



**HAL**  
open science

# Advanced NMR methods to investigate complex biological systems

Ziqing Wang

► **To cite this version:**

Ziqing Wang. Advanced NMR methods to investigate complex biological systems. Theoretical and/or physical chemistry. Sorbonne Université, 2020. English. NNT : 2020SORUS455 . tel-03771333

**HAL Id: tel-03771333**

**<https://theses.hal.science/tel-03771333v1>**

Submitted on 7 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Sorbonne Université

École doctorale 388 Chimie Physique et Chimie Analytique de Paris Centre

*Laboratoire des Biomolécules / Équipe 3 : Structure et Dynamique des Biomolécules*

# **Advanced NMR methods to investigate complex biological systems**

Par Ziqing Wang

Thèse de doctorat de Sciences de la Sorbonne Université

Dirigée par Fabien Ferrage

Présentée et soutenue publiquement le 19 mars 2020

---

	Dr. Sophie Zinn-Justin	Rapporteuse
	Pr. Nicolas Giraud	Rapporteur
	Pr. Laurence Le Moyec	Examinatrice
Jury:	Pr. Giacomo Parigi	Examineur
	Dr. Pavel Kadeřávek	Examineur
	Dr. Guillaume Bouvignies	Co-encadrant
	Dr. Fabien Ferrage	Directeur de thèse

---



## Acknowledgments

I am thankful to all the members of my thesis committee: **Dr. Sophie Zinn-Justin**, **Pr. Nicolas Giraud**, **Pr. Laurence Le Moyec**, **Pr. Giacomo Parigi** and **Dr. Pavel Kadeřávek**, for accepting reviewing my work from the last three years.

I would like to express my thanks to my supervisor, **Fabien Ferrage**, for his trust and support during my thesis. His professionalism and enthusiasm for the field lighted up the way forward for the thesis at dark moments. His understanding and tolerance encouraged me to get over the difficulties.

I would also like to acknowledge my co-supervisor, **Guillaume Bouvignies**, who was patient and willing to help at any time. Both the brain storms, and the ice-cream breaks were nice moments to share with him.

I thank the director of the lab, Dr. **Sandrine Sagan**, for taking me in the lab. I thank the team leaders, Pr. **Geoffrey Bodenhausen** and Dr. **Daniel Abergel** for including me in their team. I also want to thank **Eliane**, for taking great care of all necessary administration work, especially the contracts.

This work would not have been possible without our collaborators. **Thorsten** and **Jean-Max** from Bruker Biospin have been working on the development of the instrument. It is always nice time when they come to the lab, with their delightful smiles and newest updates on the system. It was almost like Christmas! Collaborating with **Veronica**, **Giacomo** and **Claudio** was an honor. It was pleasant visiting their laboratory in Florence, meeting them during conferences and discussing science with them.

Many team members contributed to my projects with their expertise. **Ludovic**, **Olivier** and **Jean-Jacques** were always nice and ready-to-act when I bothered them with many questions in the wet-lab. **Pavel** and **Philippe** coached me with the spectrometers and the experiments. They kept surprising me with their knowledge, skills and passion for science. **Simone** and **Duy** joined our group only very recently, but are already bringing many new thoughts, efforts and results to our projects.

Working between two campus can be really difficult to handle, unless there is good company. I was lucky enough to enjoy two different working environments. I express my appreciation for the warm working atmosphere created by my colleagues from Jussieu: **Ludovic, Olivier, Lucie, Jean-Jacques, Damien, Shadaï, Rachel** and **Soha**. I appreciate equally the crew from ENS: **Philippe, Pavel, Milan, Simone, Dennis, Diego, Mathieu, Nicolas**, and the students: **Xiao, Sina, Emmanuelle, le petit Nicolas, David, Behdad, Thomas, Morgan, Jessie, Duy** and **Vineeth**. I loved working with these brilliant scientists, as well as having after-work “seminars” with them.

My eleven-year-journey in France was always highlighted by lovely friends. It started with a group of young Chinese: Guanming, Yutian, Xiaqing, Yinan, Yi garçon and Yi fille. We shared laughs and tears and witnessed the growth of each other. I am happy that we all found our belongings. Then, I met my “Descartes Power” friends: Chayma, Fatoumia, Claire, Abby, Camélia, Victor et Fabrice, without whom I could never survive school. I treasure our long-last friendship despite that we are apart now. Sometime later, I discovered funs outdoor, and mates to share it with. I am continuously amazed and infected by their energy, especially by Rosie and Huitong. Last but not the least, some of my dearest friends are from the lab. Nicolas, Behdad, Milan, David and Jessie made me feel cared, seen and affirmed all time. I wish that I can keep sharing my next journey will all my friends, no matter where we are going to be.

I can never thank too much my parents. I am grateful everything they provided me, a big loving family, an open environment and the long-term financial support.

More importantly, I want to thank Guanming, who supported me unconditionally in all possible ways. I look forward to our lives back together.

Writing the acknowledgements was an extremely emotional process. There are many more people that I want to thank, but two pages are far from enough to tell the whole story. I am so fortunate to receive this much support. I wish to thank everyone in person in the coming days.

*Ziqing*

当你想当画家时，就怀疑自己是色盲；  
想当音乐家时，就怀疑自己是聋子；  
想当思想家，就怀疑自己是个大傻瓜。  
如果没有那种毛病，你就不会想当那种人。

——王小波，《黄金时代》

Lorsque tu veux être peintre, tu te sens que tu es daltonien ;  
Quand tu veux être musicien, tu te doutes que tu es sourd ;  
Si tu veux être un penseur, tu penses que tu es un idiot.  
Si tu n'avais pas ce genre de questionnement, tu ne voudrais pas être ce genre de personne.

Wang Xiaobo, *l'Age d'or*



The mechanisms of life rely on thousands of biomolecules and their dynamic interactions. From the simple water molecules, to small metabolites and macromolecules, they work together and form a complex network. Determining their structures, dynamics and interactions are essential information for the understanding of their diverse biological functions. Nuclear magnetic resonance spectroscopy is a unique tool to investigate molecular conformation and motions over a large range of timescales. In this thesis, we present our work of investigating complex biomolecular system using advanced NMR methods.

Metabolite-protein interactions is an emerging subject in interactomics, but lacks systematic analytical methods. We used high-resolution NMR relaxometry to identify interactions between metabolites and macromolecules in complex biological samples. The method is based on the measurement and analysis of longitudinal relaxation rates over three orders of magnetic field (from 15 mT to 14.1 T) with high-resolution detection at 600 MHz. This approach was first developed and validated on a model sample, then applied to biological fluids relevant for metabolomics studies: human blood plasma and serum. The size dispersion of macromolecules in human blood was estimated with Fast Field Cycling (FFC) NMR. Identified metabolites revealed different features in high-resolution relaxation dispersion profiles. The interactions of some metabolites with macromolecules were characterized. Furthermore, we show that relaxation dispersion profiles are sensitive to competitive binding between small molecules. High-resolution relaxometry is an efficient tool for the investigation of dynamic interactions in biological fluids, and provides valuable information complementary to conventional metabolomics approaches, which may pave the way for NMR in interactomics.

In parallel, we implemented different isotope labeling methods and NMR strategies to study two different proteins. Full length human XLF protein was analyzed by high-resolution NMR. Its intrinsically disordered region was assigned and its interaction with DNA was characterized for the first time by NMR. We also explored challenging selective isotope labeling on *S. cerevisiae* polymerase  $\eta$ .





# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b>	<b>I</b>
<b>ABSTRACT</b>	<b>V</b>
<b>RÉSUMÉ</b>	<b>9</b>
<b>1. INTRODUCTION</b>	<b>10</b>
<b>2. INVESTIGATIONS D'INTERACTIONS METABOLITE-PROTEINE PAR LA RELAXOMETRIE A HAUTE RESOLUTION</b>	<b>11</b>
<b>3. ÉTUDES DU DESORDRE INTRINSEQUE DE LA PROTEINE XLF</b>	<b>22</b>
<b>4. MARQUAGE SELECTIF D'UNE PROTEINE DE GRANDE TAILLE : L'ADN POLYMERASE II</b>	<b>27</b>
<b>CHAPTER I NUCLEAR SPIN RELAXATION AND MOTIONS</b>	<b>31</b>
<b>1 INTRODUCTION</b>	<b>32</b>
<b>2 THE BLOCH-REDFIELD-WANGSNES (BRW) THEORY</b>	<b>32</b>
<b>3 CHEMICAL EXCHANGE</b>	<b>42</b>
<b>CHAPTER II BINDING OF SMALL MOLECULES TO MACROMOLECULES PROBED BY HIGH-RESOLUTION RELAXOMETRY</b>	<b>49</b>
<b>1 INTRODUCTION</b>	<b>50</b>
<b>2 NMR EXPERIMENTS TO INVESTIGATE INTERACTIONS OF SMALL MOLECULES WITH MACROMOLECULES</b>	<b>50</b>
<b>3 A MODEL SYSTEM FOR AN INTERACTION BETWEEN SMALL AND LARGE MOLECULES</b>	<b>56</b>
<b>4 RELAXATION THEORY</b>	<b>57</b>
<b>5 EXPERIMENTAL SECTION</b>	<b>61</b>
<b>6 RESULTS AND DISCUSSION</b>	<b>65</b>
<b>7 CONCLUSION</b>	<b>70</b>
<b>CHAPTER III INVESTIGATION OF METABOLITE-PROTEIN INTERACTIONS IN BIOFLUIDS BY HIGH-RESOLUTION NMR RELAXOMETRY</b>	<b>71</b>
<b>1 INTRODUCTION</b>	<b>72</b>
<b>2 HUMAN BLOOD SAMPLE</b>	<b>77</b>

<b>3</b>	<b>RESULTS</b>	<b>80</b>
<b>4</b>	<b>DISCUSSION AND CONCLUSION</b>	<b>87</b>
<b>CHAPTER IV    <u>PROBING THE INTRINSIC DISORDER OF THE HUMAN PROTEIN XLF</u></b>		<b>91</b>
<b>1</b>	<b>GENERAL INTRODUCTION</b>	<b>92</b>
<b>2</b>	<b>BIOLOGICAL ROLE OF XLF IN DNA REPAIR</b>	<b>93</b>
<b>3</b>	<b>SAMPLE PREPARATION</b>	<b>98</b>
<b>4</b>	<b>NMR ANALYSIS</b>	<b>101</b>
<b>5</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>102</b>
<b>6</b>	<b>CONCLUSION</b>	<b>109</b>
<b>CHAPTER V    <u>DNA POLYMERASE H: PROTEIN EXPRESSION AND ISOTOPE LABELING</u></b>		<b>111</b>
<b>1</b>	<b>TRANSLESION SYNTHESIS AND POLYMERASE H</b>	<b>112</b>
<b>2</b>	<b>PROTEIN EXPRESSION AND LABELLING</b>	<b>113</b>
<b>3</b>	<b>RESULTS AND DISCUSSION</b>	<b>121</b>
<b>4</b>	<b>CONCLUSION</b>	<b>124</b>
<b><u>BIBLIOGRAPHY</u></b>		<b>127</b>

Résumé

## 1. Introduction

La spectroscopie de résonance magnétique nucléaire est une technique qui exploite des propriétés magnétiques des spins nucléaires. La RMN est appliquée dans des domaines très différents, allant de l'imagerie par résonance magnétique (IRM), à la caractérisation des substances à l'état solide, liquide ou gazeuse. Elle est aujourd'hui l'un des moyens les plus puissants pour étudier la structure et la dynamique des systèmes moléculaires complexes.

La relaxation est le retour à l'équilibre des spins nucléaires. Elle est un outil important en RMN pour sonder les mouvements moléculaires. Ces mouvements moléculaires induisent des fluctuations locales des interactions des noyaux avec leur environnement, qui conduisent à la relaxation des spins nucléaires. L'échange chimique, la diffusion rotationnelle, la diffusion translationnelle et la flexibilité moléculaire influencent les propriétés de relaxation. Ces mouvements, essentiels pour la fonction des biomolécules, couvrent une large gamme d'échelles de temps allant de quelques picosecondes à quelques secondes, et une large gamme d'amplitudes allant de la rotation très locale d'un groupe méthyle aux mouvements de domaines entiers de protéines associés à l'allostérie.

La RMN en solution fournit de nombreuses informations sur les molécules dans des conditions proches de l'environnement biologique. Le travail de thèse présenté dans ce manuscrit concerne le développement et les applications de nouvelles méthodes de RMN pour la caractérisation de la dynamique de différents systèmes biologiques :

1. Développement d'une nouvelle méthode pour étudier des interactions métabolite-protéine par la relaxométrie à haute résolution.
2. Investigation des interactions métabolite-protéine dans des milieux biologiques complexes par la relaxométrie à haute résolution.
3. L'étude du désordre intrinsèque de la protéine XLF et son interaction avec l'ADN.
4. La production d'une protéine structurée de grande taille, l'ADN polymérase  $\eta$ , l'optimisation de conditions de l'échantillon et le marquage sélectif de groupes méthyl des chaînes latérales.

## 2. Investigations d'interactions métabolite-protéine par la relaxométrie à haute résolution

### 1.1 Interaction métabolite-protéine

Les métabolites affectent les activités cellulaires de manière complexe. Non seulement les métabolites sont des substrats et des produits de réactions enzymatiques, mais ils sont également des cofacteurs et des régulateurs d'un grand nombre de protéines. La métabolomique, est la caractérisation de l'ensemble des métabolites dans un échantillon biologique. Les métabolites se lient aux macromolécules, généralement des protéines, pour exercer leur fonction dans la signalisation, la régulation et d'autres processus cellulaires. L'étude systématique des interactions métabolite-protéine est le sujet de l'interactomique. Notre compréhension des interactions métabolite-protéine (metabolite-protein interactions ou MPIs) est très limitée, faute d'analyse systématique. Des nouvelles approches sont en cours de développement, pour répondre au besoin de détermination d'une image complète du réseau biologique.

Les MPI sont souvent étudiées de manière indirecte. Les méthodes biologiques d'investigation du MPI impliquent généralement une modification chimique du métabolite ou un marquage de la protéine. Ceci limite le sujet à l'interaction entre une protéine individuelle et les métabolites. En outre, les MPI ont des origines diverses : catalytique, allostérique, affinité hydrophobe, par liaison hydrogène ou d'autres mécanismes non spécifiés. La durée de vie du complexe peut être très variable. Les interactions faibles sont souvent difficiles à conserver lors des étapes de préparation. La RMN est une technique intéressante pour détecter et caractériser ces interactions, pour sa capacité à travailler avec des échantillons biologiques en solution et l'absence de grosse sonde (par exemple fluorophore).

Pour observer les interactions entre les petites molécules (par exemple ligand) et les macromolécules (par exemple protéine), nous devons quantifier un paramètre qui diffère considérablement entre les formes libres et liées. Une telle observation peut être faite à partir d'une petite molécule ou d'une macromolécule. L'expérience des relations structure-activité (structure-activity relationship ou SAR) par RMN est une méthode classique basée sur les récepteurs protéiques. Elle consiste à surveiller la perturbation du déplacement chimique de la protéine par HSQC  $^1\text{H}$ - $^{15}\text{N}$ . Cette approche offre une cartographie d'interaction à l'échelle du résidu. Cependant, elle est limitée à un récepteur de taille relativement petite et nécessite non seulement des échantillons hautement purifiés et concentrés (plusieurs milligrammes), mais

aussi de connaître les attributions des résonances de la protéine cible, ce qui n'est pas nécessairement aisé. Pour les systèmes plus grands (> 40 kDa), la perdeutération, l'échantillon marqué au  $^{13}\text{C}$  (méthyle) et les expériences de type spectroscopie avec relaxation transversale optimisée (transverse relaxation-optimized spectroscopy ou TROSY), permettent d'augmenter la sensibilité. Cependant, le processus de préparation des échantillons est difficile et ne convient pas pour l'étude des fluides biologiques. De nos jours, les méthodes basées sur les ligands sont plus populaires dans la pratique, car il n'y a aucune exigence de marquage isotopique et peu de restrictions sur la taille de la macromolécule. Ces méthodes sont basées sur le transfert de différentes propriétés physiques des macromolécules à la petite molécule de liaison via le complexe.

## 1.2 Méthodes de RMN classiques pour étudier des interactions

Les expériences de mesure de la diffusion par la RMN sont une classe de méthodes pour étudier des interactions. Elles sont parfois appelées RMN d'affinité dans le criblage de médicaments. En mettant en œuvre des filtres de diffusion, il est possible de distinguer les molécules à diffusion rapide (petites molécules libres) des molécules à diffusion lente (macromolécules et petites molécules interagissant avec les macromolécules). Cette méthode est basée sur une différence significative des coefficients de diffusion translationnelle ( $D$ ) entre le ligand et les macromolécules.

L'expérience de RMN par diffusion est un outil populaire pour sonder les interactions. Pourtant, de nombreuses complications peuvent interférer avec l'analyse. Premièrement, le signal de la macromolécule peut interférer avec le signal de petites molécules, ce qui rend l'identification des signaux des petites molécules difficile. Il est possible d'utiliser un filtre isotopique pour simplifier le spectre, à condition que la protéine soit marquée isotopiquement. Deuxièmement, la diffusion translationnelle n'est pas très sensible à la taille de la molécule. D'après la relation de Stokes-Einstein-Debye, la diffusion translationnelle est proportionnelle à l'inverse du rayon de la particule brownienne. En comparaison, le temps de corrélation de la diffusion rotationnelle ( $\tau$ ) est proportionnel à l'inverse du cube du rayon. La diffusion translationnelle n'est pas aussi sensible que la diffusion rotationnelle (*Fig. 1*).

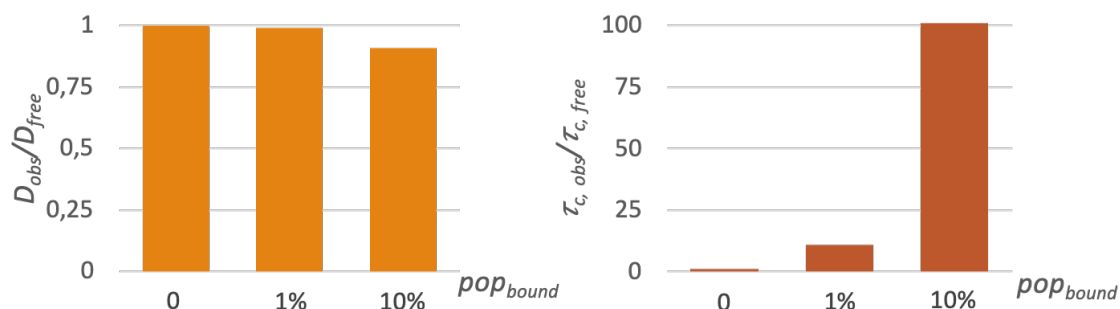


Figure 1 Comparaison des valeurs d'observation du coefficient de diffusion translationnelle ( $D$ ) et du temps de corrélation de la diffusion rotationnelle ( $\tau$ ) en fonction de la population de la forme liée à une cible. Le complexe est considéré comme ayant un rayon dix fois plus grand que celui de la petite molécule de forme libre. Toutes les valeurs sont normalisées par les valeurs des molécules libres.

