Trp-cage miniprotein, allowing characterization of structural heterogeneity by wide-line NMR measurements. *Sci Rep* **2019**, *9* (1), 2947.
18. Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R. W.; Jensen, M. R.; Blackledge, M., Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *Journal of the American Chemical Society* **2016**, *138* (19), 6240-51.
19. A ¹H nuclear magnetic relaxation study of the mechanism and solvent dependence of the motion of predominantly syndiotactic poly(methyl methacrylate) in dilute solution. *Polymer* **1980**, *21* (4), 381 - 387.
20. Siemer, A. B.; Huang, K. Y.; McDermott, A. E., Protein Linewidth and Solvent Dynamics in Frozen Solution NMR. *Plos One* **2012**, *7* (10).
21. Lipps, F.; Levy, S.; Markelz, A. G., Hydration and temperature interdependence of protein picosecond dynamics. *Phys Chem Chem Phys* **2012**, *14* (18), 6375-6381.
22. Jansson, H.; Bergman, R.; Swenson, J., Role of Solvent for the Dynamics and the Glass Transition of Proteins. *The Journal of Physical Chemistry B* **2011**, *115* (14), 4099-4109.
23. Fenwick, R. B.; Oyen, D.; Dyson, H. J.; Wright, P. E., Slow Dynamics of Tryptophan-Water Networks in Proteins. *Journal of the American Chemical Society* **2018**, *140* (2), 675-682.
24. Draganski, A. R.; Friedman, J. M.; Ludescher, R. D., Solvent-Slaved Dynamic Processes Observed by Tryptophan Phosphorescence of Human Serum Albumin. *Biophys J* **2017**, *112* (5), 881-891.

25. Adamski, W.; Salvi, N.; Maurin, D.; Magnat, J.; Milles, S.; Jensen, M. R.; Abyzov, A.; Moreau, C. J.; Blackledge, M., A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *J Am Chem Soc* **2019**, *141* (44), 17817-17829.

Chapter 4 NMR analysis of lipid content and division of labour in ants

The following is a description of work carried out in parallel to the main thesis work described in the previous chapters.

Ants are one of the major pests in agriculture, responsible alone for several billion dollars worth of damages every year. This is due to multiple factors like direct predation to different parts of many crops (roots, flowers...) and indirect damages through mutualism with sap sucking insects like aphids, which are notoriously hard to eradicate due to the protection that ants provide to them.¹ Moreover, the direct risks to human health should not be ignored.² From infection to hypersensitivity reaction, with some life-threatening cases, several species of ants cause significant damages to human populations, even in high density populated areas. They are also directly responsible for damaging ecosystems when present as invasive species, for example by predation on amphibians, other arthropods or ground nesting birds.³

These reasons are the driving force for research into the behavior and organization of ants. These (mainly) eusocial insects have conquered all available land surfaces, except the Antarctic and a few reclusive islands. They are thought to represent close to 10% of the animal biomass present on land, which places them as the second most ecologically dominant species after humans.⁴

Ants have a well defined social organization with individuals distributed amongst various tasks in the colony. This is thought to enhance efficiency of labor and permits large, populous colonies with high production. At its simplest tasks are split into those considered as ‘nursing’ (involving caring for developing brood, the queens as well as various nest maintenance behaviors) and ‘foraging’ (involving leaving the nest to go and find food, as well as other diverse roles such as scouting and colony defense). The physiology of ants is quite well documented from destructive analyses, and a correlation between lipid content and behavior is apparent, most notably with nurses having high fat and foragers low fat content. However, the dynamic interplay between physiology and behavior remain unclear, as techniques have not been available to non-destructively measure the physiology of individuals over time and correlate this with their behavior.⁴ Here we develop a methodology to analyze this problem by combining MAS-NMR and a tracking system. Initial results are presented and

ultimately the method will allow the physiology of individual ants be manipulated to find a dynamic, causative relationship between physiology and behavior.

4.1 Ant biology

Ants belongs to the family Formicidae, which split from the wasps and bees in the early Cretaceous at the latest. They are highly diverse in morphology and behavior, but most have some distinctive features. Note that due to the high diversity among ants, almost all generalization about their biology and behavior will find some exceptions, and the following introduction serves only as a general overview.

First, the most striking characteristic of ants is their eusocial behavior, displayed by the vast majority of ant species. This means that a typical ant population consists of mainly sterile female workers, with one or more fertilized queens that will lay eggs. When males (also called drones) are produced their only function is reproduction, and mating with virgin queens (gynes) after which they die. While there is much to analyze in terms of evolutionary games theory regarding the eusocial behavior, this subject will not be treated here.

We will first go through the life cycle of an ant to understand their specific biology. When an egg is laid, it will develop as female if it is diploid (two set of chromosomes like most of human cells) and as male if it is haploid (one set of chromosomes, like human spermatozooids and ovules). Therefore, only mated queens can produce both workers and alates (the cast of eusocial insects dedicated to reproduction). Eggs, like larvae and pupae, require precise conditions and constant care to develop. These are cared for by other workers in mature colonies, and the queen(s) will only play a role in their development in the earliest stages of the colony, before the nanitics (first generations of workers) are developed.

The eggs, larvae and pupae are fed, cleaned, and maintained at the right hygrometry, temperature and obscurity by the workers. Depending on a mix of genetic, environmental and nutritional/care factors, pupae will develop in a mature ant. Many ant species have different specific castes of workers (better described by a discrete distribution than a continuous one in the vast majority of cases). Since the model organism that we study here (*Myrmica rubra*) does not exhibit different castes (with the exception of nanitics) we will not go into further detail.

If the new ant is an alate, it will generally fly out of the nest during a nuptial flight to find one single mate. Drones will die quickly after reproduction, and also have a very short lifespan if prevented from mating. The newly mated queen ants will then amputate their own wings and find an appropriate place to start a colony on her own. Most workers can reproduce in certain specific conditions, but since they lack the ability to mate they will only be able to produce male alates. This explains why most ants colonies will die if their queen is removed: they cannot make new workers.

If the new ant is a worker it will start its life by caring for the queen and the brood. It will then move to nest organization and maintenance, and finally foraging. The risks of death and injuries for ants that stay inside the nest are very low, whereas foragers experience a much higher death rate. The change of job is rapid (Richardson et al - in preparation), and that there is a discrete distribution of jobs among ants. For *Myrmica rubra* only two jobs will be considered, nurses and foragers, which is the widely used method to analyze this species in an artificial setup. This distribution of labor is made possible by social stomachs, which allow ants to regurgitate food or water to each other. By this mechanism (called trophalaxis) the entire colony can be fed with only a portion of the individuals actually risking their lives by leaving the nest.⁴

As famously said by Theodosius Dobzansky, “*nothing in biology makes sense except in the light of evolution*”. Therefore, this organization where workers care for their sisters and not for their own progeniture can be much better explained by reference to the field of general evolution.

The simplified explanation is as follows: Female ants are diploid, males are haploid. When a virgin queen gyne mates with a male she combines reproductive material from her genome and his. As such any egg produced by the now mated queen will necessarily contain half her genetic material and *all* the male material. Therefore any of the queen’s daughters will be 75% related to another daughter (100% of the father’s genes and 50% of the mothers). Consequently, sisters are highly related and it is more beneficial for them to encourage the queen to make them more sisters, rather than lay their own eggs to which they would only share 50% of their genetic lineage. This is true also for the queen, all her daughters are only 50% related to her, even if the daughters are 75% related to each other. This is the evolutionary theory explanation as to why ants are eusocial and why workers forego reproduction. A more extensive and rigorous description can be found here.⁴

4.2 Ant society

The division of labor presented above is believed to be the main reason for the ecological success of ants.⁴

Division of labor is self-organized, with no knowledge about the global system. Ants behavior can be explained by reaction to local stimuli, and no global goal nor any centralization of information is required.⁵ As mentioned above, jobs are highly correlated with aging, with young ants being nurses and older ones being foragers.⁶ Nonetheless some exceptions to this general trend are observed. Precocious foraging, as well as behavioral reversion have been observed among eusocial species.^{7,8} This indicates a much more subtle progression than previously thought.

Several mechanisms driving the temporal polyethism (age correlated task allocation) have been scrutinized⁴. One relates to is the amount of fat reserve, i.e. the corpulence.⁹ It is important to notice that is not believed to be the only driving force in behavior, but only one in many (like nest architecture or colony size^{4,10}). The division of labour has previously been correlated to a division in fat reserves, with lean foragers and fat nurses.^{4,9,11} This can be also predicted from an evolutionary point of view. Since ants are able to perform trophallaxis, ants that die with reserves (like fat) are a loss for the colony. Therefore, if only the lean individuals are subjected to a high risk of death, the colony can decrease its potential loss of resources, and therefore increase its reproductive success.⁴ The negative correlation between corpulence and foraging behavior in eusocial insects has also been highlighted in honeybees, ants and wasp, i.e. the majority of eusocial species. This common trait is therefore likely to be important for their fitness as explained above.¹²

Moreover, it is likely that lean foragers will not easily regain fat (if at all). Successful foraging does not lead to consumption by the foraging ant but rather the distribution of resources to the whole colony. Furthermore, foraging is energetically costly, directly by locomotion, and indirectly by the necessity of cuticular hydrocarbon synthesis and immunity response.^{13,14}

The quantification of lipids in ants is generally done using a destructive method.^{15,16} Therefore, it is difficult to follow both the 'job' and the corpulence of an ant over time. We suggest to use NMR to measure both the corpulence of ants and their behavior without killing them and therefore analysing

the trends and correlations of those two indicators over time. Moreover, the key role of fat in insect metabolism makes it a probable intermediate in the phenomenon.¹⁶

We are asking how corpulence regulates temporal polyethism in ants. In order to analyse this problem we will monitor simultaneously the corpulence as well as the foraging behavior of all workers from full colonies of *Myrmica rubra*.

4.3 Non-lethal corpulence and social network measurement

4.3.1 Lipid reserve measurements.

In order to measure the lipid reserves of ants as described in the literature¹⁷, liquid state NMR experiments have been tried, by placing a cold anesthetised ant near to a capillary with solvent for the lock. Unfortunately, the experiments were non-reproducible, very likely due to field inhomogeneity induced by the non-uniform bulk magnetic susceptibility (BMS) of the ant. Since even at low spinning speeds MAS NMR can average this inhomogeneity, we moved to that method.

In order to test our method, we measured a few spectra of ^1H at 21 T (900 MHz ^1H frequency) using a 3.2 mm rotor spinning between 1 and 1.2 kHz. The results are presented in Figure 42. We can clearly separate the fat signal (near 1 ppm) from the water (4.3 ppm). And at this field and this spinning speed the impact of the side bands is negligible to the main fat signal. It is also important to note that the balance of the rotor is far from ideal due to the asymmetric moment of inertia of the ant. This causes spinning to be unstable and may distort the signal. To minimize this and the effect of the sidebands, we measured each ant at two different spinning speed, allowing the two measurements to be compared and errors detected easily.

To minimize the error due to the different position of the ants, they were put in a cold induced anesthetized state. They were then introduced with the gaster (abdomen; lower portion of the body, where the stinger is located) in the rotor, and moved to the bottom of the rotor by gently tapping it against the table. Neither the glue used to stick a QR code to the back of the ants, nor the QR code itself nor the paint used to mark them was detectable using our experiment, and they will therefore not influence our measurement of lipids. The QR code function is explained in the tracking of ants section.