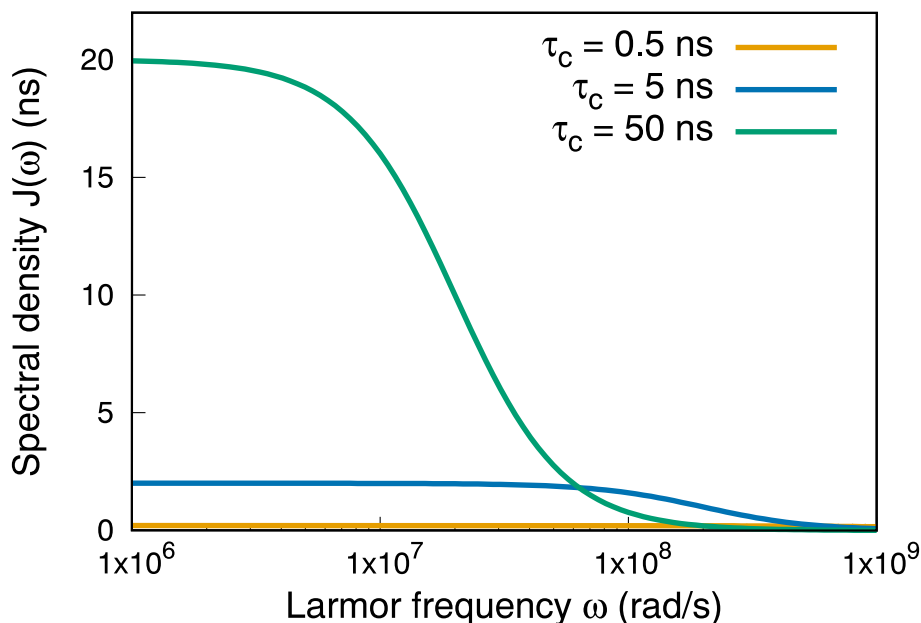
L'effet Overhauser nucléaire (nuclear Overhauser effect ou NOE) est le transfert de polarisation d'un spin à un autre à travers l'espace via la relaxation due à l'auto-corrélation d'une interaction dipôle-dipôle. Le NOE est utilisé par une classe d'expériences pour détecter les distances entre un spin nucléaire et ses voisins proche dans l'espace. La relaxation croisée due aux interactions dipôle-dipôle a une forte dépendance à la distance et aux mouvements. Quand un ligand est lié à son récepteur, les distances internucléaires courtes et une diffusion rotationnelle lente permettent un tel transfert de polarisation entre le récepteur et le ligand. Cela permet de distinguer une molécule liée des molécules libres. Les expériences de transfert de polarisation présentent de nombreux avantages : elles ont, notamment, peu de restrictions sur la taille de la macromolécule et ne nécessitent pas de marquage isotopique.

Bien que les méthodes NOE soient très populaires, elles fournissent des informations qualitatives mais pas de données quantitatives. L'efficacité du NOE dépend de plusieurs facteurs (densité de protons, dynamique locale, etc.), de sorte qu'il est possible de détecter la liaison d'une petite molécule à une macromolécule, mais plus difficile de déterminer la cinétique et de la thermodynamique de la liaison à partir de mesures de NOE seules.

Les vitesses de relaxation peuvent aussi être d'excellentes sondes pour la taille moléculaire. La relaxométrie consiste à mesurer la relaxation sur une large gamme de champs magnétiques. La relaxométrie est une sonde unique de la dynamique moléculaire. La variation des vitesses de relaxation avec le champ magnétique aide à identifier les mouvements sur une large gamme d'échelles de temps. Le cyclage de champ rapide (fast field cycling ou FFC) est une technique populaire employée par la relaxométrie. Un relaxomètre FFC est un système où



le champ peut être commuté entre plusieurs kHz et 42 MHz ( $\sim 100 \mu\text{T}$  à 1 T) en quelques millisecondes. L'échantillon est polarisé et détecté à un champ relativement élevé pour une meilleure sensibilité. Le champ magnétique est diminué rapidement pour le délai de relaxation. La mesure d'une série de vitesses de relaxation sur des ordres de grandeur de champs magnétiques permet de sonder intensivement la fonction de densité spectrale, et donc le temps de corrélation rotationnel (*Fig. 2*).



*Figure 2* Tracé de la fonction de densité spectrale sur une fonction de la fréquence de Larmor pour différentes valeurs du temps de corrélation de la diffusion rotationnelle  $\tau_c$ .

La relaxométrie FFC est une technique à basse résolution. La relaxation collective mesurée provient des noyaux de l'ensemble de l'échantillon. L'obtention d'informations sur les sites spécifiques dans une macromolécule ou d'informations sur les différents composants d'un mélange complexe est difficile, et souvent impossible avec la relaxométrie FFC. Pour déterminer les mouvements à l'échelle nanoseconde avec une résolution atomique, la relaxométrie haute résolution est une approche unique. Ici, on présente une nouvelle méthode d'identification et de quantification des interactions dans un milieu complexe. La méthode a été conçue avec un échantillon modèle, puis appliquée sur des fluides biologiques.

### 1.3 Relaxometrie de haute résolution

Le principe de la relaxométrie haute résolution consiste à utiliser le champ de fuite de l'aimant d'un spectromètre RMN à champ élevé comme champ magnétique variable pour la

relaxométrie. La combinaison de la relaxation aux champs magnétiques faibles et de la RMN à haute résolution au champ magnétique fort est obtenue avec une navette qui déplace rapidement l'échantillon entre le centre magnétique et les positions choisies dans le champ de fuite. Nous avons utilisé un système de navette pneumatique, développé par nos collaborateurs de Bruker Biospin, que nous avons installé sur un spectromètre RMN 600 MHz, équipé d'une sonde spécialement conçue (Fig. 3). Le tube contenant l'échantillon a été adapté afin de résister aux chocs mécaniques induits par les arrêts aux positions haute et basse de son trajet. Les détecteurs optiques placés en position haute et basse envoient leurs informations au contrôleur de la navette (via des satellites). Un script de paramètres prédéfini implémenté dans le contrôleur de navette principal ajuste le débit d'air requis pour déplacer l'échantillon à l'intérieur d'un guide navette. Le gradient de champ couvre tous les champs entre 14,1 T et environ 15 mT sur la longueur du guide-navette. Notez que les ferroschims installés dans le trou de l'aimant modifient localement le champ magnétique et créent un plateau à 0,33 T. Ces ferroschims ont été installés pour effectuer des expériences de RMN à deux champs.

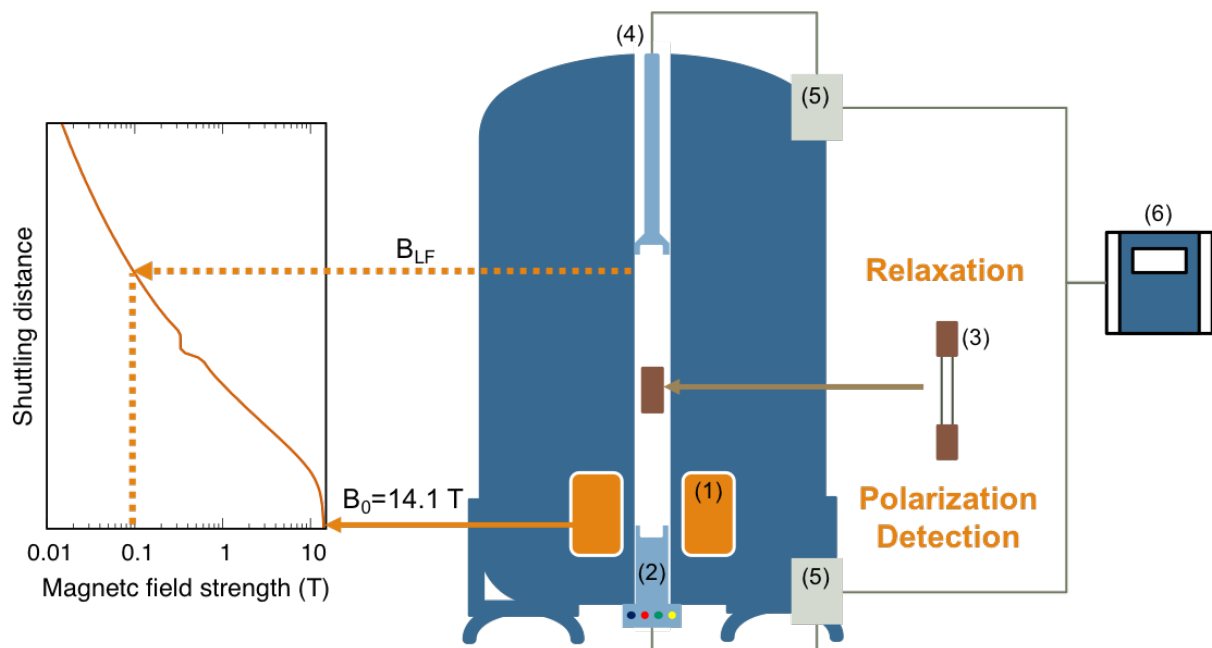
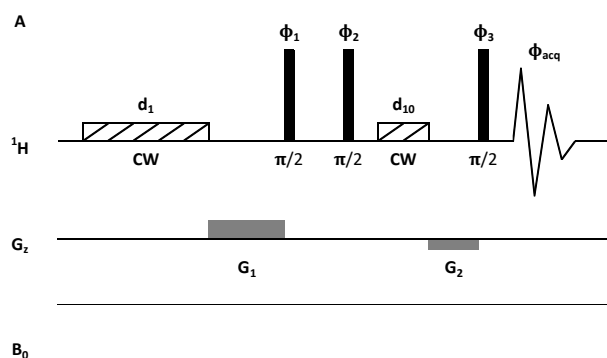


Figure 3 Vue schématique du système relaxomètre navette. (1) aimant fonctionnant à 14,1 T ; (2) Sonde navette à triple résonance  $^1\text{H} / ^{13}\text{C} / ^{15}\text{N}$  équipée d'un gradient d'axe z ; (3) tube navette avec un volume actif de 60  $\mu\text{L}$  et ses bouchons d'extrémité (rectangles bruns) protégeant les tubes pendant le transfert et l'atterrissage ; (4) stoppeur pour ajuster la position de relaxation dans l'alésage de l'aimant ; (5) navettes contrôleurs satellites pour position haute et basse ; (6) contrôleur de navette principal.

Pour étudier des interactions, nous avons adapté une séquence d'impulsions 1D standard utilisée pour la métabolomique et inséré des transferts de navette entre le centre magnétique et un plus bas champ, ainsi que les délais de relaxation. La séquence d'impulsions 1D NOESY (Fig. 4A) est l'une des expériences conventionnelles utilisées en métabolomique par RMN. Le schéma d'impulsions est similaire à l'expérience NOESY 2D mais le temps de mélange est fixé à  $\tau_{\text{mix}} = 10$  ms. L'expérience de relaxométrie a été dérivée de cette séquence d'impulsions 1D NOESY conventionnelle en introduisant un délai de relaxation à faible champ entre le temps de mélange et la dernière impulsion  $\pi/2$  (Fig. 4B). L'échantillon est transféré vers le bas-champ  $B_{\text{LF}}$  souhaité (de 0,015 à 10 T) avec un temps de transfert  $\tau_{\text{HF-LF}}$  ( $\sim 55 - 210$  ms), maintenu dans cette position pendant un délai de relaxation variable  $T_{\text{rel}}$  et transféré de nouveau vers le haut-champ avec un temps de transfert  $\tau_{\text{LF-HF}}$  ( $\sim 60 - 260$  ms) pour la détection. Pour mesurer les vitesses de relaxation à 14,1 T, les transferts de navette ont été simplement omis, de sorte que l'échantillon reste à haut-champ pendant toute l'expérience. La polarisation de l'eau peut se récupérer lors des transferts navettes et du délai de stabilisation, de sorte qu'il est nécessaire d'effectuer une suppression d'eau supplémentaire avant la détection. Pour cela, un bloc Watergate (WG) a été ajouté avant la détection. A noter que la présence d'un signal d'eau a des conséquences importantes dans le contexte de la navette : les petites vibrations de l'échantillon après la navette conduisent à une faible modulation de signal conduisant à des artefacts sur tout le spectre. Une suppression appropriée du signal fort du solvant supprime efficacement ces artefacts.



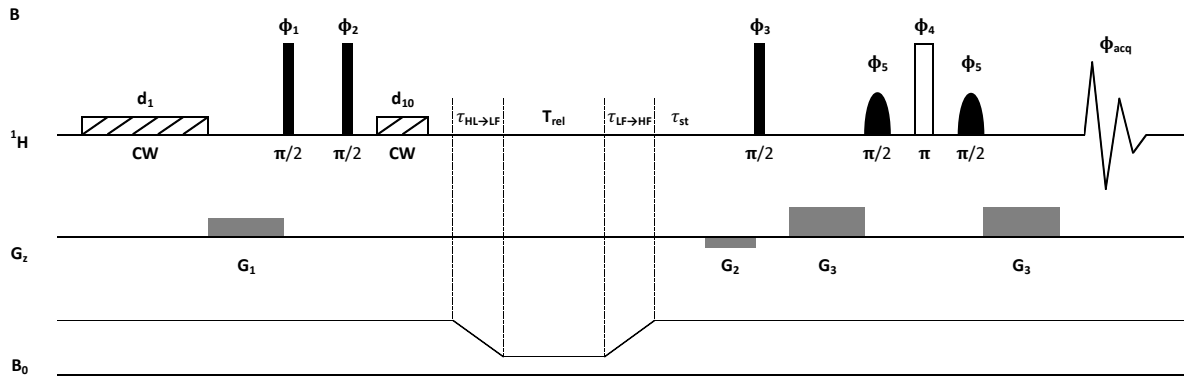


Figure 4 A. Séquence d'impulsion de l'expérience NOESY 1D standard utilisée en métabolomique. B. Séquence d'impulsion de l'expérience 1D NOESY adaptée à l'expérience de navette de relaxométrie : un délai de relaxation est inséré entre le temps de mélange et le bloc Watergate et la détection pour une meilleure suppression du signal du solvant. Cette séquence d'impulsions 1D NOESY-WG a été utilisée pour mesurer les profils de dispersion de relaxation. L'échantillon est déplacé du centre magnétique à haut-champ du spectromètre vers une position choisie à champ faible pendant  $\tau_{HL-LF}$  ( $\sim 55$ - $210$  ms) et le retour vers le haut-champ est effectué pendant  $\tau_{LF-HF}$  ( $\sim 60$ - $260$  ms). Le délai de relaxation variable à faible champ magnétique est  $T_{rel}$ , le délai de stabilisation après le retour de l'échantillon à la position de champ élevé est  $\tau_{st} = 100$  ms. Les rectangles étroits et larges représentent respectivement les impulsions proton  $\pi/2$  et  $\pi$ ; et l'impulsion  $\pi/2$  sélective de 1,5 ms pour le signal de l'eau est représentée par un demi-cercle. Les rectangles rayés représentent l'irradiation de pré-saturation à onde continue appliquée pour des durées  $d_1 = 4$  s et  $d_{10} = 10$  ms avec une faible amplitude (75 Hz). Les cycles de phase étaient:  $\phi_1 = (x, -x)$ ;  $\phi_2 = 8 (x), 8 (-x)$ ;  $\phi_3 = 2 (x), 2 (-x), 2 (y), 2 (-y)$ ;  $\phi_4 = 2 (y), 2 (-y), 2 (-x), 2 (x)$ ;  $\phi_5 = 2 (-y), 2 (y), 2 (x), 2 (-x)$ ;  $\phi_{acq} = (x, -x, -x, x, y, -y, -y, y, -x, x, x, -x, -y, y, y, -y)$ . Les gradients de champ pulsé de 1 ms  $G_1$ ,  $G_2$  et  $G_3$  sont appliqués le long de l'axe z ont pour amplitude 8,5, 4 et 20 G / cm, respectivement.

#### 1.4 L'échantillon modèle

Des études de faisabilité ont été réalisées à l'aide d'un échantillon modèle. Cet échantillon est conçu pour simuler l'interaction de petites molécules et macromolécules avec trois molécules généralement présentes dans le plasma sanguin ou le sérum, contenant 0,5 mM d'alanine, 50 mM de 3- (Triméthylsilyl) propionique-2,2,3,3-d4 (TSP), utilisé comme référence de déplacement chimique, et 0,5 mM d'albumine sérique bovine (BSA), comme substitut de l'albumine sérique humaine.

Nous utilisons l'alanine comme référence. L'alanine est un métabolite naturellement présent dans de nombreux fluides biologiques, comme le sérum sanguin, pour lequel aucune interaction avec les macromolécules n'a été détectée. Le TSP est souvent introduit comme standard interne dans les études de métabolomique. Son interaction avec les macromolécules, en particulier l'albumine, est connue. L'albumine sérique est la protéine la plus abondante dans

le sérum et le plasma sanguin. Elle est connue pour se lier à de nombreuses molécules différentes. Pour cela, nous avons utilisé l'albumine sérique bovine (bovine serum albumin ou BSA) comme macromolécule de référence (MW = 66,5 kDa), qui donne également un échantillon de viscosité similaire au plasma sanguin. Nous avons utilisé du BSA hautement purifié et sans lipides afin d'éviter les interférences ou la compétition résultant d'impuretés. L'alanine et la BSA ont été utilisées à leur concentration physiologique. Tous les composés ont été dissous dans un tampon phosphate (pH = 7,4) contenant 35 mM de  $\text{Na}_2\text{HPO}_4$ , 10% de  $\text{D}_2\text{O}$  et 0,0125% de  $\text{NaN}_3$  comme agent bactériostatique.

Le profil de dispersion de relaxation magnétique nucléaire (NMRD) d'alanine a été obtenu pour le signal de son groupe méthyle à  $\sim 1,4$  ppm (*Fig. 5*). La vitesse de relaxation longitudinale de ce groupe méthyle n'a pas de dépendance notable avec le champ magnétique. La vitesse de relaxation est proche de  $0,85 \text{ s}^{-1}$  sur toute la gamme des champs magnétiques. Cette absence de dépendance de la vitesse de relaxation longitudinale montre que la dynamique rotationnelle de l'alanine est dans le régime « extreme narrowing », en l'absence de liaison aux macromolécules comme prévu.

Au contraire, le profil NMRD du TSP démontre clairement l'interaction avec une macromolécule. Lorsque nous diminuons le champ magnétique, la vitesse de relaxation du groupe triméthylsilyle ( $-0,1$  ppm) augmente d'abord de  $1,5 \text{ s}^{-1}$  à  $14,1 \text{ T}$  à  $4 \text{ s}^{-1}$  autour de  $0,2 \text{ T}$  (*Fig. 5*). En dessous de  $0,2 \text{ T}$ , la vitesse de relaxation longitudinale augmente plus rapidement après à champ allant jusqu'à un plateau à  $0,03 \text{ T}$  et moins avec un taux  $\sim 12 \text{ s}^{-1}$ . Les deux points d'inflexion autour de  $3 \text{ T}$  et  $0,07 \text{ T}$  correspondent approximativement à des mouvements avec des temps de corrélation autour de  $1 \text{ ns}$  et de l'ordre de quelques dizaines de ns. Nous pouvons attribuer ces deux temps de corrélation aux temps de corrélation globaux du TSP et du complexe du TSP et du BSA. Une analyse quantitative est discutée en détail dans le chapitre correspondant.