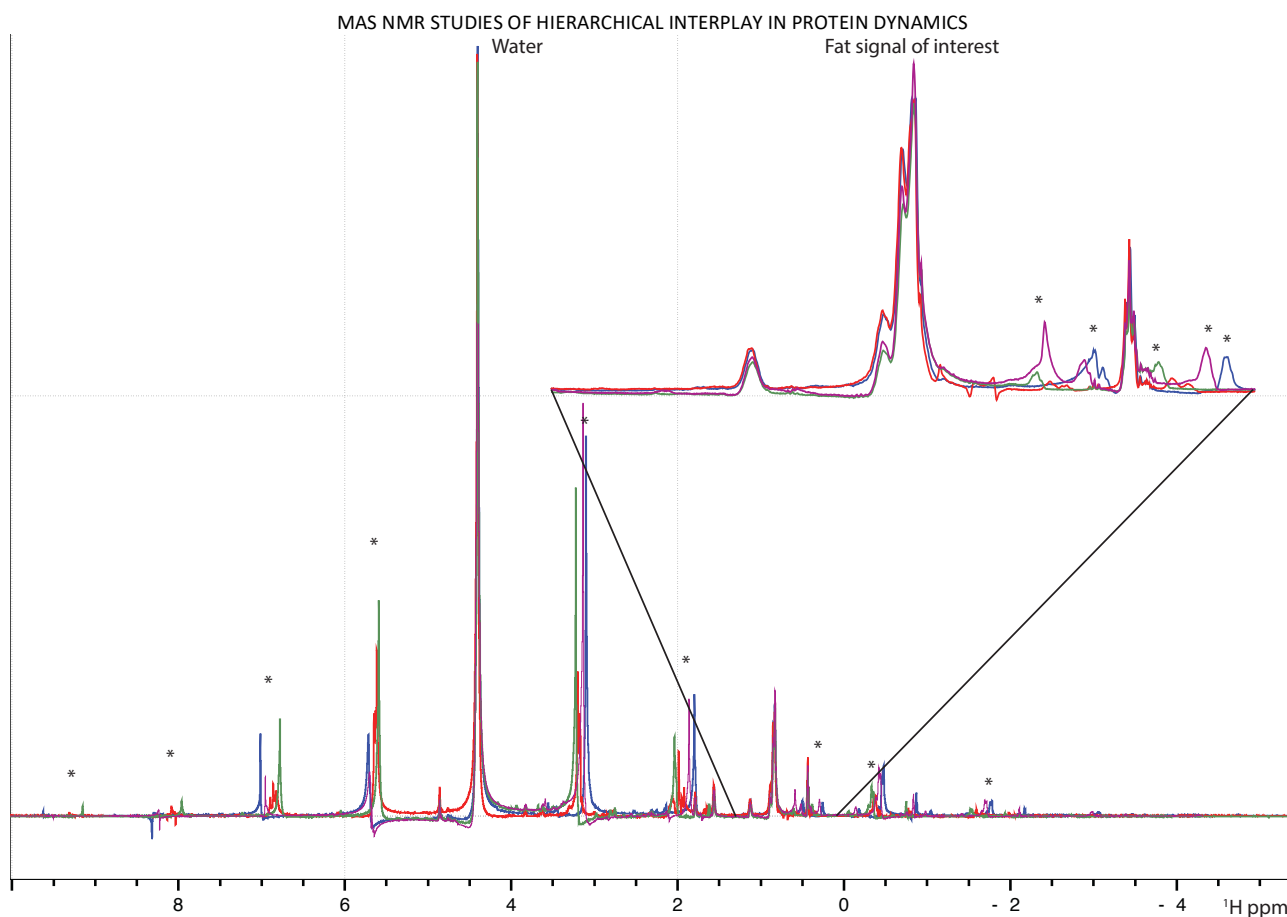


Figure 42: ^1H NMR spectra of the same *Myrmica rubra* ant. The asterisks indicate spinning sidebands, in black for the water and in red for the lipids. We observe that at this spinning speed, water sidebands can overlap with the lipid signal.

To assure the reproducibility of the method, we measured the same ant with different measures, letting her wake up and changing rotor between measurements, in order to identify any variation. In Figure 43 we present the spectra for a single ant across multiple sessions. The same peaks are observed with minimal variation in intensity, indicating that the method is indeed reproducible.