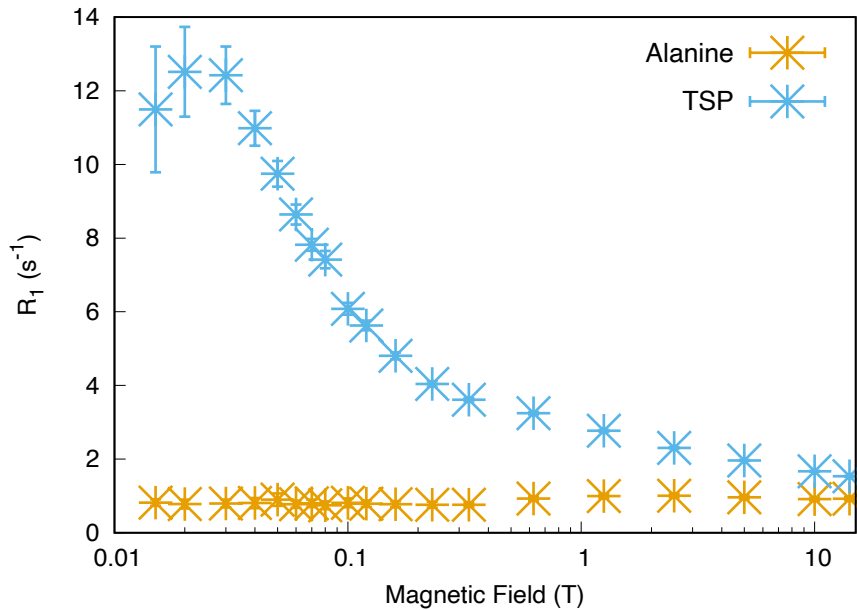


Figure 5 Profils de NMRD d'alanine (jaune) et de TSP (azur) dans l'échantillon de modèle.

### 1.5 Milieux biologiques complexes

Les fluides biologiques utilisés pour l'étude sont des sérums sanguins humains. Le sérum est une solution aqueuse obtenues à partir de sang total centrifugé. Il contient à la fois des protéines (principalement l'albumine sérique, la globuline et les lipoprotéines, le plasma contient des protéines de coagulation, absentes du sérum), des peptides, des métabolites (tels que les lipides et les acides aminés) et des électrolytes. Les expériences de relaxométrie ont été mesurées avec des sérums prélevés sur 3 individus. Une vingtaine de métabolites différents ont été attribués et analysés, dont 7 molécules ont un rapport signal / bruit relativement bon pour déterminer les vitesses de relaxation. Les données de relaxation de 4 métabolites communs sont présentées sur la Fig. 6. Les vitesses de relaxation de 3 sérums différents étaient raisonnablement proches les uns des autres.

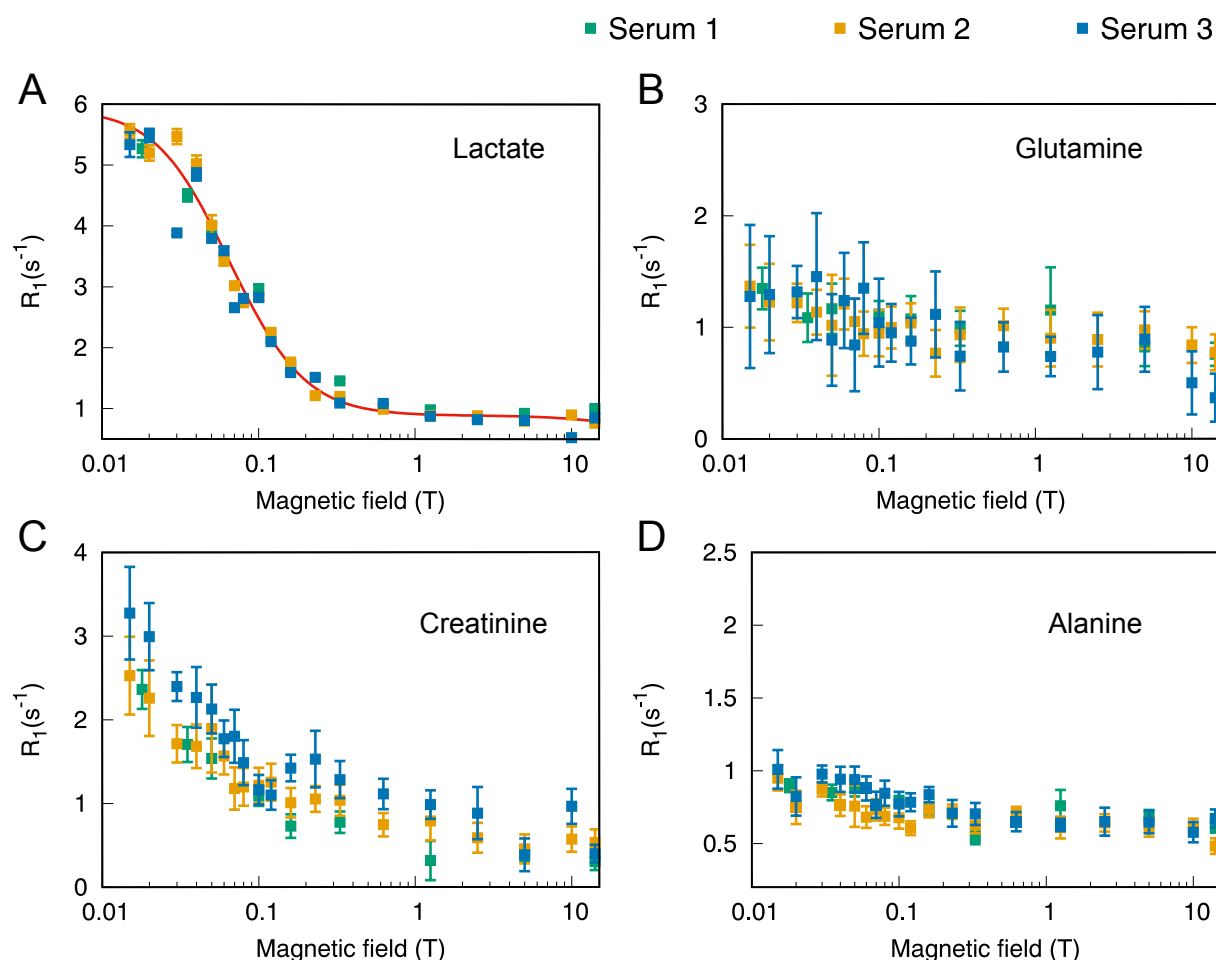


Figure 6 RMN des métabolites prélevés sur trois individus, RMN des métabolites prélevés sur trois individus, présentés avec des points de différentes couleurs. Les profils de lactate (A) et de créatinine (C) montraient une dépendance au champ magnétique, donc signes d'interactions avec les macromolécules. Les profils de la glutamine (B) et de l'alanine (D) n'ont révélé aucune liaison aux macromolécules dans la gamme de magnitude du champ magnétique explorée.

Les vitesses de relaxation mesurées des métabolites étaient très similaires pour les trois différents échantillons de sérum, ce qui reflète les compositions similaires de ces sérums chez les sujets sains, comme prévu.

La NMRD du lactate a montré une interaction entre ce dernier et les macromolécules (Fig. 6A). Nous avons pu estimer le temps de corrélation pour la diffusion rotationnelle  $\tau_c$  du complexe autour de 40 ns, ce qui correspond à l'interaction connue avec l'albumine. Le profil NMRD de la créatinine indique également des signes d'interaction (Fig. 6C) avec les macromolécules. Les vitesses de relaxation ont des grandes barres d'erreur et il ne nous a pas été possible d'atteindre la valeur limite de la vitesse de relaxation longitudinale aux champs faibles, ce qui rend l'analyse quantitative difficile. Mais cela ne nous empêche pas à conclure

que la créatinine interagit probablement avec de grandes particules dans le sérum, telles que la macroglobuline, les protéines C-réactives ou les lipoprotéines. La glutamine (*Fig. 6B*) et l'alanine (*Fig. 6D*) avaient des profils NMRD relativement plats. Ils n'ont pas montré de signe clair d'interaction.

Nous avons développé une nouvelle approche d'investigation MPI dans les milieux complexes en utilisant la relaxométrie RMN haute résolution. Cette méthode est fondée sur la mesure et l'analyse des vitesses de relaxation sur plusieurs ordres de grandeur de champ magnétique (de 0,015 à 14 T) couplés à une détection haute résolution. Cette méthode présente de nombreux avantages par rapport à celles actuellement utilisées. Tout d'abord, il n'y a pas d'étape de séparation pendant la préparation des échantillons. Une fois le sérum obtenu à partir du sang par une procédure traditionnelle, l'échantillon est directement scellé dans un tube avec un minimum de manipulation. Ensuite, la relaxométrie est capable de sonder les interactions faibles et transitoires, qui sont extrêmement difficiles à détecter par les méthodes existantes. Troisièmement, l'analyse est à la fois qualitative et quantitative. Grâce à la large gamme de champ magnétique fournie par le relaxomètre, nous sommes capables de détecter des mouvements de quelques nanosecondes à quelques centaines de nanosecondes. La taille des complexes des métabolites peut être estimée à partir de leur temps de corrélation rotationnelle. (La base de la théorie de relaxation est décrite dans le Chapitre I. Différentes approximations doivent être utilisées pour des différents systèmes de spin. Des modèles de relaxation ont été proposés, testés et comparés pour le système de la molécule de TSP en Chapitre II. Les analyses quantitatives des métabolites dans sérums sont présentées dans le Chapitre III.) Nous avons également identifié la compétition entre métabolites. Dans un environnement biologique complexe comme le sang, cette méthode est efficace pour détecter des liaisons discrètes et fournir des estimations de la taille du complexe.

Malgré une large marge d'amélioration, liée à la fois à la sensibilité et à l'étendue du champ magnétique, ce travail montre l'énorme potentiel de la relaxométrie RMN haute résolution, qui peut sûrement devenir un outil d'analyse crucial pour les mélanges de fluides complexes comme le sérum sanguin humain. Cette méthode a également été testée sur d'autres types d'échantillons, comme des interactions des médicaments avec sa molécules cible, la relaxation par des agents paramagnétiques, et détermination de la taille des particules dans des produits raffinés de pétrole.



### 3. Études du désordre intrinsèque de la protéine XLF

Les protéines sont des polymères d'acides aminés liés par des liaisons peptidiques. L'organisation dans l'espace du squelette polypeptidique et des chaînes latérales est sa conformation ou structure. La conformation d'une protéine est essentielle à sa fonction. L'étude de la relation structure-fonction dans les protéines est le fondement de la biologie structurale. Cependant, les protéines sont dynamiques et explorent un ensemble de conformations sur de nombreuses échelles de temps.

La connaissance croissante des protéines ou des régions intrinsèquement désordonnées (IDP ou IDR) montre que des régions longues de protéines sans structure tridimensionnelle stable peuvent être fonctionnelles. Les IDPs et les IDRs sont particulièrement abondants dans le protéome des organismes eucaryotes. Des prévisions à partir d'une étude du génome montrent qu'environ 40% des gènes codant pour les protéines humaines contiennent des segments désordonnés de 30 résidus ou plus.

Comparée avec la cristallographie et la microscopie cryo-électronique, la RMN a le pouvoir unique de fournir des informations sur les propriétés conformationnelles et dynamiques dans les protéines désordonnées à l'échelle atomique. Nous avons choisi de mieux comprendre la dynamique et les fonctions de la protéine XLF, qui possède un IDR C-terminal.

La protéine XLF est un des facteurs importants impliqués dans un mécanisme de réparation des cassures double brin d'ADN (double-strand break ou DSB) : la jonction d'extrémité non homologue (NHEJ). Elle est la voie de réparation des DSB sans modèle d'homologie, qui peut se produire pendant toutes les phases du cycle cellulaire. La protéine XLF a été signalée comme un site de mutation dans l'immunodéficiência humaine avec des anomalies du développement. Le complexe ligase IV-XRCC4-XLF intervient sur la dernière étape du processus de réparation NHEJ, et est en charge de l'assemblage final des extrémités DSB. XLF améliore l'affinité de la ligase IV-XRCC4 aux extrémités d'ADN non compatibles et interagit directement avec XRCC4 au niveau de leurs terminaisons N. Pourtant, une grande partie des mécanismes qui sous-tendent la fonction de XLF ne sont pas bien compris. En particulier, on a peu de connaissance sur les propriétés de son extrémité C désordonnée. La structure d'un complexe de Ku-ADN avec un peptide court composé des dix derniers résidus C-terminaux de Ku a été récemment résolue. Des études phosphomimétiques ont montré que la

phosphorylation des régions désordonnées C-terminales de XLF et XRCC4 accélère la dissociation des filaments XLF-XRCC4 de l'ADN.

Pour cette raison, nous visons à mieux comprendre les mécanismes moléculaires impliquant la région C-terminale désordonnée de XLF, et son interaction avec l'ADN. La RMN haute résolution est un outil de choix pour étudier cette région désordonnée dans le contexte de la XLF humaine entière.

XLF a été exprimé et marqué  $^{13}\text{C}$  et  $^{15}\text{N}$  en utilisant *E. coli* comme hôte. L'attribution des résonances de la chaîne principale et des chaînes latérale a été réalisée à partir d'un échantillon de la protéine purifiée et d'une série d'expériences RMN à triple résonance : HNCO, HN(CA) CO, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, CC(CO)HN et H(CCO)NH.

Les résonances de tous les résidus non-proline 232 à 293 correspondants à 91.2% de la région désordonnée de XLF ont été attribués (*Fig. 7*). Les résidus 295 à 305, y compris le site de liaison de Ku et l'étiquette poly-histidine, étaient difficiles à attribuer en raison de leurs faibles intensités dans tous les spectres. Les domaines repliés de la protéine n'ont pas été attribués : les dimères de 60 kDa sont trop grands pour être étudiés sans deutération.

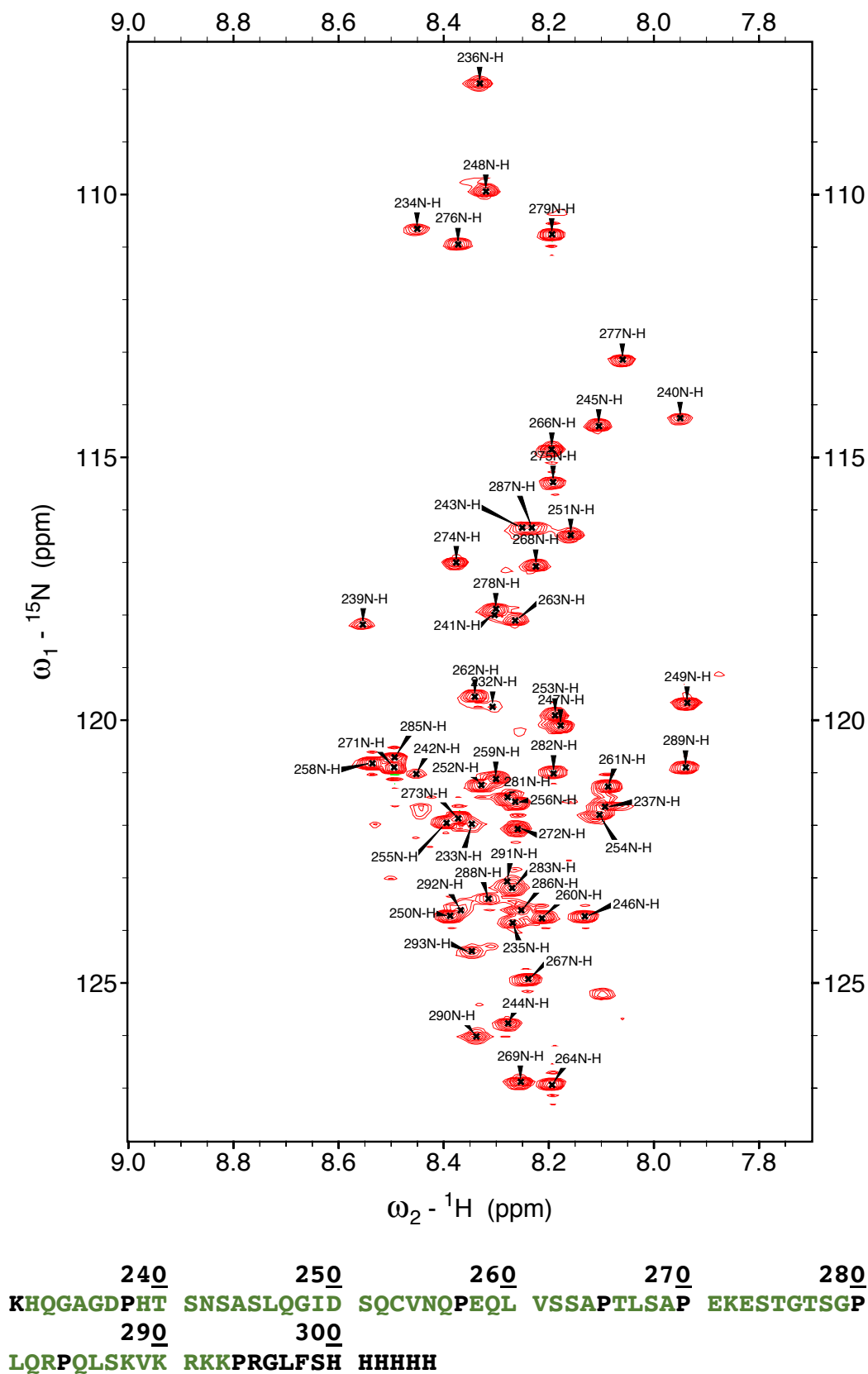


Figure 7 Spectre HSQC  $^{15}\text{N}$ - $^1\text{H}$  de XLF. Les attributions de résidus IDR sont indiqués sur le spectre. Les résidus non attribués sont affichés en gris et les résidus attribués en vert dans la séquence.

Les propensions à former des structure secondaire (SSP) ont été calculées à partir des déplacements chimiques de  $C\alpha$  et  $C\beta$ , (Fig. 8). On peut voir que l'IDR C-terminal a des préférences de conformation locales limitées. Entre les résidus 240 et 260, de courtes hélices  $\alpha$  transitoires pourraient se former avec des populations très faibles ( $\sim 10\%$ ). Cependant, il est préférable de ne pas sur interpréter des valeurs de SSP si proches de zéro.