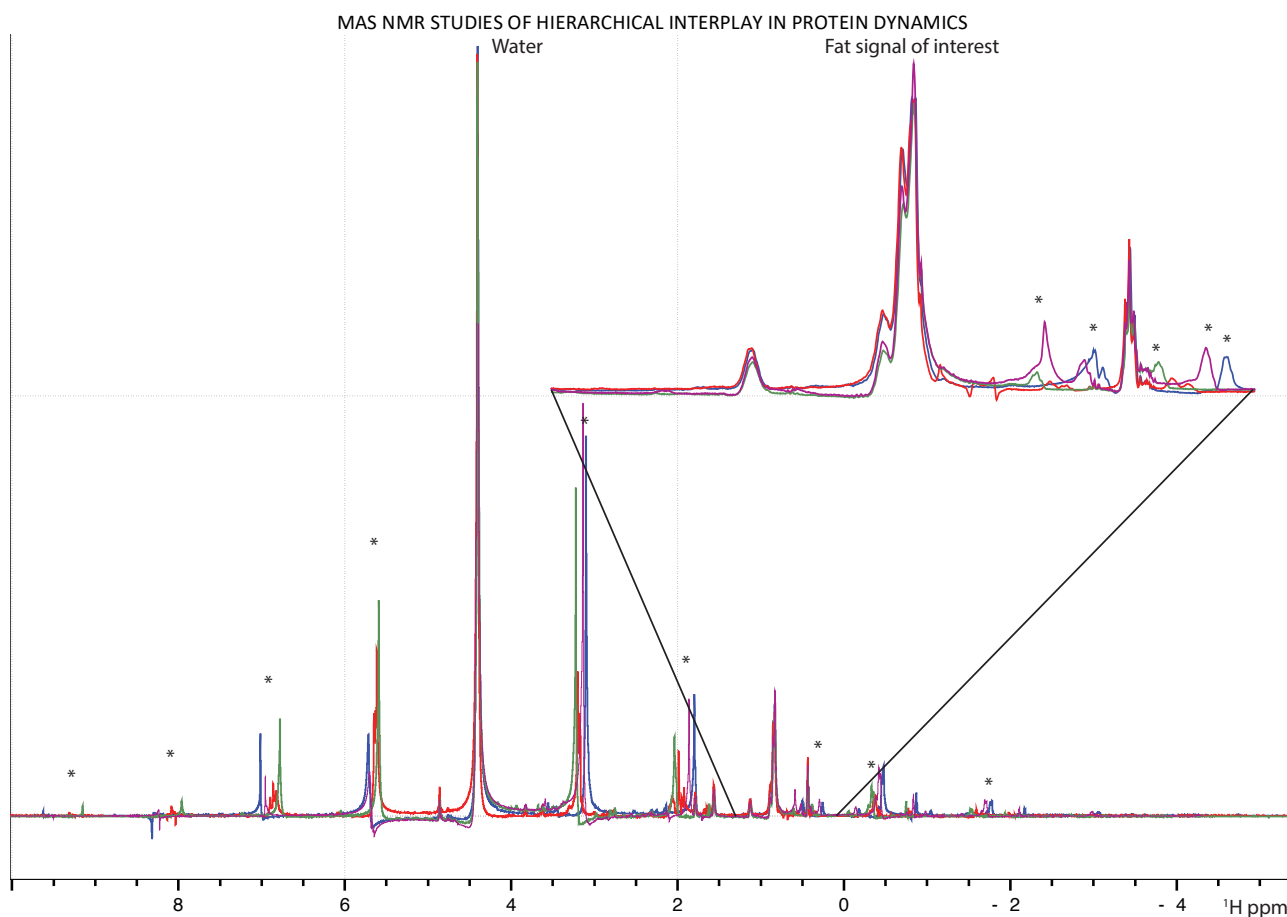


Figure 43 : comparison of the ^1H signal of the same ant in in different rotors.

To quantitatively compare the spectra, we measured the integral of the biggest peak of the lipid region only, as it is difficult to control the spinning speed precisely at such low speed, and the smaller peaks are much more sensitive to superposition with side bands (Figure 44). The main contribution to this peak comes from the CH_2 groups from the lipid chains. We observe that the reproducibility is high, and that there is influence of the packing. One of the rotors was extremely unstable, nevertheless we kept these data to assure a maximum estimate of our error. The results are represented in Figure 44.

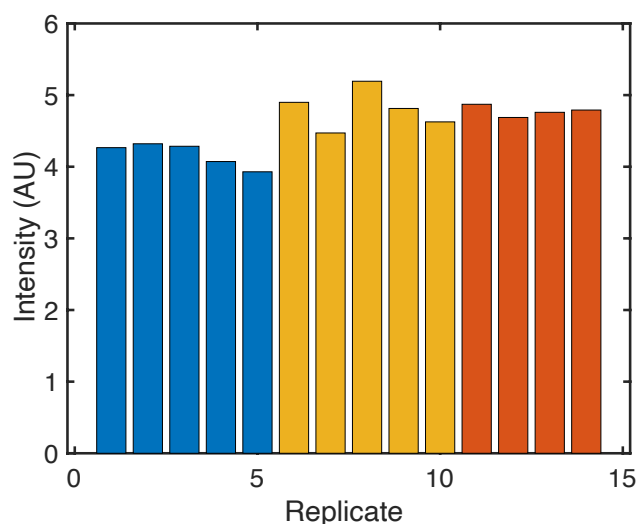


Figure 44: Integrated intensity of the lipid signal at 0.9 ppm, for multiple trials within three different sessions (blue, yellow and red). For the session in yellow, the rotation was particularly unstable. Even if the results are concordant, packing seems to have a small influence on the results. The standard deviation is approximately 10% of the signal in this case.

In order to determine if the measured lipid content for an ant is simply determined by its size, we investigated the correlation between the integrated intensity for the water and lipid signals for two ant colonies, since the intensity of the water signal should scale with the size of the ant (Figure 45). There is a large scatter between the intensity of the water and lipid signals, indicating that they are not just determined by the size of the ant, and that NMR can therefore be used to measure the corporality of the ants. As a control experiment, the integrations of the lipid CH_2 and methyl resonances were compared; in contrast, strong positive correlation is now observed, indicating that a greater lipid content increases the intensity of both.