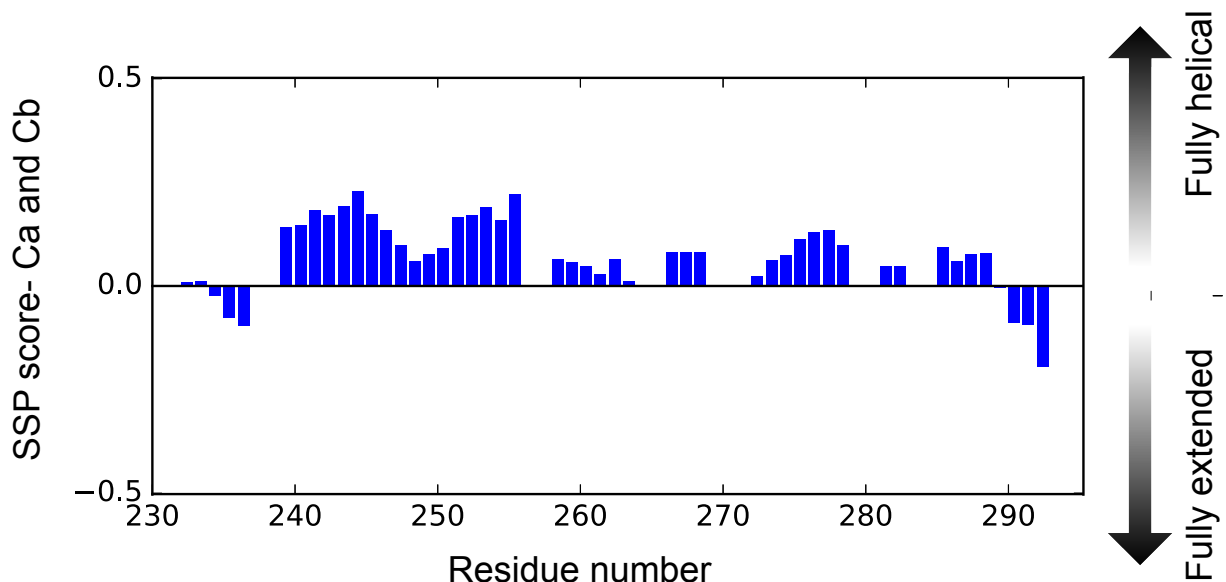


Figure 8 SSP des résidus attribués de XLF, calculés à partir des déplacements chimiques  $CA$  et  $CB$ . Les scores SSP sont 1 pour une hélice  $\alpha$  et -1 pour un feuillet  $\beta$ .

Les vitesses de relaxation longitudinales,  $R_1$  et de relaxation transversale,  $R_2$  ont été mesurés à 800 MHz avec 6 délais de relaxation chacun. Les valeurs  $R_1$  sont uniformes, variant entre 1,2 et 1,5  $s^{-1}$  pour tous les résidus (Fig. 9A). C'est souvent le cas pour les protéines désordonnées dans ce domaine de champ magnétique. Les vitesses de relaxation transversale  $R_2$  (Fig. 9B) varient beaucoup plus le long de la séquence avec des valeurs comprises entre 3 et 14  $s^{-1}$  obtenues à partir de l'expérience CPMG. Les valeurs  $R_2$  les plus faibles, observées dans la région 235-280 avec une décroissance lente indiquent que cette région est très dynamique. Le  $R_2$  des résidus C-terminaux 280-293 sont beaucoup plus élevés et similaires à ceux de la région 232-235 qui relie l'IDR au domaine structuré. Cela peut impliquer que l'extrémité C se lie de manière transitoire à la partie globulaire structurée de la protéine avec un échange rapide (sous-ms) entre les formes libres et liées. Nous ne pouvons pas exclure que cet effet soit artefactuel et provoqué par le His-tag chargé positivement. Cette relaxation transversale rapide explique probablement la difficulté d'obtenir des signaux de résidus au-delà du résidu 293 dans des expériences de triple résonance.

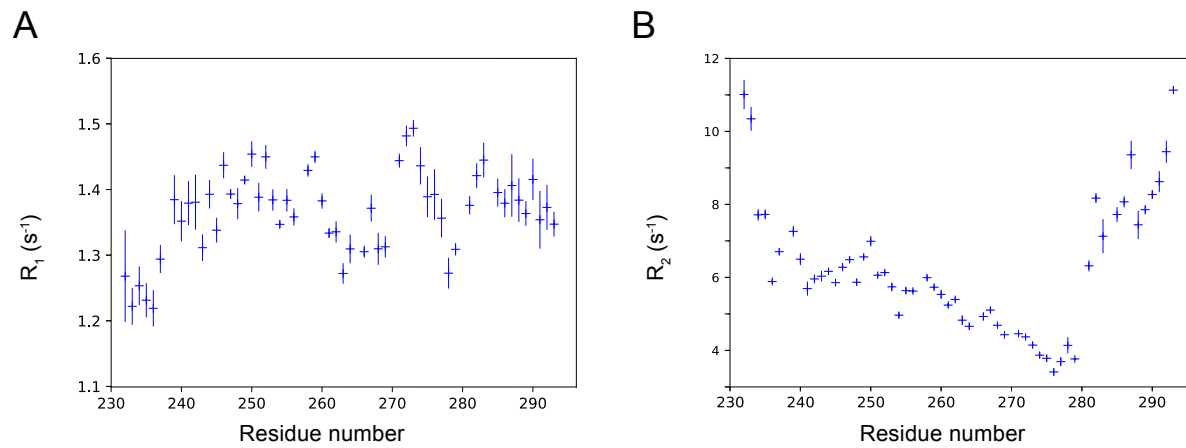


Figure 9 Profils de vitesses de relaxation longitudinales  $R_1$  (A) et vitesses de relaxation transversales  $R_2$  (B) de la région 232-293 de XLF recombinant avec le profil  $R_2$ -CPMG en croix bleue et le profil  $R_2$ -Echo en croix verte. Les délais de relaxation étaient de 0,02 s à 1,8 s pour  $R_1$  0,004 s à 0,1 s pour  $R_2$ -CPMG, et de 0,0096 s à 0,192 s pour  $R_2$ -Echo.

Ces résultats nous ont conduit à concevoir une nouvelle construction de XLF dans laquelle l'étiquette histidine est positionnée en N-terminal et clivée avant les expériences de RMN. L'interaction de cette nouvelle construction de XLF avec l'ADN a été étudiée par titrage RMN et mesure de vitesses de relaxation. L'association avec l'ADN se fait comme attendu au niveau de la région chargée positivement (résidus 288-293). Une constante de dissociation de l'ordre de la dizaine de  $\mu\text{M}$  a été déterminée, qui explique l'affinité très forte entre les filaments de XLF-XRCC4 et l'ADN. Ces résultats ouvrent la voie à une étude approfondie des mécanismes d'interaction entre XLF et l'ADN et de leur régulation.

#### 4. Marquage sélectif d'une protéine de grande taille : l'ADN polymérase $\eta$

La stabilité de l'ADN est essentielle aux organismes vivants, car il est le véhicule l'information génétique. Cependant, l'ADN est constamment exposé à des dommages dus à des facteurs externes et internes, qui entraînent des lésions différentes telles que des décalages de paires de bases et des cassures simple ou double brin. S'ils ne sont pas réparés, les dommages peuvent entraîner des mutations, l'apoptose ou des cancers. De nombreux mécanismes cellulaires permettent de réparer ou tolérer divers types de dommages à différents stades cellulaires, tel que la réparation par excision de base ou de nucléotide. Dans le cas où les lésions restent non réparées en phase S, les voies de tolérance aux dommages à l'ADN aident les cellules à survivre. Les polymérases de synthèse d'ADN translesion (TLS) sont le dernier recours pour surmonter le blocage de la réplication.

L'ADN polymérase  $\eta$  (Pol $\eta$ ) est une polymérase de translesion multi-domaine, qui appartient à la plus grande famille de toutes les polymérases TLS, la famille Y. Elle assure une réplication précise à travers diverses lésions chimiques de l'ADN telles que les dimères cyclobutyliques (CPD) induits par le rayonnement ultraviolet dans les cellules eucaryotes. Les patients ayant une Pol $\eta$  non-fonctionnelle souffrent d'une incidence accrue de cancers cutanés induits par le soleil (une forme variante de xeroderma pigmentosum).

La famille Y a une topologie de structure commune avec d'autres familles de polymérases qui ressemble à une main droite avec les domaines de la paume, du pouce et des doigts. Ces domaines sont hautement conservés dans la famille Y. L'extrémité C-terminale des polymérases de la famille Y, appelée domaine associé à la polymérase (polymerase-associated domain PAD) ou domaine du petit doigt, les distingue des autres. Ce domaine détermine également les propriétés biochimiques de chaque membre de la polymérase. Des études par transfert d'énergie de résonance de fluorescence de molécule unique (single molecule fluorescence resonance energy transfer ou smFRET) sur un autre membre dans la famille de l'ADN polymérase, polymérase I, ont montré qu'il existait un équilibre conformationnel entre une forme ouverte et une forme fermée. Cet équilibre est modifié lorsque l'ADN se lie à la protéine. Les structures de la forme apo ouverte et de la forme holo fermée (avec un modèle d'ADN) de Pol $\eta$  ont été résolues par cristallographie (*Fig. 10*). Le domaine PAD (vert) se déplace vers le domaine du pouce (jaune) lors de la liaison avec l'ADN.

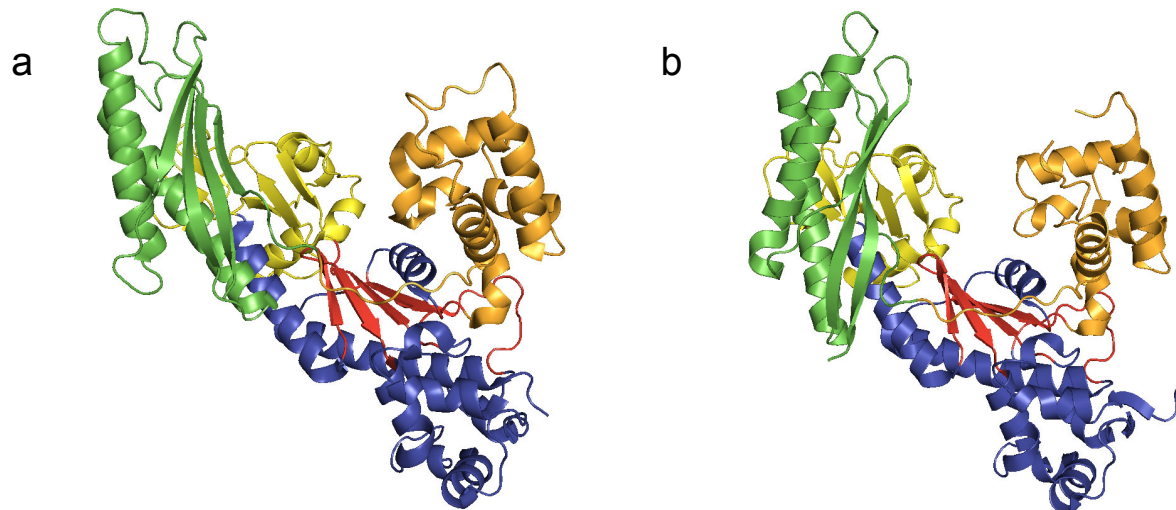


Figure 10 Structures cristallographiques des formes apo (PDB: 1JIH) (a) et holo (PDB: 3MFH) (b) de Pol $\eta$  de *S. cerevisiae*. La paume, le pouce, les doigts et le domaine PAD sont représentés respectivement en bleu, jaune orange et vert. Les feuillet beta rouges sont un sous-domaine de la paume, essentiel pour la liaison des ions métalliques. Le domaine PAD se déplace relativement aux autres domaines de la protéine d'environ 1,3 nm lorsque la matrice d'ADN est liée.

L'objectif du projet était de caractériser les mouvements internes et de domaines dans la Pol $\eta$  de *S. cerevisiae* (513 résidus, 58 kDa). Les méthodes RMN développées récemment permettent d'étudier les fluctuations des conformations et les transitions entre différentes conformations qui ont lieu sur des échelles de temps allant de la picoseconde à la seconde. Au lieu d'un marquage  $^{13}\text{C}$  uniforme, seulement un groupe méthyle sur les chaînes latérales des acides aminés porteuses de méthyle (Ala, Leu, Val, Ile, Thr, Met) est sélectivement marqué avec  $^{13}\text{CH}_3$ . Cela permet d'améliorer la qualité des spectres en diminuant des vitesses de relaxation (combinaison de la deutération et d'expériences TROSY), qui seraient sinon trop élevées dans le cas d'une molécule de grande taille. Ces groupes méthyle marqués sélectivement serviront de sondes aux mouvements de la protéine multidomaine et amélioreront la résolution spectrale pour le système de grande taille. Le marquage spécifique nécessite une production de protéines dans du  $\text{D}_2\text{O}$ . Cette première étape est essentielle au projet.

Pour obtenir Pol $\eta$  à partir de la souche d'*E. coli*, différentes combinaisons de transformation et conditions d'expression ont été testées. Deux types d'*E. coli* BL21 compétentes ont été transformées avec deux vecteurs portant la séquence codante de la protéine recombinante. Des colonies des quatre transformations ont été induites pour l'expression des protéines dans un petit volume de milieu LB. Les essais ont été réalisés dans différentes

conditions telles que les durées d'incubation, les températures et les concentrations d'IPTG. Après des sélections de colonie et de conditions, la combinaison de pGEX-6p-3 et BL21 Star (DE3) pLysS a pu produire la protéine cible avec un rendement satisfaisant de 10 mg par litre de culture. Le protocole de purification a aussi été optimiser, afin d'obtenir un échantillon de Pol $\eta$  stable sans contaminant. Un fluorimètre à balayage différentiel (DSF) a été utilisé pour améliorer la stabilité de la protéine dans différents tampons RMN. Grâce au DSF, nous avons pu stabiliser suffisamment la protéine pour que sa température de fusion ( $T_m$ ) augmente de 7,5 °C (Fig. 11) par rapport au tampon initialement utilisé.

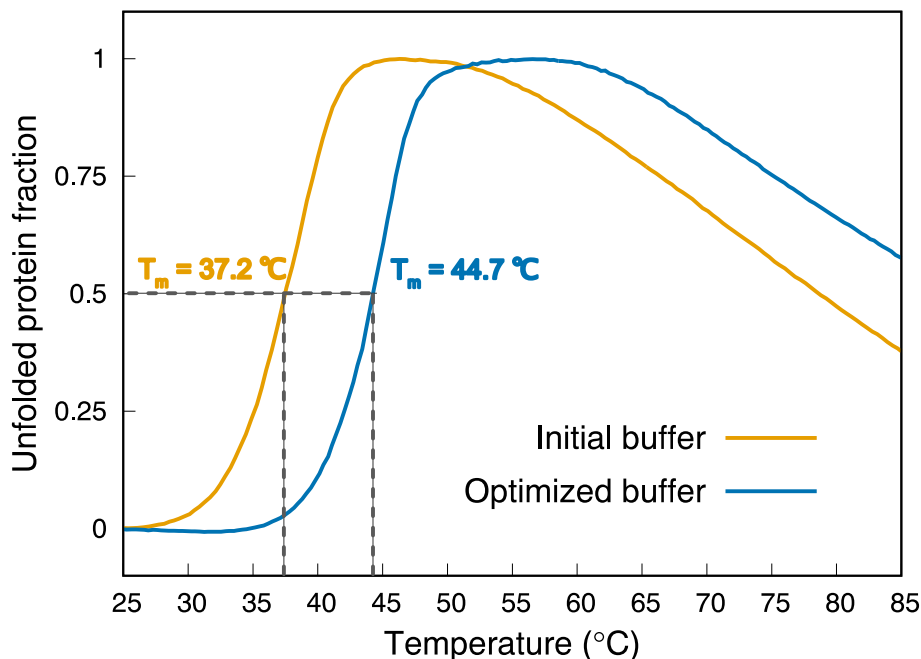


Figure 11 Courbes de fusion de Pol $\eta$  dans deux tampons différents. La  $T_m$  était de 37,2 °C dans le tampon initial et 44,7 °C dans le tampon optimisé.

Malheureusement, deux problèmes nous ont empêché d'effectuer des études conformationnelles sur la Pol $\eta$  par RMN haute résolution. Premièrement, la protéine était de très faible stabilité en présence, même faible, de D<sub>2</sub>O (2%). Deuxièmement, la souche tolérait difficilement le milieu de culture D<sub>2</sub>O et le rendement a diminué de 95%.





## Chapter I Nuclear spin relaxation and motions

A solution-state Nuclear Magnetic Resonance (NMR) spectrum depends on interactions such as the chemical shift (scaling of the interaction between the nuclear magnetic moment and the external magnetic field by electrons) or spin-spin couplings (interactions between the nuclear magnetic moment of nuclear spins mediated by electrons). The linewidth of signals comes from relaxation. Relaxation is the irreversible return of the spins to their thermal equilibrium state. A simple description of relaxation will distinguish two types of relaxation: longitudinal (historically referred to as spin-lattice relaxation) and transverse relaxation (historically referred to as spin-spin relaxation). Nuclear spin relaxation has been extensively used to gain information, both on the structure (with the Nuclear Overhauser Effect, NOE). Because the origin of relaxation lies fluctuations of interactions due to motions at the molecular scale, relaxation probes the dynamics of the systems under study. Before describing the different uses of NMR spin relaxation in biological NMR applications, we will introduce briefly the theoretical approaches and some relaxation experiments in this first chapter.

## 1 Introduction

An NMR experiment usually starts from thermal equilibrium or a state close to it, where the nuclear magnetization lies along the axis of the magnetic field. The magnetization is then perturbed using radiofrequency pulses. The evolution towards the equilibrium magnetization is called relaxation. Two time-constants are often used to quantify relaxation processes: the longitudinal relaxation time ( $T_1$ ) and the transverse relaxation time ( $T_2$ )<sup>1-4</sup>, which characterize the recovery of population distribution and the decay of coherences respectively.

The Bloch equations<sup>1</sup> were introduced to describe empirically the relaxation behavior of an isolated spin  $-1/2$  system. The Solomon equations<sup>5</sup> allow to describe more complex systems with more spins. They are indispensable to study of nuclear Overhauser effect (NOE). Nuclear spin relaxation depends on many factors, including the nature of the nuclei, the spin system and its surroundings.

Relaxation is an important tool in NMR to probe molecular motions. Indeed, molecular motions induce local fluctuations of nuclear magnetic interactions, which in turn lead to the relaxation nuclear spins. Chemical exchange, rotational diffusion, translational diffusion on surfaces, and internal molecular flexibility (in small and large molecules) lead to relaxation. These motions are functionally important for biomolecules.<sup>6-9</sup> These motions cover a broad range of time-scale from few picoseconds to seconds, and a large range of amplitudes from the very local rotation of a methyl group to domain motions associated with allostery.

The Bloch-Wangsness-Redfield (BWR) theory<sup>10-11</sup> allows us to make the link between the parameters of motions at the molecular scale and NMR accessible quantities: relaxation rates. The BWR theory relies on an efficient semi-classical approach where the nuclear spin system from a quantum mechanical point of view while the surrounding is treated classically. For a better understanding the theoretical foundation in the context of this thesis, some of the most essential features of the BWR theory will be reviewed in this chapter, which is based on books and reviews.<sup>12-16</sup>

## 2 The Bloch-Redfield-Wangsness (BRW) theory

### 2.1 The master equation

In the semi-classical theory, the Hamiltonian of the nuclear spin system  $\mathcal{H}(t)$ , is expressed as the sum of a stationary time-independent part  $\mathcal{H}_0$  and time-dependent part  $\mathcal{H}_1(t)$ :

$$\mathcal{H}(t) = \mathcal{H}_0 + \mathcal{H}_1(t). \quad [1.1]$$

The *Liouville-von Neumann* equation of the evolution of density operator  $\rho(t)$  is written as:

$$\frac{d\rho(t)}{dt} = -i[\mathcal{H}_0 + \mathcal{H}_1(t), \rho(t)]. \quad [1.2]$$

When transforming this equation from the rotating frame to the interaction frame of the Hamiltonian  $\mathcal{H}_0$ , the effect of stationary part  $\mathcal{H}_0$  is directly included in the evolution of the frame. The density operator in the interaction frame can be written as:  $\rho^I(t) = \exp(i\mathcal{H}_0 t)\rho(t)\exp(-i\mathcal{H}_0 t)$  and the Hamiltonian operator reduces to  $\mathcal{H}^I(t) = \mathcal{H}_1^I(t) = \exp(i\mathcal{H}_0 t)\mathcal{H}_1(t)\exp(-i\mathcal{H}_0 t)$ . The Liouville-von Neumann equation in the interaction frame can be written as:

$$\frac{d\rho^I(t)}{dt} = -i[\mathcal{H}_1^I(t), \rho^I(t)]. \quad [1.3]$$

The integral of equation [1.3] yields:

$$\rho^I(t) = \rho^I(0) - i \int_0^t [\mathcal{H}_1^I(t'), \rho^I(t')] dt'. \quad [1.4]$$

Substituting the expression of  $\rho^I(t)$  of equation [1.4] into [1.3] results in:

$$\frac{d\rho^I(t)}{dt} = -i[\mathcal{H}_1^I(t), \rho^I(0)] - \int_0^t [\mathcal{H}_1^I(t), [\mathcal{H}_1^I(t'), \rho^I(t')]] dt'. \quad [1.5]$$

We now define  $\tau = t - t'$ , leading to:

$$\frac{d\rho^I(t)}{dt} = -i[\mathcal{H}_1^I(t), \rho^I(0)] - \int_0^t [\mathcal{H}_1^I(t), [\mathcal{H}_1^I(t - \tau), \rho^I(t - \tau)]] d\tau, \quad [1.6]$$

This equation can be further solved by averaging over the stochastic Hamiltonians as:

$$\frac{d\rho^I(t)}{dt} = -i\overline{[\mathcal{H}_1^I(t), \rho^I(0)]} - \int_0^t \overline{[\mathcal{H}_1^I(t), [\mathcal{H}_1^I(t-\tau), \rho^I(t-\tau)]]} d\tau, \quad [1.7]$$

The ensemble average of the Hamiltonians fluctuation  $\overline{\mathcal{H}_1^I(t)}$  is zero. The first term on the right side of the equation vanishes, yielding:

$$\frac{d\rho^I(t)}{dt} = - \int_0^t \overline{[\mathcal{H}_1^I(t), [\mathcal{H}_1^I(t-\tau), \rho^I(t-\tau)]]} d\tau. \quad [1.8]$$

We can define a time  $t$  such that:

1.  $t$  is short enough that the evolution of the spin system is negligible. This requires that  $t$  is much smaller than all relaxation times (usually in the ms to s range). Under this condition,  $\rho^I(t-\tau)$  on the right side of the equation may be replaced by  $\rho^I(t)$ .
2.  $t$  is long compared to the correlation times of the motions under consideration. In liquid,  $\tau_c$  is the rotational diffusion correlation time for a molecule. It is on the order of  $10^{-12}$  to  $10^{-7}$  s depending on the size of the molecule, the viscosity of the medium and the temperature. Under this assumption, the upper limit of the integral  $t$  may be replaced by  $+\infty$ .
3. There is no correlation between  $\mathcal{H}_1^I(t)$  and  $\rho^I(t)$ . The ensemble average can be calculated separately.
4. To ensure that the spin system relaxes toward equilibrium (finite temperature), we replace  $\rho^I(t)$  by  $\rho^I(t) - \rho_0^I$ , where  $\rho_0^I$  is the density operator at equilibrium in the interaction frame.

Equation [1.8] then simplifies as:

$$\frac{d\rho^I(t)}{dt} = - \int_0^\infty \overline{[\mathcal{H}_1^I(t), [\mathcal{H}_1^I(t-\tau), \rho^I(t) - \rho_0^I]]} d\tau. \quad [1.9]$$

## 2.2 From Master equation to correlation function

Now we want to find an explicit expression of the stochastic Hamiltonian  $\mathcal{H}_1(t)$ . For this,  $\mathcal{H}_1(t)$  is decomposed as:

$$\mathcal{H}_1(t) = \sum_q (-1)^q K_{-q}(t) \mathbf{A}_q, \quad [1.10]$$

where  $K_q(t)$  is a random function of spatial variables describing the lattice;  $\mathbf{A}_q$  is a spin operator that can be expanded into basis of eigenoperators of the Hamiltonian  $\mathcal{H}_0 \sum_p \mathbf{A}_q^p$ .

In the interaction frame, we have:

$$\begin{aligned} \mathbf{A}_q^l &= \exp(i\mathcal{H}_0 t) \mathbf{A}_q(t) \exp(-i\mathcal{H}_0 t) \\ &= \sum_p \exp(i\mathcal{H}_0 t) \mathbf{A}_q^p(t) \exp(-i\mathcal{H}_0 t) \\ &= \sum_p \mathbf{A}_q^p(t) \exp(i\omega_q^p t) \end{aligned} \quad [1.11]$$

Similarly,

$$\mathbf{A}_{-q}^l = \sum_p \mathbf{A}_{-q}^p(t) \exp(-i\omega_q^p t), \quad [1.12]$$

where  $\omega_{-q}^p = -\omega_q^p$ . Substituting [1.10] and [1.11] to [1.9] yields:

$$\begin{aligned} \frac{d\rho^l(t)}{dt} &= - \sum_{q,q'} \sum_{p,p'} (-1)^{q+q'} \exp(i\omega_q^p t + i\omega_{q'}^{p'} t) [\mathbf{A}_{q'}^{p'}, [\mathbf{A}_q^p, \rho^l(t) - \rho_0^l]] \\ &\quad \times \int_0^\infty \overline{K_{-q'}(t) K_{-q}(t-\tau) \exp(-i\omega_q^p \tau)} d\tau. \end{aligned} \quad [1.13]$$

Here we define the correlation function as:

$$G_{qq'}(\tau) = \overline{K_q(t) K_{q'}(t-\tau)}. \quad [1.14]$$

It describes the time dependence of the random motion that leads to the fluctuations of the stochastic Hamiltonian.

Equation [1.13] can be simplified further under the following conditions. First, if  $q' \neq -q$ ,  $K_q(t)$  and  $K_{q'}(t)$  are statistically independent, and  $G_{qq'}(\tau) = 0$ . Thus, we only keep non-zero correlation functions with  $q' = -q$ , which we write  $G_q(\tau)$ . Second, oscillating terms average faster to zero than any evolution under relaxation. So,  $\exp(i\omega_q^p t + i\omega_{q'}^{p'} t)$  quickly averages to 0 unless  $q' = -q$  and  $p = p'$ : all terms in Equation [1.11] vanish if the eigenfrequencies are not degenerated, that is when  $p \neq p'$ . This is called the secular approximation. Now, we can write Equation [1.13] as:

$$\frac{d\rho^l(t)}{dt} = - \sum_q \sum_p [\mathbf{A}_{-q}^p, [\mathbf{A}_q^p, \rho^l(t) - \rho_0^l]] \int_0^\infty \overline{K_q(t)K_{-q}(t-\tau)} \exp(-i\omega_q^p \tau) d\tau. \quad [1.15]$$

Meanwhile, we define the power spectral density function as the real part of the integral in Equation [1.15]:

$$j_q(\omega) = RE \left\{ \int_{-\infty}^\infty G_q(\tau) \exp(-i\omega\tau) d\tau \right\} \quad [1.16]$$

and the relaxation superoperator  $\hat{\Gamma}$  gives:

$$\hat{\Gamma}(\rho^l(t) - \rho_0^l) = \frac{1}{2} \sum_q \sum_p [\mathbf{A}_{-q}^p, [\mathbf{A}_q^p, \rho^l(t) - \rho_0^l]] j_q(\omega_q^p) \quad [1.17]$$

### 2.3 From correlation function to spectral density

In isotropic liquids in the high-temperature limit, the random function of space variables  $K_q(t)$  can be written as:

$$K_q(t) = K_0(t) = c_0(t)Y_0[\Omega(t)]. \quad [1.18]$$

where  $c_0(t)$  is a function of physical constants and spatial variable (*Table I-1*),  $Y_0[\Omega(t)]$  is a modified second-order spherical harmonic function, and  $\Omega(t)$  is a set of Euler angles that

describe the orientation of the principal axis of the interaction in the laboratory frame. So, the correlation function yields:

$$G_q(\tau) = \overline{c_0(t)c_0(t+\tau)Y_0[\Omega(t)]Y_0[\Omega(t+\tau)]}. \quad [1.19]$$

For a rigid molecule undergoing rotational diffusion,

$$c_0(t) = c_0 \quad [1.20]$$

and

$$Y_0[\Omega(t)]Y_0[\Omega(t+\tau)] = \frac{1}{5} \exp(-\tau/\tau_c). \quad [1.21]$$

Table I-1 Spatial Function for Relaxation Mechanisms

Interaction	$c_0(t)$
Dipolar	$-\sqrt{6}(\mu_0/4\pi)\hbar\gamma_I\gamma_S r_{IS}(t)^{-3}$
CSA	$\Delta\sigma\gamma_I B_0/\sqrt{3}$
Quadrupolar	$e^2qQ/[4\hbar I(2I-1)]$

Therefore,

$$j(\omega) = \frac{2}{5} d_{00} \frac{\tau_c}{1 + \omega^2 \tau_c^2}, \quad [1.22]$$

where  $d_{00} = \frac{1}{6} c_0^2$ . The corresponding spectral density function is

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{1 + \omega^2 \tau_c^2}. \quad [1.23]$$

Fig. 1.1 shows series of plots of this function as a function of the angular frequency  $\omega$  for different values of  $\tau_c$ . The function is rather flat when  $\omega^2 \tau_c^2 \ll 1$ , as  $J(\omega) = 2\tau_c/5$ . This



condition defines the fast motion or extreme narrowing regime. When  $\omega^2\tau_c^2 \gg 1$ , it is the slow tumbling regime, where  $J(\omega) \propto \omega^{-2}$ . The inflection point of the Lorentzian function occurs at  $\omega^2\tau_c^2 = 1$ , i.e.  $\omega = 1/\tau_c$ . At this point,  $J(\omega) = \tau_c/5$ .

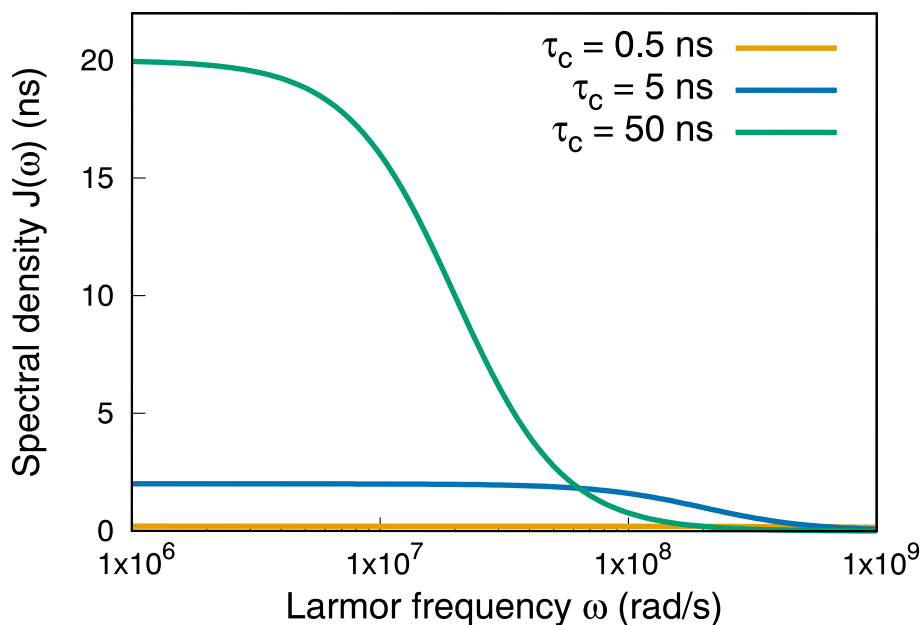


Figure 1.1 Plot of the spectral density function on a function of the Larmor frequency for different values of  $\tau_c$ .

#### 2.4 Model-free approach

In practice, most molecules are not rigid. Thus, it is important to include internal motions, with a correlation time  $\tau_i$ , in the description of molecular motions, in addition to the overall tumbling of the molecule. This is done in a simple form in the Model-Free approach<sup>17-18</sup>. The effective correlation time  $\tau_e$  is:

$$\frac{1}{\tau_e} = \frac{1}{\tau_c} + \frac{1}{\tau_i}. \quad [1.24]$$

The amplitude of internal motions is described by an order parameter ( $0 \leq S^2 \leq 1$ ). Lower order parameter indicates higher local flexibility. The spectral density is given by:

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2\tau_c}{1 + \omega^2\tau_c^2} + \frac{(1 - S^2)\tau_e}{1 + \omega^2\tau_e^2} \right] \quad [1.25]$$

An example of this approach will be discussed in the next chapter.

## 2.5 Relaxation mechanisms

Most often, for spin -1/2 nuclei, the fluctuations of the dipole-dipole and Chemical Shift Anisotropy (CSA) interaction are the most important sources of relaxation. For protons, the CSA is often small, particularly when carbon-bound hydrogen is considered, while for other nuclei (such as carbon-13 or nitrogen-15), CSA is more dominant, especially at high fields. For nuclei with a spin larger than 1/2, the electric quadrupole relaxation mechanism is the main contribution to relaxation.

In the context of this thesis, only dipolar relaxation will be considered.

Dipolar relaxation exists in systems of two spins or more. The dipole-dipole coupling is the interaction of one nuclear spin dipole with the magnetic field generated by a second nuclear spin. In liquid state, the orientation and the magnitude of the dipolar field changes with the rotational diffusion of the molecule and internal motions, leading to fluctuations of the dipole-dipole interaction (*Fig I.2*).

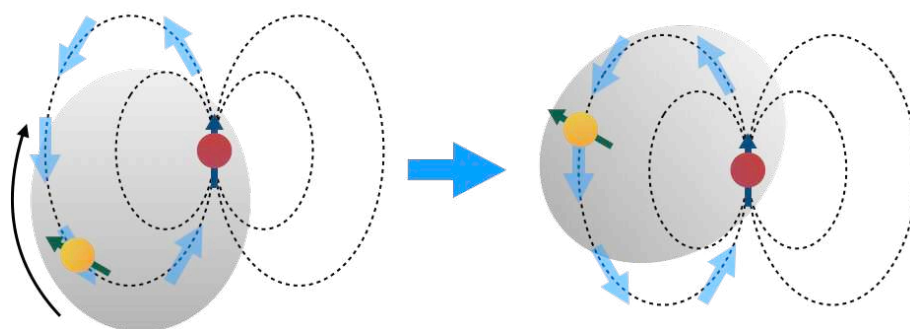


Figure I.2 Dipolar relaxation modulated by molecular tumbling

Table I-2 Tensor operators for the dipole-dipole coupling

q	p	$A_q^{2p}$	$A_{-q}^{2p} = (-1)^q A_q^{2p\dagger}$	$\omega_q^p$
0	0	$(2/\sqrt{6})I_z S_z$	$(2/\sqrt{6})I_z S_z$	0
0	-1	$(-1/2\sqrt{6})I^- S^+$	$(-1/2\sqrt{6})I^- S^+$	$\omega_S - \omega_I$
0	1	$(-1/2\sqrt{6})I^+ S^-$	$(-1/2\sqrt{6})I^+ S^-$	$\omega_I - \omega_S$
1	0	$(-1/2)I_z S^+$	$(1/2)I_z S^-$	$\omega_S$
1	1	$(-1/2)I^+ S_z$	$(1/2)I^- S_z$	$\omega_I$
2	0	$(1/2)I^+ S^+$	$(1/2)I^- S^-$	$\omega_I + \omega_S$

We consider the dipole-dipole interaction in a two-spin system  $IS$ . We assume that the scalar-coupling constant  $J_{IS} = 0$  and  $\omega_I \gg \omega_S$ . The tensor operators for the dipolar interaction are listed in *Table I-2*. We use the values of these operators  $\mathbf{A}_q^p$  in Equation [1.17] as well as the definition of the spectral density function from Equations [1.22-23] to yield all values of dipole-dipole relaxation rates (*Table I-3*).

*Table I-3 Relaxation rate constants for IS dipolar interaction.*

Coherence level	Operator	Relaxation rate constant
Populations	$I_z$	$(d_{00}/4)[J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)]$
	$S_z$	$(d_{00}/4)[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)]$
	$I_z \leftrightarrow S_z$	$(d_{00}/4)[-J(\omega_I - \omega_S) + 6J(\omega_I + \omega_S)]$
0	$2I_z S_z$	$(3d_{00}/4)[J(\omega_I) + J(\omega_S)]$
	$ZQ_x, ZQ_y$	$(d_{00}/8)[2J(\omega_I - \omega_S) + 3J(\omega_I) + 3J(\omega_S)]$
	$I^+, I^-$	$(d_{00}/8)[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_S) + 6J(\omega_I + \omega_S)]$
$\pm 1$	$S^+, S^-$	$(d_{00}/8)[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)]$
	$2I^+ S_z, 2I^- S_z$	$(d_{00}/8)[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)]$
	$2I_z S^+, 2I_z S^-$	$(d_{00}/8)[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)]$
$\pm 2$	$DQ_x, DQ_y$	$(d_{00}/8)[3J(\omega_I) + 3J(\omega_S) + 12J(\omega_I + \omega_S)]$

## 2.6 Relaxation under strong scalar coupling

In this work, we will consider low magnetic fields, where strong scalar couplings are encountered more frequently than at high magnetic fields. The effect of strong scalar couplings on the relaxation behavior of systems of spins has been studied theoretically<sup>17</sup> and experimentally<sup>18</sup> in the groups of H. M. Vieth, K. Ivanov and A. Yurkovskaya,.