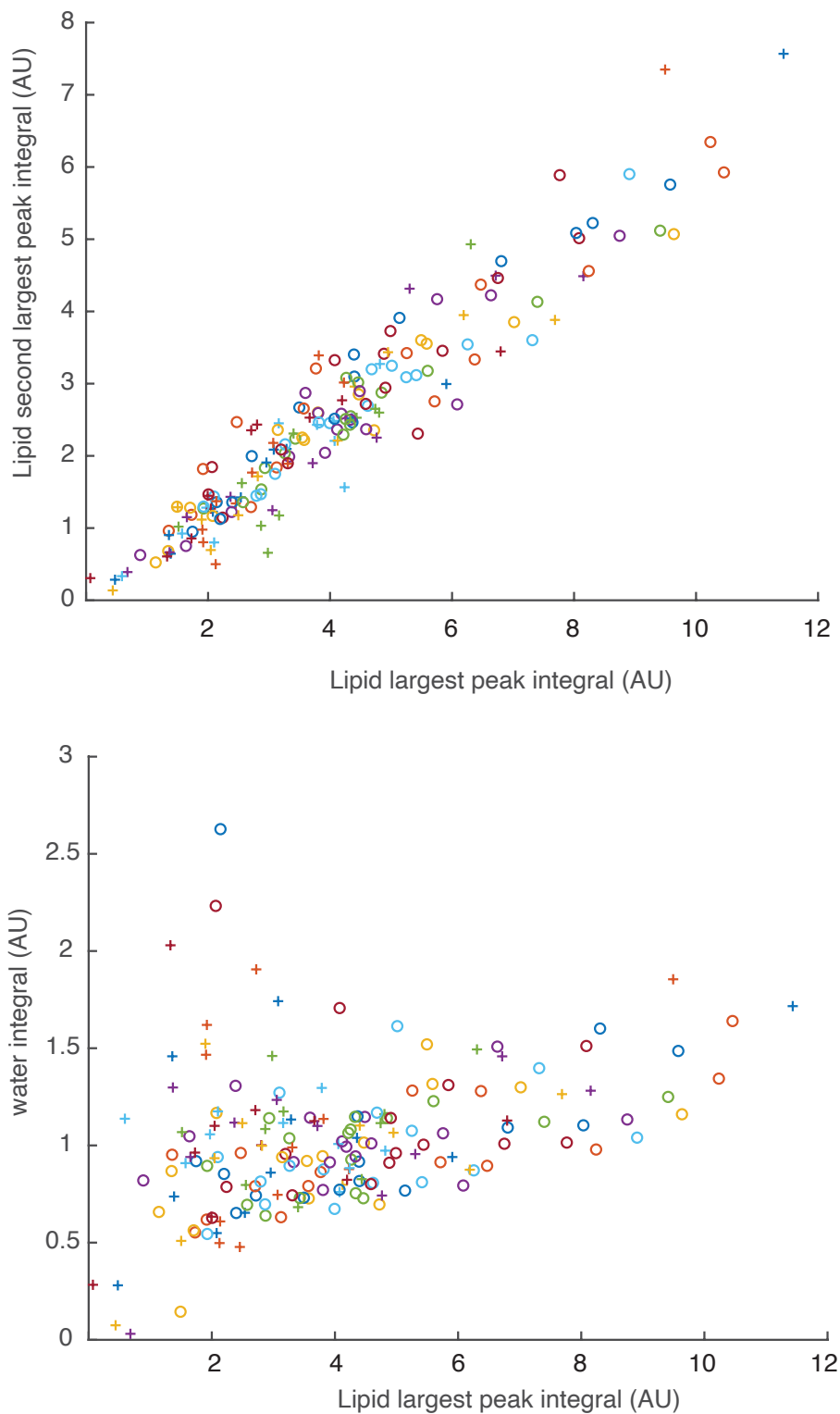


Figure 45: Integrated intensities of the lipid CH_2 (0.9 ppm) and methyl (0.5 ppm) signals for different ants. b) Integrated intensities of the water (4.3 ppm) and lipid CH_2 (X ppm) signals. In the first panel the high correlation between the two lipids signal seems to indicate that they probe the same phenomenon. The correlation is much less clear when it comes to comparing water versus lipids signal. We are not therefore not just observing the effect of the size of the ant. Each ant is represented by a color, and one colony members by crosses and the other by circles. The largest peak contains resonances from CH_2 fatty chains of lipids, and the smaller one contains resonances from methyl groups.

For the MAS measurements, 7 mm rotors were not able to achieve stable rotation, neither to preserve the ants inside, even at very moderate speed. 3.2 mm rotors have proven to be the more stable. Survival tests have been made with several species (with variable survival rate) and we decided to focus on *Myrmica rubra*.

Myrmica rubra ants are able to resist for a minute to a spinning speed of 1.2 kHz in a 3.2 mm rotor, which allows us to prevent the side bands of the water signal interacting with the lipids signals. Moreover, their stinger does not penetrate the skins of human hands which is important for such careful manipulation.

Another particularity is their recognition pattern. Most ants, will recognize the smell of the colony where they were born and raised, regardless of genetic proximity and this smell will vary with time (typically through consumption of varying edible material)⁴. This means that if separated from the main colony for too long they may be unable to recognize their sisters and will fight. For *Myrmica rubra*, this recognition has a very strong genetic component, which allows us to isolate at will any worker or group or workers without fearing future aggression.¹⁸

4.3.2 Metabolic correlation

In order to test our hypothesis stating the changes in lipids reserve drive task allocation we tried direct injection. Unfortunately, injection of lipids directly into the haemolymph, or chemical perturbation the metabolism of liquid are not compatible with MAS NMR as the injured ants does not survive the centrifugal forces.

The method that we choose was therefore starvation. We isolated identical sub population of the three colonies, and starved half on them while the others had an unrestricted access to food.

In order to test the method, we measured the lipids signals from two sub populations, foragers and nurses.

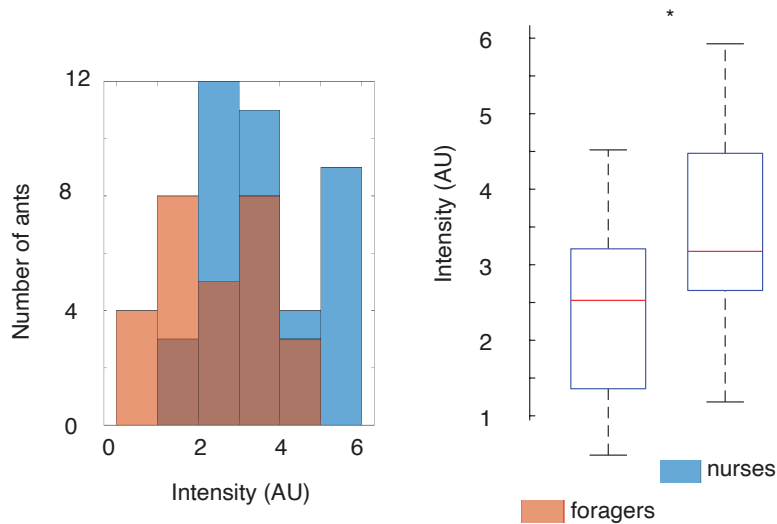


Figure 46: Lipid content measured by NMR for nurses and foragers. Left: Histogram and right: boxplot showing the median, 25th and 75th percentiles and the extreme non outlier values. The difference is significant ($p < 0.05$).

As described in Figure 46 we reproduced the results from an extensive literature by observing lean foragers and corpulent nurses.⁴

We then observed if starving them depleted their fat reserve, as it is common in many animals. The results presented in Figure 47 were positive. We therefore conclude that the method is adapted to our study.