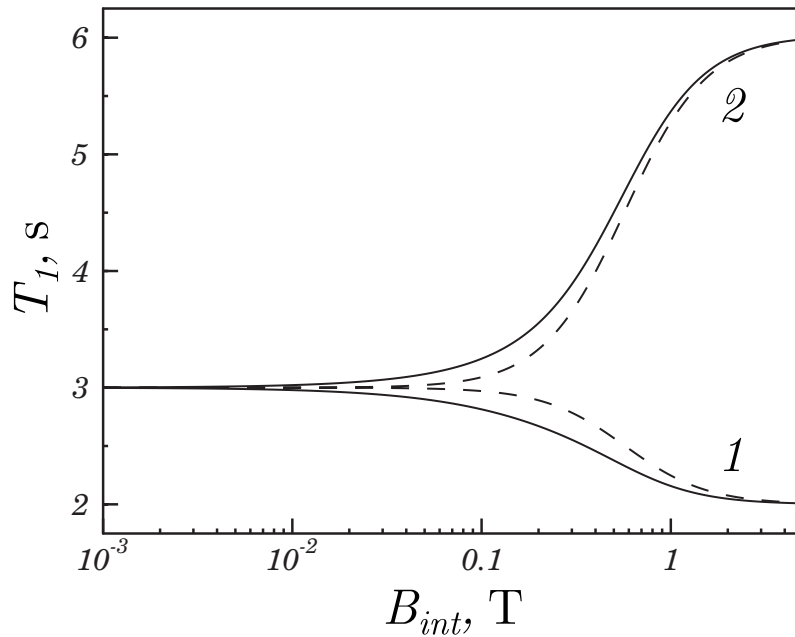
When the external field is low enough, scalar coupling interaction  $\pi J_{IS}$  becomes comparable with or larger than the difference of the resonance frequencies of the two spins ( $\omega_I - \omega_S$ ). In the limit of a strong coupling when the scalar coupling is much larger than the difference in resonance frequencies, the coupled spins share a common effective relaxation rate, despite their different intrinsic relaxation rate at high field. For a two-spin system, the common

longitudinal relaxation rate  $R_{common}$  is the average of the individual relaxation rates of two isolated spins:

$$R_{common} = \frac{1}{2}(R_I + R_S). \quad [1.26]$$

Another condition for this phenomenon to be observed, is that  $J_{IS} \geq R_I$  (or  $R_S$ ). Otherwise the spins follow the intrinsic decay faster than the mixing of longitudinal polarization by the scalar interaction.

An example is presented in *Fig. I.3*<sup>17</sup> for a two-spin system, where  $J_{IS}=15$  Hz,  $\omega_I - \omega_S=1$  ppm. The calculated NMR relaxation dispersions (NMRD) profiles of spin I and S are shown by curves 1 and 2, respectively. The spin-lattice relaxation time of individual spins is considered field-independent. At high field, two different relaxation rates can be obtained: are  $1/2 \text{ s}^{-1}$  and  $1/6 \text{ s}^{-1}$ . At low field, the common relaxation time of the system is  $1/3 \text{ s}^{-1}$ .



*Figure I.3* NMRD profiles of a coupled two-spin system when the field switch  $B_{int} \rightarrow B_0$  is either slow (solid lines) or fast (dashed lines).<sup>17</sup>

## 2.7 Relaxation in a like-spin system

Until here, relaxation theory was demonstrated for systems consisting by two spins of different frequencies. Here, we present briefly the case of like spin system, particularly the differences with the unlike spin systems presented above.

To simplify equation [1.13],  $\exp(i\omega_q^p t - i\omega_q^{p'} t) = 1$ , the assumption  $p = p'$  is too restrictive in like-spin systems, as it does not consider the fact that resonance frequencies are identical in like spin systems (*e.g.* zero-quantum operator have associated 0 frequency), which makes new terms in the relaxation matrix secular. Therefore, when  $\omega_q^p t = \omega_q^{p'} t$  is applied, the relaxation superoperator  $\hat{\Gamma}$  becomes:

$$\hat{\Gamma}(\rho(t) - \rho_0) = \frac{1}{2} \sum_q \sum_{p,p'} [A_{-q}^{p'} [A_q^p, \rho(t) - \rho_0]] j_q(\omega_q^p) \quad [1.27]$$

In biological molecules the most common like-spin systems, are  $A_2$  and  $A_3$  systems, such as protons in  $-CH_2$  and  $-CH_3$  groups. The relaxation rate constants<sup>5, 19</sup> are listed in *Table I-4*.

*Table I-4 Relaxation rate constants for equivalent spin system*

Spin system	Relaxation rate constant
$A_2$	$\frac{3}{64} \frac{\mu_0^2 \gamma^4 \hbar^2}{\pi^2 r^6} [J(\omega_H) + 4J(2\omega_H)]$
$A_3$	$\frac{3}{32} \frac{\mu_0^2 \gamma^4 \hbar^2}{\pi^2 r^6} [J(\omega_H) + 4J(2\omega_H)]$

## 3 Chemical exchange

Now we turn the discussion to the chemical exchange and its effect on the evolution of magnetization. In NMR, the dynamic equilibrium of a nucleus between two or more chemical environments is called chemical exchange. It is due to either chemical reactions or conformational fluctuations. The chemical exchange influences the chemical shifts and pic width of the nucleus in the spectrum. By monitoring these features, we can study chemical exchange process.

## 3.1 Kinetics of chemical exchange

We consider chemical exchange between two isomeric sites A and B:



where  $k_1$  and  $k_{-1}$  are the first order rate constant of forward and reverse reaction. The exchange rate constant  $k_{ex}$  is

$$k_{ex} = k_1 + k_{-1}. \quad [1.29]$$

The average exchange time  $\tau_{ex}$  is

$$\tau_{ex} = \frac{1}{k_{ex}}. \quad [1.30]$$

The concentrations evolution of A and B can be described by rate constants as

$$\begin{cases} \frac{d[A]}{dt} = -k_1[A] + k_{-1}[B] \\ \frac{d[B]}{dt} = k_1[A] - k_{-1}[B] \end{cases}, \quad [1.31]$$

or in matrix form as

$$\frac{d}{dt} \begin{bmatrix} [A](t) \\ [B](t) \end{bmatrix} = \begin{bmatrix} -k_1 & k_{-1} \\ k_1 & -k_{-1} \end{bmatrix} \begin{bmatrix} [A](t) \\ [B](t) \end{bmatrix}. \quad [1.32]$$

At equilibrium,

$$\frac{d[A]_{eq}}{dt} = \frac{d[B]_{eq}}{dt} = 0. \quad [1.33]$$

So, we can get the concentration ratio of A and B

$$\frac{[A]_{eq}}{[B]_{eq}} = \frac{k_{-1}}{k_1}. \quad [1.34]$$

The population fractions of A and B are respectively

$$p_A = \frac{k_{-1}}{k_1 + k_{-1}}, \quad [1.35]$$

and

$$p_B = 1 - p_A = \frac{k_1}{k_1 + k_{-1}}. \quad [1.36]$$

### 3.2 Bloch-McConnell equation

The evolution of magnetization of expressed by the Bloch equation is

$$\frac{d}{dt} \begin{bmatrix} E \\ M_x \\ M_y \\ M_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -R_2 & -\omega & 0 \\ 0 & \omega & -R_2 & 0 \\ R_1 M_{eq} & 0 & 0 & -R_1 \end{bmatrix} \begin{bmatrix} E \\ M_x \\ M_y \\ M_z \end{bmatrix}, \quad [1.37]$$

where  $E$  is the identity operator,  $R_1$  and  $R_2$  are longitudinal and transverse relaxation rates and  $M_{eq}$  is the equilibrium magnetization along  $z$  direction. If we include chemical exchange with kinetics described as in equation [1.32], the effect of chemical exchange is called Bloch-McConnell equation<sup>20</sup>, which can be written under the following form:

$$\frac{d}{dt} \begin{bmatrix} E \\ M_x^A \\ M_y^A \\ M_z^A \\ M_x^B \\ M_y^B \\ M_z^B \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_{2A} - k_1 & -\omega_A & 0 & k_{-1} & 0 & 0 & 0 \\ 0 & \omega_A & -R_{2A} - k_1 & 0 & 0 & k_{-1} & 0 & 0 \\ R_{1A} M_{eqA} & 0 & 0 & -R_{1A} - k_1 & 0 & 0 & 0 & k_{-1} \\ 0 & k_1 & 0 & 0 & -R_{2B} - k_{-1} & -\omega_B & 0 & 0 \\ 0 & 0 & k_1 & 0 & \omega_B & -R_{2B} - k_{-1} & 0 & 0 \\ R_{1A} M_{eqA} & 0 & 0 & k_1 & 0 & 0 & -R_{1B} - k_{-1} & 0 \end{bmatrix} \begin{bmatrix} E \\ M_x^A \\ M_y^A \\ M_z^A \\ M_x^B \\ M_y^B \\ M_z^B \end{bmatrix}. \quad [1.38]$$

This equation allows to calculate the trajectory of the magnetization for a one-spin system and that it can easily be generalized to include the presence of a radiofrequency field, thereby

offering an easy way to simulate pulse sequences in presence of chemical exchange. The observed component during an experiment is the transverse magnetization after a  $\pi/2$  pulse  $\mathbf{M}^+ = \mathbf{M}_x + i\mathbf{M}_y$ . The Bloch-McConnell equation can be simplified to:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{M}_A^+ \\ \mathbf{M}_B^+ \end{bmatrix} = \begin{bmatrix} i\omega_A - R_{2A} - k_1 & k_{-1} \\ k_1 & i\omega_B - R_{2B} - k_{-1} \end{bmatrix} \begin{bmatrix} \mathbf{M}_A^+ \\ \mathbf{M}_B^+ \end{bmatrix}. \quad [1.39]$$

The solutions to the equation are given by:

$$\begin{bmatrix} \mathbf{M}_A^+(t) \\ \mathbf{M}_B^+(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \mathbf{M}_A^+(0) \\ \mathbf{M}_B^+(0) \end{bmatrix}, \quad [1.40]$$

where

$$\begin{aligned} a_{11}(t) &= \frac{1}{2} \left[ \left( 1 - \frac{i\Delta\omega - \Delta R_2 + k_{ex}(p_A - p_B)}{\lambda_+ - \lambda_-} \right) e^{-\lambda_- t} \right. \\ &\quad \left. + \left( 1 + \frac{i\Delta\omega - \Delta R_2 + k_{ex}(p_A - p_B)}{\lambda_+ - \lambda_-} \right) e^{-\lambda_+ t} \right] \\ a_{22}(t) &= \frac{1}{2} \left[ \left( 1 + \frac{i\Delta\omega - \Delta R_2 + k_{ex}(p_A - p_B)}{\lambda_+ - \lambda_-} \right) e^{-\lambda_- t} \right. \\ &\quad \left. - \left( 1 - \frac{i\Delta\omega - \Delta R_2 + k_{ex}(p_A - p_B)}{\lambda_+ - \lambda_-} \right) e^{-\lambda_+ t} \right] \\ a_{12}(t) &= \frac{k_{ex}p_A}{\lambda_+ - \lambda_-} (e^{-\lambda_- t} - e^{-\lambda_+ t}) \\ a_{21}(t) &= \frac{k_{ex}p_B}{\lambda_+ - \lambda_-} (e^{-\lambda_- t} - e^{-\lambda_+ t}) \end{aligned} \quad [1.41]$$

with  $\Delta R_2 = R_{2B} - R_{2A}$ ,  $\Delta\omega = \omega_B - \omega_A$  and

$$\lambda_{\pm} = \frac{1}{2}(-i\omega_B - i\omega_A + R_{2B} + R_{2A} + k_{ex}) \pm \frac{1}{2}[(i\Delta\omega - \Delta R_2 + k_1 - k_{-1})^2 + 4k_1k_{-1}]^{\frac{1}{2}}.$$

The regime of chemical exchange effects the shape of NMR spectrum.

In the case of fast exchange,  $k_{ex} \gg |i\Delta\omega + \Delta R_2|$ , equation [1.40] simplifies to:

$$\mathbf{M}^+ = \mathbf{M}^+(0) e^{(i\bar{\omega} - \bar{R}_2 - p_A p_B \Delta\omega^2 / k_{ex})t} \quad [1.42]$$



where  $\overline{R_2} = p_A R_2^A + p_B R_2^B$  and  $\overline{\omega} = p_A \omega_A + p_B \omega_B$ . A single line at the population-averaged position,  $\overline{\omega}$ , is observed in the spectrum with a linewidth corresponding to the population-averaged transverse relaxation,  $\overline{R_2}$ , plus an exchange contribution equal to  $p_A p_B \Delta\omega^2 / k_{ex}$ . Note that, as  $\Delta\omega$  is proportional to the static magnetic field,  $B_0$ , the exchange contribution to the transverse relaxation rate increases quadratically with  $B_0$ .

In the case of slow exchange,  $k_{ex} \ll |i\Delta\omega + \Delta R_2|$ , the off-diagonal terms are negligible compared to diagonal terms. Equation [1.40] can be then reduced to:

$$\begin{bmatrix} \mathbf{M}_A^+(t) \\ \mathbf{M}_B^+(t) \end{bmatrix} = \begin{bmatrix} -R_{2A} - k_1 & 0 \\ 0 & i\Delta\omega - R_{2B} - k_{-1} \end{bmatrix} \begin{bmatrix} \mathbf{M}_A^+(0) \\ \mathbf{M}_B^+(0) \end{bmatrix}. \quad [1.43]$$

The spectrum consists of two peaks arising separately from A and B forms at position  $\omega_A$  and  $\omega_B$ . The peak widths are determined by  $R_{2A} + k_1$  and  $R_{2B} + k_{-1}$ . In the case where one form is much more populated than the other one,  $p_A \gg p_B$  and  $k_{-1} \gg k_1$ , the exchange contribution to the linewidth of **B** is much larger than the one to **A**. For example, assuming  $k_{ex} = 1000 \text{ s}^{-1}$  and  $p_B = 2\%$ , we have  $k_{-1} = 20 \text{ s}^{-1}$  and  $k_1 = 980 \text{ s}^{-1}$ , so that peak **B** is too broad to be detected in the NMR spectra.

### 3.3 Experiments to quantify chemical exchange

Here, we present two NMR approaches for characterizing chemical exchange process, provided that exchange lies in the intermediate regime and one of the exchanging states is visible in the spectrum.

One such approach, the Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment<sup>21-24</sup>, consists in measuring the effective transverse spin relaxation rate,  $R_{2\text{eff}}$ , over a series of experiments where a variable number of refocusing pulses is applied to the magnetization in the transverse plane during a fixed period of time,  $T_{\text{CPMG}}$  (*Fig I. 4A*). The chemical shift fluctuations arising from the stochastic interconversion of the nuclear spin between states with different chemical shifts induce magnetization dephasing, that results in an increase of  $R_{2\text{eff}}$ . When refocusing  $\pi$  pulses are applied, the magnetization dephasing is reduced. The more frequently refocusing pulses are applied, the more the deviations due to chemical exchange are compensated and the lower  $R_{2\text{eff}}$  becomes (*Fig I.4 B*). The resulting relaxation dispersion profile of  $R_{2\text{eff}}$  values plotted as a function of pulsing frequency,  $\nu_{\text{CPMG}} = 1/4\tau_{\text{CPMG}}$ , where  $2 \tau_{\text{CPMG}}$  is the delay between two refocusing pulses, can then be analyzed to obtain

information about the kinetic and the thermodynamic of the exchange process as well as site-specific chemical shifts of the exchanging forms. The CPMG relaxation dispersion experiment are most effective on systems exchanging on the micro to millisecond timescale.

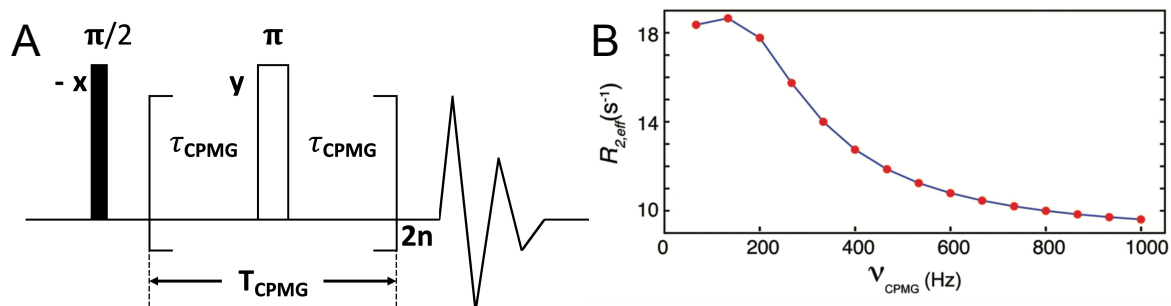


Figure 1.4 A Basic CPMG experiment sequence. Narrow and wide rectangles represent  $\pi/2$  and  $\pi$  proton pulses, respectively. The refocusing frequency is one fourth of the refocusing delay  $\nu_{\text{CPMG}} = 1/4\tau_{\text{CPMG}}$ . The relaxation delays,  $T_{\text{cpmg}} = \frac{n}{\nu_{\text{CPMG}}}$ , remain constant for all refocusing frequency. B Typical relaxation dispersion curve of CPMG experiment. Figure B is adapted from reference.<sup>25</sup>

Chemical exchange saturation transfer (CEST) is another method that allows to quantify chemical exchange involving sparsely populated, “invisible” states. The method was developed more than fifty years ago to characterize proton exchange between two chemical compounds<sup>26</sup> and applied to different biological systems thereafter<sup>25, 27-31</sup>. The CEST approach consists in recording a series of experiments in which a weak radio-frequency (rf) irradiation is applied at different positions of the spectrum while monitoring the intensity of the peaks associated to the major, “visible” state (I). When the carrier is positioned near a major state resonance, a huge drop of intensity of the associated signal is observed due to spin saturation. Interestingly, when the rf field is applied at the position of the minor, “invisible” resonances, the perturbation induced by the rf field is transferred to the major state magnetization via the chemical exchange process, which also decreases the intensity of the major state peak. The intensities are normalized by the intensity from an experiment recorded without the CEST block ( $I_0$ ). Tracing the intensity of the main form peak against the position of rf irradiation allow to localize and visualize the signal of the forms that are little populated and invisible in a normal spectrum. In addition, analysis of the resulting profile using the Bloch-McConnell equation provides information about the exchange rate between the states as well as the fractional population of the different states.