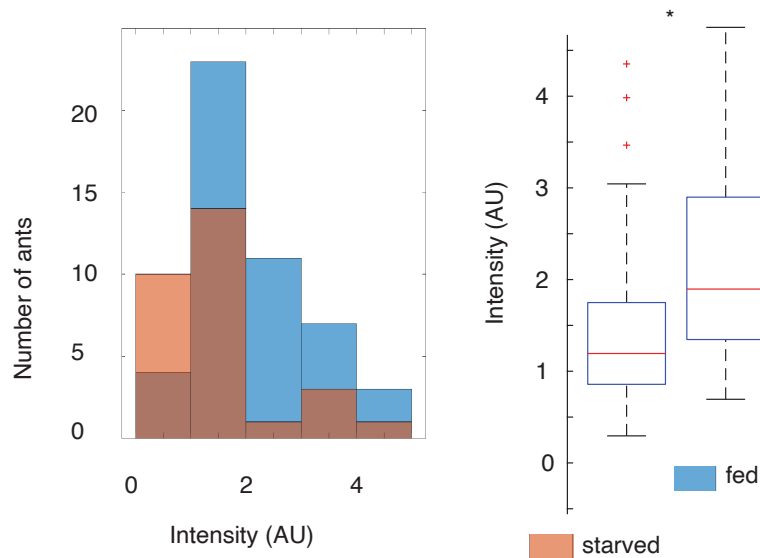


Figure 47 : Lipid content measured by NMR for fasted and well fed workers. Left: Histogram and right: boxplot showing in each case the median, 25th and 75th percentiles and the extreme non outlier values. The difference is significant ($p < 0.05$).

Using paint for age marking ant in conjunction with color analysis that is extremely specific for this species, we were able to have four colonies of 70 workers with a similar age distribution.

4.3.3 Tracking system

The method to track ants have been developed by a research group at UNIL¹⁹ and is extensively explained. This approach was first used to identify cleaners, nurses and foragers, in *Camponotus fellah*.

QR codes from an ARTag library with 36 bits of information (more than 2000 tags for a colony) were used. Our colonies were 70 workers strong at the beginning of our experiment, and we only expect less than 10 new births (ants emerging from the pupal stage to the worker stage) for the season and duration of the experiment. The QR codes weigh less than 1mg, which is at most 10% of the body-weight of our ants, and does not significantly change their behavior or death rates.¹⁹ For tagging, ants are immobilized by flexible plastic blades and the QR code is glued on their back using Sauer skin glue. This specific glue is used because ants do not try to remove it from themselves or their sisters, which happens with many man made compounds. The tags are changed after each NMR measurement

since the force exerted during the manipulation greatly damages the glue. The QR codes are presented in Figure 48.

Individually tagged ants are then placed in an artificial nest/outside box, as represented in Figure 49. No specific unusual behavior was observed after tagging. No attacks on the QR codes were detected, except in one of our four colonies. This behavior was partially related to physical constraints at the entrance of the nest, modification decreased the QR code degradation but did not totally suppress it.

To record the QR codes, a monochrome camera 4560*3048 pixel, sensitive in the infrared spectrum, was used. The lighting was assured by 5 ms infrared flashes above the nest, which are not detected by ants and are synchronized with the data acquisition.¹⁹ Such short flashes increase the resolution and have a minimal heating. Ants position are detected every half second, and video and images are saved for post processing. The angle between the ant and its tag is manually corrected to better identify the position of the ant. To discriminate a meaningful contact between two ants and just spatial proximity, the angle between the two ants bodies and the time of the interaction are recorded. The parameter describing a contact can be modified to better identify the social network.

The week long time-lapse is then analysed by an algorithm to allocate a job to each worker, as well as constructing a social network.¹⁹

The humidity as well as the temperature are controlled for the well-being of the ants and the longevity of the QR code by placing the colony (presented in Figure 49) in a insulated foam box. The temperature is kept constant at 20°C and the humidity at 60%. This is less than the optimal humidity for this species (80%) but prevents tags degradation



Figure 48: reproduced from reference²⁰, picture of ants with their QR code on the A panel. Alongside eggs, larvae and pupae, a queen and several workers of *Lassius niger*.

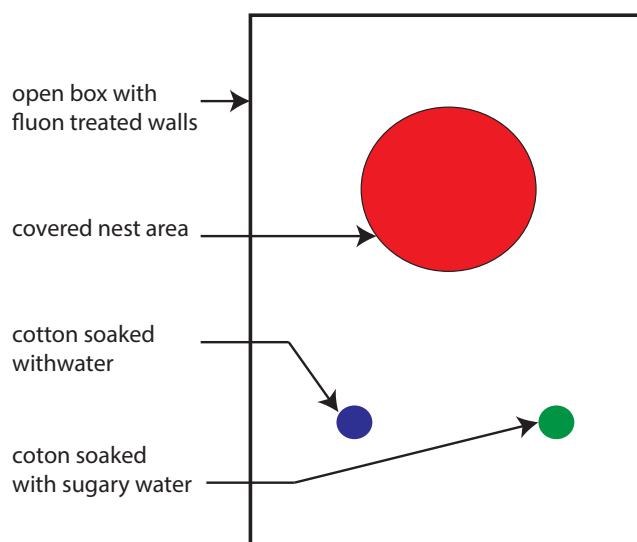


Figure 49: Schematic representation of the nest used for detection. The cover of the nest is transparent to infrared light. View from above. The fluon prevents the ants from climbing outside the restricted area.

4.3.4 Behavioural/metabolic correlation

Experiments are still underway. One example of the data being collected is presented in Figure 50. In this example one ant (blue) seems to have become leaner over the three week period, while the two others do not seem to change meaningfully. Our objective is to record those data for three months and observe the correlations with the task allocation.