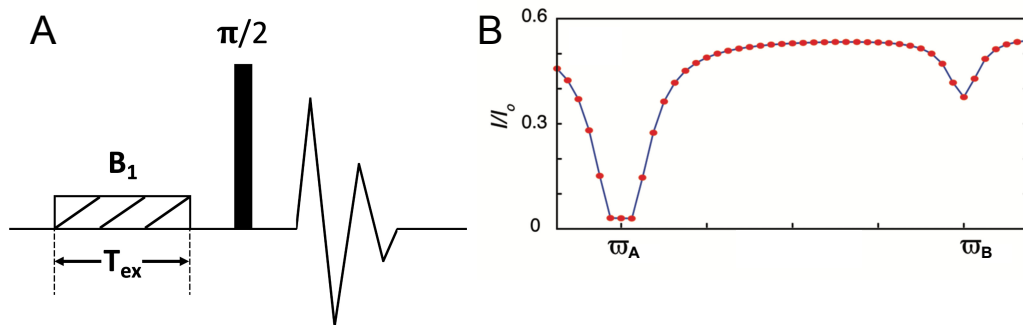


Figure 1.4 A Schematic of CEST experiment. A weak radio frequency irradiation  $B_1$  is applied to saturate certain region of the spectrum. B Intensity profile are obtained by quantifying the intensity of the visible-state peak as a function of the weak  $B_1$ . The duration of the weak  $B_1$  is relaxation delay,  $T_{ex}$ . Intensities are normalized by  $I_0$ , the intensity from an experiment recorded without the CEST block. In case of chemical exchange, this saturation will be transferred to signals of other forms of the molecule. Figure B is adapted from reference<sup>25</sup>.

## Chapter II Binding of small molecules to macromolecules probed by high-resolution relaxometry

The power of NMR is undisputed to analyze complex mixtures. Advances in pulse sequence development and new instrumentation keep pushing further the limitations of NMR. In this chapter, we present the development of a new method to investigate molecular interactions, especially weak interactions, by high-resolution NMR relaxometry. This method will be applied to complex mixtures in the following chapter.

The principle of relaxometry is to measure relaxation rates over several orders of magnitude of magnetic fields. High-resolution relaxometry provides the measurement of these relaxation rates for each individual signal in a high-resolution spectrum. Thus, each molecule will show a different relaxometry signature due to its correlation time for tumbling. When the components of a mixture that tumble at very different timescales interact, high-resolution relaxometry can, in principle identify this interaction. Here, we describe the development of the method on a model sample. Thanks to the relaxation dispersion obtained, we were able to interpret interactions between small molecules and macromolecules qualitatively and quantitatively.

## 1 Introduction

Understanding ligand binding is important to understand protein function. In drug design, quantifying the binding of a large number of molecules with the target protein is a major step. NMR is an attractive technique to detect and characterize these interactions, for its capacity of working with biological samples in solution and the absence of large probe (*e.g.* fluorophore).

To observe interactions between small molecules (*e.g.* ligand) and macromolecules (*e.g.* protein), we must quantify a parameter that differs significantly between the free and bound forms. Such observation may be made from a small molecule or a macromolecule. Structure-activity relationships (SAR) experiment<sup>32</sup> is a classic receptor-based method. It monitors chemical shift perturbation of the protein by <sup>1</sup>H-<sup>15</sup>N HSQC. This approach offers a residue-specific interaction mapping. However, it is limited to relatively small size receptor and needs highly purified and concentrated sample (several milligrams) and resonance assignments of the protein target, which is not necessarily well-behaved for NMR. For larger systems (>40 kDa), perdeuteration, <sup>13</sup>C(methyl)-labeled sample and transverse relaxation-optimized spectroscopy (TROSY)-HSQC experiment, allow to increase the sensitivity.<sup>33-38</sup> However, the sample preparation process is challenging and are not suitable for investigation of biological fluids. Nowadays, ligand-based methods are more popular in practice, for the reason that there is no requirement of isotope labelling and few restrictions on the size of the macromolecule.<sup>39-42</sup> These methods are based on the transfer of different physical properties from the macromolecules to the binding small molecule via the complex.

## 2 NMR experiments to investigate interactions of small molecules with macromolecules

### 2.1 Diffusion-edited experiment

Diffusion-edited NMR experiments are often referred as affinity NMR in drug screening. By implementing diffusion filters, it is possible to distinguish fast diffusing molecules (free small molecules) from slow diffusing molecules (macromolecules and the small molecules interacting with macromolecules). This method is based on significant difference of translational diffusion coefficients (*D*) between the ligand and the macromolecules.<sup>43-44</sup> In drug screening, it is not always necessary to make qualitative measurements of diffusion coefficient. The experiments can be carried out in 1D or 2D, depending on the context of the study.

To introduce the principle of diffusion measurements by NMR, we describe one of the simplest diffusion experiments: the stimulated-echo (STE), illustrated in Fig. II. 1. A first pulsed field gradient is applied to encode the magnetization spatially. Then, molecules are free to move during a time delay ( $\Delta$ ) records. A couple of  $\pi/2$  pulses instead of a  $\pi$  pulse allow the magnetization to have longitudinal relaxation during the diffusion encoding, which avoids the fast transverse relaxation. A second pulsed field gradient decodes the spins before the detection: it is the spin echo. Ideally, in the absence of motion, all magnetization is refocused as the magnetic field will be the same for each molecule during the encoding and decoding gradients. Molecular motions during the delay  $\Delta$  may lead to a change of the position-dependent magnetic field during the two gradient pulses. This will reduce the amplitude of the echo.

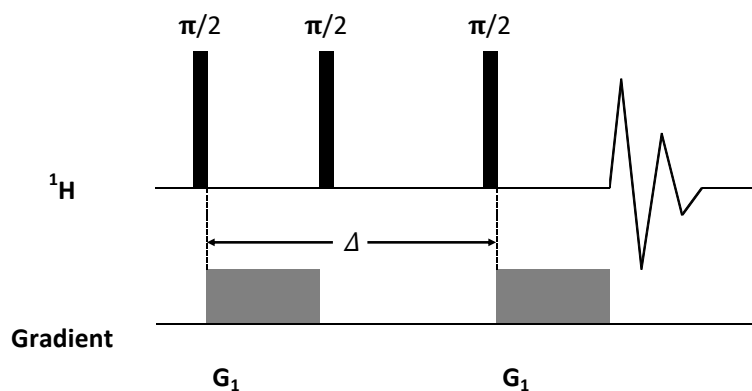


Figure II.1 The STE diffusion experiment. The black squares on the proton channel are  $\pi/2$  pulses, and that on the gradient channel are PFGs. The delay  $\Delta$  between two PFGs is the diffusion encoding time.

Diffusion-edited NMR experiments are popular tools for probing interactions. Yet, many complications may interfere with the analysis.

First, the problem arises from the signals of the protein, which are always present in spectra. The signal of a large protein may be larger than the signal of small molecules, which makes it difficult to identify the signals of the small molecules. It is possible to use an isotope filter to simplify the spectrum, under the condition that the protein needs to be isotopically labeled.<sup>45</sup>

Second, translational diffusion is not very sensitive to the size of molecule. Quantitative relations between diffusion and molecular dimension were described by the Stokes-Einstein-Debye relation:

$$D = \frac{k_B T}{6\pi\eta r_h} \quad [2.1]$$

and

$$\tau = \frac{4\pi\eta r_g^3}{3k_B T}. \quad [2.2]$$

$D$  is the translational diffusion coefficient, and  $\tau$  is the rotational diffusion coefficient, often called overall tumbling time or rotational diffusion time  $\tau$  in NMR.  $k_B T$  is the thermal energy,  $\eta$  is the viscosity of the medium, and  $r_h$  and  $r_g$  are the hydrodynamic radius and the radius of gyration of the Brownian particle, respectively. Here, we assume that  $r_h$  and  $r_g$  are approximately the same. The observed diffusion is translational diffusion, proportional to  $r^{-1}$ . Compare  $\tau$ , which is to the order -3 of the radius,  $D$  is not as sensitive to the molecular dimension.

The more restricting factor comes from the averaging of observables under fast exchange of the small molecule between its free and bound form. The expected value of the diffusion coefficient ( $D_{obs}$ ) is the population averaged value of the bound ( $D_{bound}$ ) and free form ( $D_{free}$ ):

$$D_{obs} = pop_{bound} D_{bound} + pop_{free} D_{free}, \quad [2.3]$$

where  $pop_{bound}$  and  $pop_{free} (= 1 - pop_{bound})$  are population fractions of the small molecule in its bound and free forms, respectively. If the free form fraction is much larger than that of the bound form, the observed average diffusion coefficient is then not very different from the diffusion coefficient of the free form. Signals of the small molecule are inseparable in the diffusion dimension. *Fig. II. 2* gives a graphic presentation of quantitative comparison of  $D$  and  $\tau$  on function of binding population. The change of the average  $\tau$  with the increase of the bound population is much steeper than the change of the average  $D$ . In other cases, where  $pop_{bound}$  is large enough to distinguish  $D_{bound}$  and  $D_{free}$ , the line broadening may deteriorate the spectra.

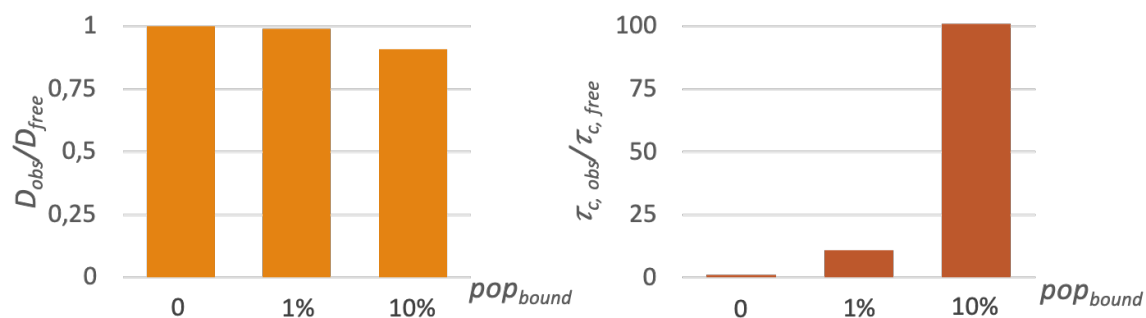


Figure II.2 Comparison of observation values of translational diffusion coefficient and rotational correlation time on function of binding population. The complex is considered to be ten times larger than the free form small molecule. All values are normalized by the values of free molecules.

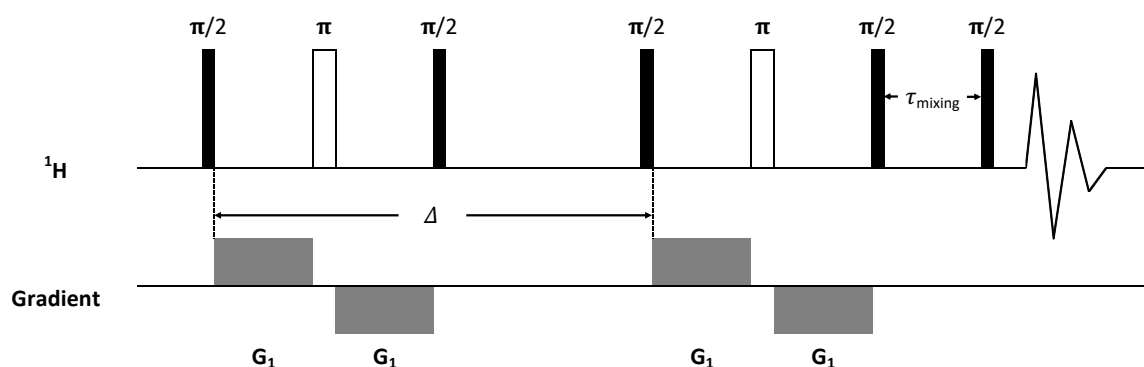
## 2.2 Polarization and saturation transfer experiments based on nuclear Overhauser effects

The nuclear Overhauser effect (NOE) is the transfer of polarization from one spin to another through space via relaxation due to the auto-correlation of a dipole-dipole interaction. The NOE is employed by many experiments to detect binding.<sup>46-47</sup> Cross-relaxation due to dipole-dipole interactions has a strong distance dependence (the rate is proportional to  $r_{IS}^{-6}$ , where  $r_{IS}$  is the internuclear distance. When a ligand is bound to its receptor, short internuclear distances make such polarization transfer possible. This allows one to distinguish a binding molecule from the non-binding ones. Depending on the timescale of molecular motions, the NOE cross-relaxation rate can be either negative or positive. When the molecular motions are much slower than the Larmor frequency, the probability of zero quantum transition is dominant, the intensity of the resonance peaks decreases due to cross relaxation. This is a negative NOEs. When the molecular motions are faster than the Larmor frequency, double quantum transition is dominant. The resonance peak is pumped up by cross and lead to positive NOEs. The correlation time of overall rotational diffusion for biological macromolecules is large enough to be in the slow-motion regime at high magnetic field. This leads to negative NOEs in complexes between receptors and ligands. On the contrary, NOE cross-relaxation rates in small molecules are generally positive. The change of sign of cross-relaxation between small molecules and their complexes with macromolecules can be used to detect binding.

Polarization transfer experiments have many advantages, such as little restriction on size of the macromolecule, no need of isotopic labelling and potentially, some information about the orientation of the small molecule relative to the macromolecule. This class of experiment has many variations. Two popular experiments are presented in the following paragraphs.



NOE pumping experiment is a good combination between diffusion filter and NOE transfer. It overcomes shortcoming (interferences resulted from the receptor) of classic affinity NMR by adding a NOE block after a diffusion experiment and before detection<sup>48</sup> (*Fig II.3*). The experiment first suppresses the polarization of all fast diffusing small molecules with a standard bipolar gradient pulse pair stimulated echo (BPP-STE). The BPPs allow to reduce artefact arising from the induced eddy currents. Then, the preserved polarization of macromolecules is used to pump up the polarization of interacting small molecules with NOE. As results, all detectable signals are from small molecules in exchange between a free state and a state bound to a macromolecule.



*Figure II.3 Pulse sequence of BPP-STE NOE-pumping experiment. The sequence is composed by a BPP-STE diffusion filter to suppress coherences of all fast diffusing molecules and a NOE experiment to pump up polarization of binding ligand. Typical encoding delay  $\Delta$  is about 50 ms, and NOE mixing time increase from several milliseconds to seconds.*

Saturation transfer difference (STD) experiments are another type of popular NMR method for detecting ligand interactions, particularly in the context of drug screening.<sup>49-52</sup> *Fig. II.4* illustrates the STD pulse sequence. Selective saturation pulses are applied to the methyl group region, which is specific to the protein. Fast spin diffusion leads to efficient saturation of the whole protein, and any bound small molecule. The exchange between the free and bound states for the small molecules leads to saturation transfer to the detectable free state of the ligand. A reference experiment is obtained by moving the selective irradiation far off resonance. The difference between the reference and saturated spectra only shows the signals left from the binding small molecules.<sup>49, 51</sup> For higher resolution, STD can be combined with different NMR experiment such as HSQC or TOCSY.<sup>49-51</sup>

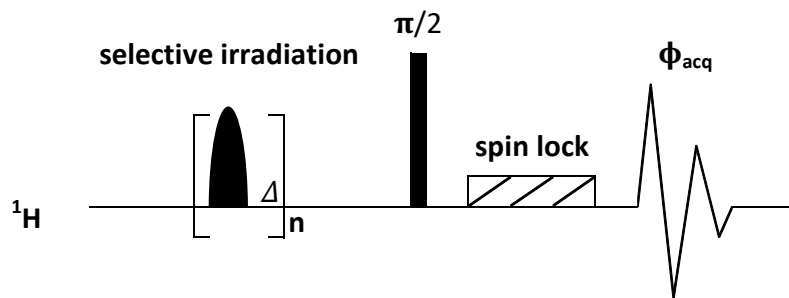


Figure II.4 Pulse sequence of STD experiment. The selective pulse is switched between the methyl group region and off resonance frequency after every scan. The subtraction is performed after every scan by phase cycling. The typical selective pulse is 20-50 ms gauss pulse. The standard delay  $d$  between the pulses ( $\Delta$ ) is 1 m and the number of repetitions of the selective pulses ( $n$ ) is around 50.

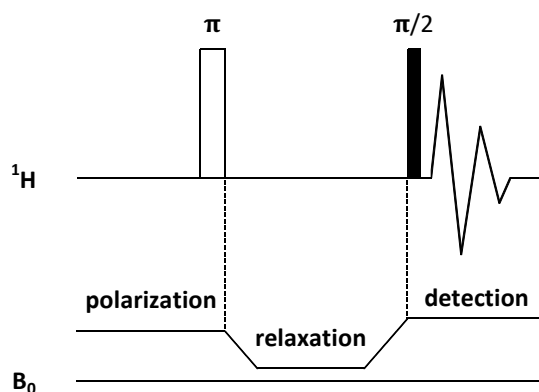
Although this type of methods is very popular<sup>53</sup>, they provide qualitative information but not quantitative data, as diffusion methods do. The efficiency of the NOE depends on several factors (proton density, local dynamics, etc.), so that this experiment only identifies binding of a small molecule to a macromolecule, but the determination of the kinetics and thermodynamics of binding from NOE-based data is more challenging.

### 2.3 Relaxation times

Relaxation rates can be excellent probes for molecular size. At high magnetic fields, transverse relaxation rates are strongly correlated to molecular size. A direct observation of fast transverse relaxation is line-broadening. When binding to a macromolecule, the linewidth of the signals in the small increases. In complex mixture, this method is rather limited due to signal overlap. A way to suppress the signal of small molecules interacting with macromolecules is to apply relaxation filter, so that signals with short relaxation times vanish from the spectrum. Relaxation filters can be integrated in many experiments mentioned above.

Relaxometry consists in measuring relaxation over a broad range of magnetic fields. Relaxometry is a unique probe of molecular dynamics. The variation of relaxation rates with the magnetic field helps to identify motions over a broad range of timescales. Fast field-cycling (FFC) is a popular technique for relaxometry. An FFC relaxometer is a system where the field can be switched between several kHz and 42 MHz ( $\sim 100 \mu\text{T}$  to 1 T) within a few milliseconds. The sample is polarized and detected at a relatively high field for better sensitivity. The magnetic field is rapidly switched down to lower magnetic fields during the relaxation delay.

The measurement of a series of relaxation rates over orders of magnitude of magnetic fields allows one to extensively probe the spectral density function (*Fig II. 5*).



*Figure II.5 Schematic presentation of the principle of FFC experiment. The polarization and detection take places at highest possible fields. The durance and field magnitude of relaxation processes can be varied as needed.*

FFC relaxometry is a low-resolution technique. The measured collective relaxation comes from the nuclei of the whole sample. FFC relaxometry has been used to probe motions in a host of systems<sup>54-56</sup> including the characterization of paramagnetic agents<sup>57</sup>. Relaxometry is also applied to studies of protein dynamics. By dissolving protein in  $\text{D}_2\text{O}$  on millimolar concentration level, it is possible to measure directly the collective relaxation of protons in a protein and define a collective order parameter<sup>58-60</sup>. It is also possible to measure the effects of protein dynamics on the relaxation of the solvent (in particular deuterium in  $\text{D}_2\text{O}$ <sup>61</sup>). Obtaining site-specific information in a macromolecule or information on the different components of a complex mixture is difficult and often impossible with FFC relaxometry. To determine motions on nanosecond to microsecond time scales with atomic resolution, high-resolution relaxometry is a unique approach<sup>62-63</sup>. The high-resolution relaxometry will be presented in detail in the experimental section.