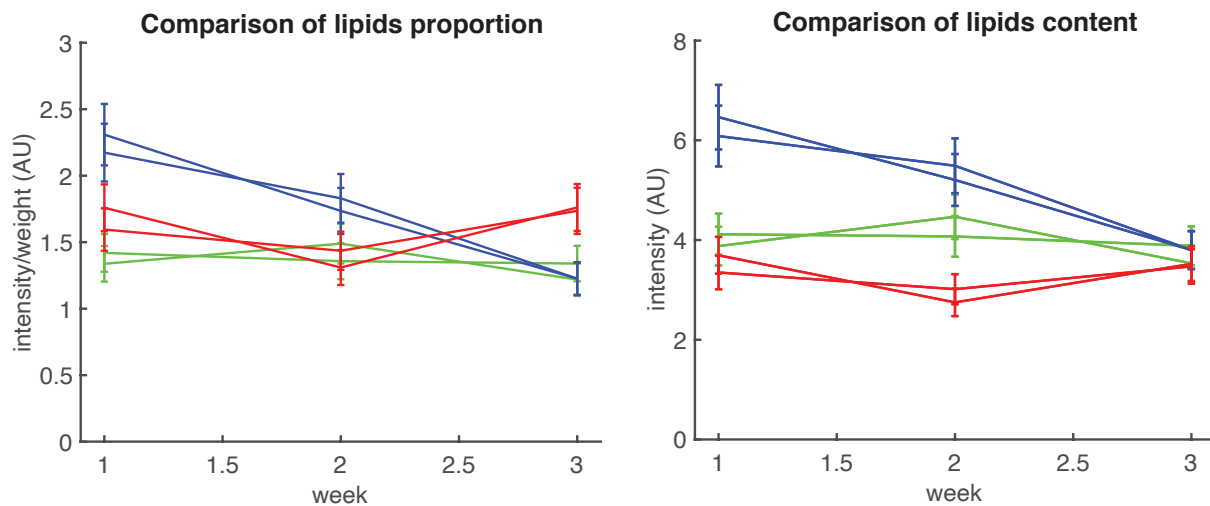


Figure 50: Absolute lipid content (right) and normalized by weight (left) for three ants of the same colony over the course of three weeks. Each week two measurements were performed.

4.4 References

1. Chan, K. H.; Gu?nard, B., Ecological and socio-economic impacts of the red import fire ant, *Solenopsis invicta* (Hymenoptera: Formicidae), on urban agricultural ecosystems. *Urban Ecosyst* **2020**, *23* (1), 1-12.
2. Kemp, S. F.; deShazo, R. D.; Moffitt, J. E.; Williams, D. F.; Buhner, W. A., 2nd, Expanding habitat of the imported fire ant (*Solenopsis invicta*): a public health concern. *J Allergy Clin Immunol* **2000**, *105* (4), 683-91.
3. Campomizzi, A. J.; Morrison, M. L.; Farrell, S. L.; Wilkins, R. N.; Drees, B. M.; Packard, J. M., Red Imported Fire Ants Can Decrease Songbird Nest Survival. *The Condor* **2009**, *111* (3), 534-537.
4. Hölldobler, B.; Wilson, E. O., *The Ants*. 1990.
5. Deneubourg, J. L.; Aron, S.; Goss, S.; Pasteels, J. M., The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior* **1990**, *3* (2), 159-168.
6. Wilson, E. O., Behavioral discretization and the number of castes in an ant species. *Behavioral Ecology and Sociobiology* **1976**, *1* (2), 141-154.
7. Huang, Z. Y.; Robinson, G. E., Honeybee colony integration: worker-worker interactions mediate hormonally regulated plasticity in division of labor. *Proceedings of the National Academy of Sciences* **1992**, *89* (24), 11726-11729.
8. Kuszewska, K.; Woyciechowski, M., Reversion in honeybee, *Apis mellifera*, workers with different life expectancies. *Animal Behaviour* **2012**, *85*.
9. Blanchard, G. B.; Orledge, G. M.; Reynolds, S. E.; Franks, N. R., Division of labour and seasonality in the ant *Leptothorax albigipennis*: worker corpulence and its influence on behaviour. *Animal Behaviour* **2000**, *59*, 723-738.
10. Pinter-Wollman, N., Nest architecture shapes the collective behaviour of harvester ants. *Biol Letters* **2015**, *11* (10).
11. Smith, C. R.; Suarez, A. V.; Tsutsui, N. D.; Wittman, S. E.; Edmonds, B.; Freau, A.; Tillberg, C. V., Nutritional Asymmetries Are Related to Division of Labor in a Queenless Ant. *Plos One* **2011**, *6* (8).
12. Odonnell, S.; Jeanne, R. L., Implications of Senescence Patterns for the Evolution of Age Polyethism in Eusocial Insects. *Behav Ecol* **1995**, *6* (3), 269-273.
13. Koto, A.; Motoyama, N.; Tahara, H.; McGregor, S.; Moriyama, M.; Okabe, T.; Miura, M.; Keller, L., Oxytocin/vasopressin-like peptide inotocin regulates cuticular hydrocarbon synthesis and water balancing in ants. *P Natl Acad Sci USA* **2019**, *116* (12), 5597-5606.
14. Robinson, E. J.; Feinerman, O.; Franks, N. R., Flexible task allocation and the organization of work in ants. *Proc Biol Sci* **2009**, *276* (1677), 4373-80.
15. Roma, G. C.; Bueno, O. C.; Camargo-Mathias, M. I., Chemical detection of the proteins and lipids in the fat body cells from workers of *Attini* ants (Hymenoptera: Formicidae). *Cell Biol Int* **2008**, *32* (4), 406-16.
16. Roma, G. C.; Bueno, O. C.; Camargo-Mathias, M. I., Morpho-physiological analysis of the insect fat body: a review. *Micron* **2010**, *41* (5), 395-401.

17. Schilling, F.; Dworschak, K.; Schopf, R.; Kuhn, R.; Glaser, S. J.; Haase, A., Non-invasive lipid measurement in living insects using NMR microscopy. *J Exp Biol* **2012**, *215* (18), 3137-3141.
18. Guillem, R. M.; Drijfhout, F. P.; Martin, S. J., Species-Specific Cuticular Hydrocarbon Stability Within European *Myrmica* Ants. *J Chem Ecol* **2016**, *42* (10), 1052-1062.
19. Mersch, D. P.; Crespi, A.; Keller, L., Tracking Individuals Shows Spatial Fidelity Is a Key Regulator of Ant Social Organization. *Science* **2013**, *340* (6136), 1090-1093.
20. Stroeymeyt, N.; Grasse, A. V.; Crespi, A.; Mersch, D. P.; Cremer, S.; Keller, L., Social network plasticity decreases disease transmission in a eusocial insect. *Science* **2018**, *362* (6417), 941-945.

Curriculum Vitae

3 Avenue de la Poste 1020
Renens
Switzerland

Phone: +33674452381
Email: baptiste.busi@epfl.ch ; busi.baptiste@gmail.com

Birth: July 07, 1993

Nationality: French

Education

- | | |
|-----------|---|
| 2016-2020 | PhD thesis (Chemistry)
<i>École Polytechnique Fédérale de Lausanne, Switzerland</i>
Dynamics of protein using Nuclear Magnetic Resonance Relaxometry, directed by Lyndon Emsley |
| 2013-2016 | Master of Science (Bioscience)
<i>École Normale Supérieure de Lyon , France</i>
Focus on biophysics, biochemistry, Nuclear Magnetic Resonance and metabolomics |
| 2012-2013 | Bachelor of Science (Biology)
<i>École Normale Supérieure de Lyon , France</i> Broad introduction to biology |
| 2010-2012 | Classe préparatoire aux grandes écoles (BCPST)
<i>Lycée Victor Hugo, Besançon, France</i>
Two-year intensive course preparing for the competitive entrance examination to French institution of higher education, with Major on biology, chemistry, physics and geology |
| 2009-2010 | Scientific Baccalauréat (Biology major)
<i>Lycée Paul Émile Victor, Champagnole, France</i> equivalent to high school diploma, <i>summa cum laude</i> |

Research

- | | |
|-----------|--|
| 2016-2020 | Dynamics of protein using Nuclear Magnetic Resonance Relaxometry <i>École Polytechnique Fédérale de Lausanne, Switzerland</i>
directed by Lyndon Emsley |
| 2015 | Metabolomic signature studies of thyroid receptor deficiency on mice <i>Centre de Résonance Magnétique Nucléaire, France</i>
directed by Gilles Rautureau |

2014-2015	Natural Deep Eutectic Solvent for solid state Nuclear Magnetic Resonance studies on β -amyloid peptide <i>Technische Universität München</i> , Germany directed by Bernd Reif
2014	Preliminary structural data about the La related protein 6 RNA interaction <i>King's College London</i> , United Kingdom directed by Sasi Conte
2013	Biochemical characterisation of the phospholipase D <i>Université Claude Bernard Lyon 1</i> , France directed by Alexandre Noiriél

Teaching

2014-2020	<i>École Polytechnique Fédérale de Lausanne, Université de Lausanne</i> , Switzerland Practical for PhD student in Nuclear Magnetic Resonance Practical for 1 st years students in chemistry Guidance for a bachelor project Exercises for bachelor chemistry students
-----------	---

Skills

French driving license

Languages

French	Native
English	Fluent
German	Medium Intermediate

Computer

OS	Linux, Windows, Macintosh
Coding	Python, Matlab
Edditor	Latex, Office

Other experiences

2012-2013	Communication manager <i>Association Sportive de l'École Normale Supérieure de Lyon</i> , France as.ens-lyon.fr
2009-2013	Camp counselor <i>Ville de champagnole</i> , France Children summer camp

Interests

Sports Contact sport: Judo (black belt and 10 years of competition), Muay-thai & MMA, Climbing, Hiking, Skiing

Literature Maupassant, Bradley, Vian, Orwel, Jaworski, Pratchett

Publications & talks

Journal articles

1. Luigi Martino, Simon Pennell, Geoff Kelly, **Baptiste Busi**, Paul Brown, R Andrew Atkinson, Nicholas JH Salisbury, Zi-Hao Ooi, Kang-Wei See, Stephen J Smerdon, Caterina Alfano, Tam TT Bui, Maria R Conte, *Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module*
Nucleic acids research, **2015** 43, 645-660
<https://doi.org/10.1093/nar/gku1287>
2. Riddhiman Sarkar, Andi Mainz, **Baptiste Busi**, Emeline Barbet-Massin, Maximilian Kranz, Thomas Hofmann, Bernd Reif, *Immobilization of soluble protein complexes in MAS solid- state NMR: Sedimentation versus viscosity*
Solid state nuclear magnetic resonance, **2016** 76, 7-14
<https://doi.org/10.1016/j.ssnmr.2016.03.005>
3. **Baptiste Busi**, Jayasubba Reddy Yarava, Albert Hofstetter, Nicola Salvi, Diane Cala- de Paepe, Józef R Lewandowski, Martin Blackledge, Lyndon Emsley, *Probing Protein Dynamics Using Multifield Variable Temperature NMR Relaxation and Molecular Dynamics Simulation*,
The Journal of Physical Chemistry B, **2018** 122, 9697-9702
<https://doi.org/10.1021/acs.jpcb.8b08578>
4. Houda Boumaza, Suzy Markossian, **Baptiste Busi**, Gilles JP Rautureau, Karine Gauthier, Bénédicte Elena-Herrmann, Frederic Flamant, *Metabolomic Profiling of Body Fluids in Mouse Models Demonstrates that Nuclear Magnetic Resonance Is a Putative Diagnostic Tool for the Presence of Thyroid Hormone Receptor $\alpha 1$ Mutations*
Thyroid, **2019** 29, 1327-1335
<https://doi.org/10.1089/thy.2018.0730>
5. **Baptiste Busi**, Jayasubba Reddy Yarava, Andrea Bertarello, François Freymond, Wiktor Adamski, Damien Maurin, Matthias Hiller, Hartmut Oschkinat, Martin Blackledge, and Lyndon Emsley, *Similarities and Differences Among Protein Dynamics Studied by Variable Temperature NMR Relaxation*.
The Journal of Physical Chemistry, **submitted**