### 3 A model system for an interaction between small and large molecules

Interactions between proteins and their ligands have diverse binding modes and kinetics. Here, we present our model of binding. We consider two molecules with significant different radii. The radius (hydrodynamic or gyration) of the macromolecule  $r_{macro}$  is much larger than the radius of the small molecule  $r_{small}$ :  $r_{macro} \gg r_{small}$ . Both the macromolecule and the small

molecule are considered as rigid for the sake of simplify. The correlation time for overall rotational diffusion of the macromolecule  $\tau_{macro}$  is much larger than that of the small molecule  $\tau_{small}$ . In the case of binding, the small molecule is assumed to be in fast exchange between one single free form and one single bound form (Fig.II. 6).

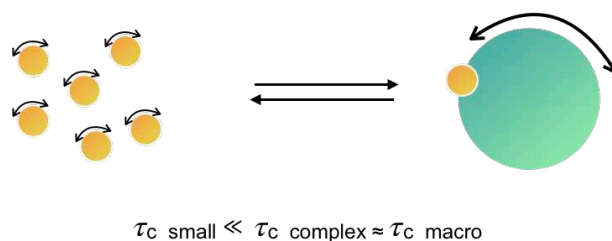


Figure II.6 Schematic presentation of the simple interaction model

A model sample is designed to test the theory and to implement the experimental method. The interaction to be investigated is between 3-(Trimethylsilyl)propionic-2,2,3,3-d4 (TSP) and bovine serum albumin (BSA)<sup>64-65</sup>. Alanine is also added into the same sample as a reference of non-binding molecule. More details are presented in experimental part below.

## 4 Relaxation theory

Here, we aim to develop a method to characterize interactions that is applicable to biological samples in the absence of isotopic labeling. Therefore, we concentrate on relaxation properties of proton spins. In addition, relaxometry is more conveniently performed on longitudinal relaxation. Thus, we will only consider longitudinal relaxation properties in proton spin systems. Relaxation is assumed to arise from proton dipole-dipole interactions. Here, we neglect cross-relaxation effects as well as other relaxation mechanisms. We will introduce the basics of the model on a simple case, where the molecules are considered rigid. We will then include local motions in the model.

### 4.1 Rigid model

Following the Bloch - Wangsness - Redfield semi-classical theory introduced in the first chapter, we describe the relaxation of protons in the small molecule. Proton longitudinal relaxation due to dipolar interactions with the environment can be described as:

$$R_{env} = \frac{1}{64} \frac{\mu_0^2 \gamma_H^4 \hbar^2}{\pi^2 r_{env}^6} [J(0) + 3J(\omega_H) + 6J(2\omega_H)]. \quad [2.4]$$

The collective longitudinal relaxation due to dipolar interaction within identical protons, such as the three protons in a methyl group, can be described as:

$$R_{equ} = 2 * \frac{3}{4} \frac{\mu_0^2 \gamma_H^4 \hbar^2}{\pi^2 r_{intra}^6} [J(\omega_H) + 4J(2\omega_H)], \quad [2.5]$$

where  $\mu_0$  is the permeability of vacuum,  $\gamma_H$  is the proton gyromagnetic ratio,  $r_{env}$  is the effective proton-proton distance,  $r_{intra}$  is the distance between equivalent protons,  $\omega_H$  is the proton Larmor frequency, and  $J(\omega)$  is the value of the spectral density function at frequency  $\omega$ . To simplify the equations, we replace the constants part in equation [2.4] and [2.5] as  $A = \frac{1}{64} \frac{\mu_0^2 \gamma_H^4 \hbar^2}{\pi^2 r_{env}^6}$  and  $B = \frac{3}{2} \frac{\mu_0^2 \gamma_H^4 \hbar^2}{\pi^2 r^6}$ , respectively. B can be calculated, because the distance between protons in a group is known.

Here, we use a simple model of the spectral density function, assuming that it has the Lorentzian form defined with a single correlation time  $\tau$  (a model corresponding to tumbling of a rigid spherical molecule in solution):

$$J(\omega, \tau) = \frac{2\tau}{5(1 + \omega^2 \tau^2)}. \quad [2.6]$$

We consider that, in the free form, the relaxation is mainly due to intramolecular dipolar interaction with equivalent protons,

$$R_{free} = R_{equ-free} = B \times [J(\omega_H, \tau_{free}), 4J(2\omega_H, \tau_{free})]. \quad [2.7]$$

In the bound form, the relaxation is due to dipolar interaction with equivalent protons and the environment,

$$\begin{aligned} R_{bound} &= R_{env-bound} + R_{equ-bound} \\ &= A \times [J(0, \tau_{bound}) + 3J(\omega_H, \tau_{bound}) + 6J(2\omega_H, \tau_{bound})] \\ &\quad + B \times [J(\omega_H, \tau_{bound}) + 4J(2\omega_H, \tau_{bound})], \end{aligned} \quad [2.8]$$

where  $\tau_{free}$  and  $\tau_{bound}$  are the correlation times of the free and bound form molecule.

We assume that the metabolite is in fast exchange between its free form and bound form. In the presence of transient binding to a macromolecule, relaxation rates were considered to be a population weighted average of the relaxation rates in the free and bound states (similar to observed diffusion coefficient mentioned in the beginning of this chapter). The final form of the equation used to fit the relaxation dispersion profile then followed:

$$\begin{aligned}
 R_1 &= (1 - pop) \times R_{free} + pop \times R_{bound} \\
 &= (1 - pop) \times B \times [J(\omega_H, \tau_{free}) + 4J(2\omega_H, \tau_{free})] \\
 &\quad + pop \times \{A \times [J(0, \tau_{bound}) + 3J(\omega_H, \tau_{bound}) + 6J(2\omega_H, \tau_{bound})] \\
 &\quad \quad + B \times [J(\omega_H, \tau_{bound}) + 4J(2\omega_H, \tau_{bound})]\}. \tag{2.9}
 \end{aligned}$$

where the correlation time of the free molecule ( $\tau_{free}$ ), correlation time of the complex of the small molecule with a macromolecule ( $\tau_{bound}$ ), population of the small molecules bound to the macromolecule ( $0 \leq pop \leq 1$ ), and an additional parameter ( $A$ ) are to be optimized.  $B$  can be calculated based on the distance between protons in a group.

#### 4.2 Trimethylsilyl model including internal motions

In the simple model introduced above, both the small molecule and the complex with the macromolecule are considered rigid. For more accurate quantitative analysis, two local motions of the small molecule (*i.e.* TSP) can be considered: methyl group rotation and trimethylsilyl group rotation. The structure of partially deuterated TSP is shown in *Fig. II.7*. When TSP is in its free form, both internal motions should be considered.

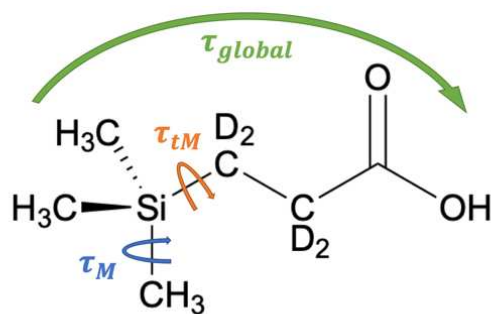


Figure II.7 Schematic presentation of the TSP molecule structure. Two internal motions are considered: methyl group rotation around the Si-CH<sub>3</sub> bond (in blue) with a correlation time  $\tau_M$  and an order parameter  $S_M^2$  and trimethylsilyl group rotation around the Si-CD<sub>2</sub> bond (in orange) with a correlation time  $\tau_{tM}$  and an order parameter  $S_{tM}^2$ .

The order parameters  $S_M^2$  and  $S_{tM}^2$  for the methyl and trimethylsilyl rotations are equal for their similar geometry<sup>15</sup>:

$$S_M^2 = S_{tM}^2 = \left[ \frac{1}{2} (3 \cos^2 \theta - 1) \right]^2 = \frac{1}{4}, \quad [2.10]$$

where  $\theta = 90^\circ$  is the angle between the axis of rotation and the principal axis of the dipolar interaction (internuclear vector). The rotation of methyl groups is assumed to be very fast. The correlation function for the dipole-dipole interaction within the trimethylsilyl group in the free form of TP is then:

$$J_{free}(\omega, \tau) = S_M^2 \times [S_{tM}^2 \times J(\omega, \tau_{global}) + (1 - S_{tM}^2) \times J(\omega, \tau_{eff})] \quad [2.11]$$

where  $\tau_{free}$  is the rotation correlation time of the trimethylsilyl group:

$$\tau_{eff} = \frac{\tau_{tM} * \tau_{global}}{\tau_{tM} + \tau_{global}}, \quad [2.12]$$

and  $\tau_{global}$  is the global rotational correlation time of TSP molecule.

The trimethylsilyl group of TSP is a hydrophobic region. When binding to macromolecules, we consider that the trimethylsilyl group loses the freedom to rotate, while the rotation of methyl groups is still very fast. The corresponding correlation function is

$$J_{bound}(\omega, \tau) = S_M^2 \times J(\omega, \tau_{bound}). \quad [2.13]$$

By inserting equations [2.10], [2.11] and [2.13] into [2.9], the obtained complete proton relaxation rate for TSP is:

$$\begin{aligned}
 R_1 = & \frac{B(1 - pop)}{4} \times \left[ \frac{1}{4}J(\omega_H, \tau_{global}) + J(2\omega_H, \tau_{global}) + \frac{3}{4}J(\omega_H, \tau_{eff}) + 3J(2\omega_H, \tau_{eff}) \right] \\
 & + \frac{pop}{4} \times \{B \times [J(0, \tau_{bound}) + 3J(\omega_H, \tau_{bound}) + 6J(2\omega_H, \tau_{bound})] \\
 & \quad + A \times [J(\omega_H, \tau_{bound}) + 4J(2\omega_H, \tau_{bound})]\} \tag{2.14}
 \end{aligned}$$

## 5 Experimental section

### 5.1 Instrumentation

The principle of high-resolution relaxometry is to use the stray field of a high-field high-resolution NMR spectrometer as a variable magnetic field for relaxometry. Combining relaxation at low magnetic field and high-resolution NMR at high magnetic field is achieved with a sample shuttle that moves the sample fast between the magnetic center and chosen positions in the stray field. We have used a sample shuttle developed by our collaborators from Bruker BioSpin.<sup>24</sup> The pneumatic shuttle system<sup>27</sup> is installed on a commercial 600 MHz NMR spectrometer, equipped with a specially designed probe (*Fig II.8*). A customized sample container was designed. Optical-detectors placed at the top and bottom positions send their information to the shuttle controllers (*via* satellites). A pre-set parameter script implemented in the main shuttle controller adjusts the air flow required to move the sample. The sample can move inside a shuttle guide (tube) that is a little longer than the bore of the magnet. The stray field covers all fields between 14.1 T and about 15 mT over the length of the shuttle guide. Note that ferroschims installed in the bore of the magnet locally modify the magnetic field and create a plateau at 0.33 T. These ferroschims were installed to perform two-field NMR experiments.<sup>66-68</sup>

A high-resolution relaxometry experiment starts with high-field polarization ( $B_0 = 14.1$  T). The sample is then shuttled to low field at a chosen position determined by the transfer system. The sample stays at low field during the relaxation delay and is shuttled back to the high-field position for detection. This scheme preserves the sensitivity and resolution offered by high-field magnets.



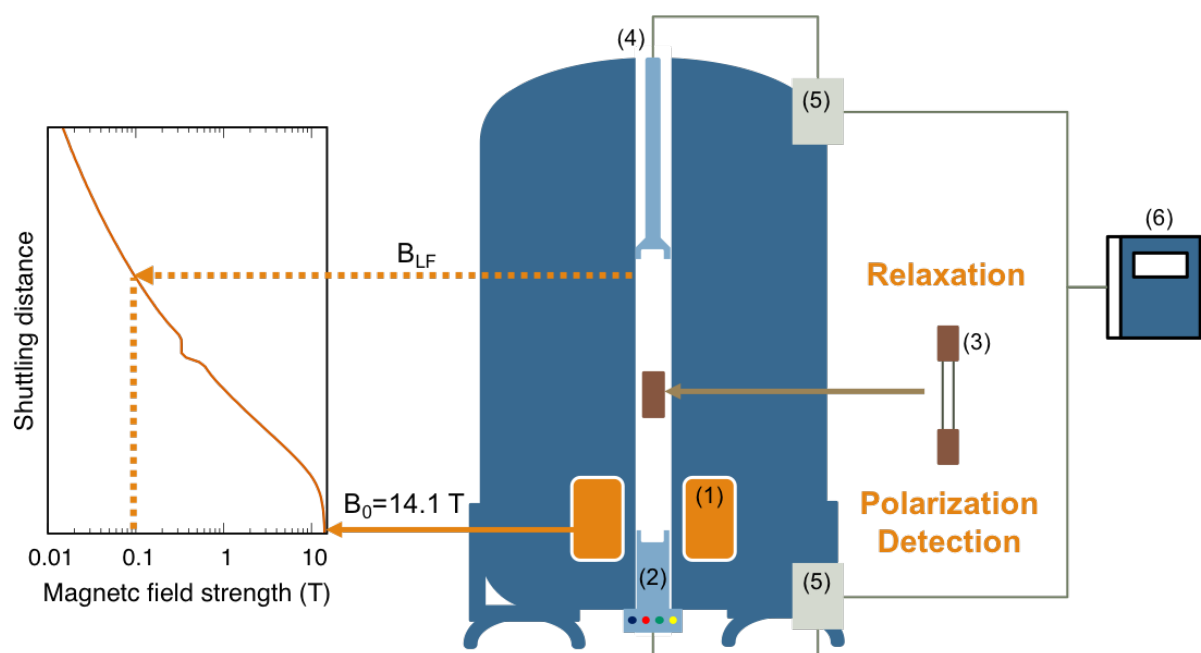


Figure II.8 Schematic view of the shuttle relaxometer system. (1) magnet operating at 14.1 T; (2)  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple-resonance shuttle probe equipped with z-axis gradient; (3) shuttle tube with an active volume of  $60\ \mu\text{L}$  and end plugs (brown rectangles) protecting the tubes during the transfer and landing; (4) maneuverable stopper to adjust the position for relaxation in the stray field; (5) shuttle satellites controllers for top and bottom position; (6) main shuttle controller.

## 5.2 Pulse sequence

We have adapted a standard 1D pulse sequence used for metabolomics and inserted shuttle transfers and relaxation delays. The 1D NOESY pulse sequence (Fig. II.9 A) is one of the conventional experiments used in metabolomics by NMR, for its efficient water signal suppression<sup>69</sup>. The pulse scheme is similar to the standard 2D NOESY but the mixing time is fixed  $\tau_{\text{mix}} = 10\ \text{ms}$ .<sup>69-70</sup> As explained before, distinguishing between large and small molecules is done using relaxation rates profiles such that this mixing time is kept constant in our experiments to a value of 10 ms.

The relaxometry experiments were derived from the conventional 1D NOESY pulse-sequence by introducing a relaxation delay at low field after the mixing time ( $\tau_{\text{mix}}$ ) and before the last  $\pi/2$  pulse (Fig. II.9 B). The sample is transferred to a desired low field  $B_{LF}$  (from 0.015 to 10 T) with a transfer time  $\tau_{HF-LF}$  ( $\sim 55 - 210\ \text{ms}$ ), held in that position for a variable relaxation delay  $T_{rel}$  and transferred back to high field with a transfer time  $\tau_{LF-HF}$  ( $\sim 60 - 260\ \text{ms}$ ) for detection. To measure relaxation rates at 14.1T, the shuttle transfers are simply omitted, such that the sample stays at high field during the whole experiment. The polarization of water can

recover during the shuttle transfers and the stabilization delay  $\tau_{st}$  so that it is necessary to perform an additional water suppression before detection. A Watergate (WG) block is added before detection. Note that the presence of a strong water signal has significant consequences in shuttling context: The small vibrations of sample after shuttling lead to a modulation of this strong signal leading to artefacts throughout the spectrum. Proper suppression of the strong signal of the solvent efficiently suppresses these artefacts.

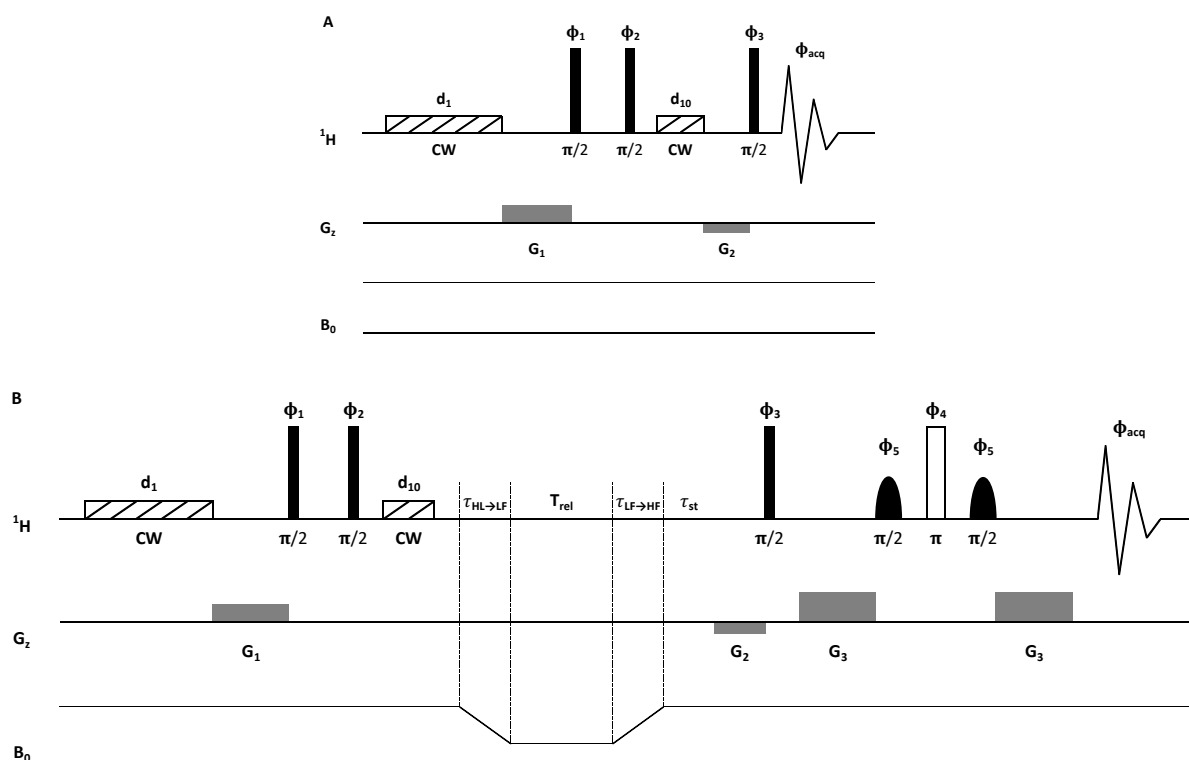


Figure II.9 A. Standard 1D NOESY experiment employed in metabolomics. B. 1D NOESY experiment adapted for relaxometry shuttle experiment: a relaxation delay is added after the mixing time and Watergate block is inserted before detection for better solvent signal suppression. 1D NOESY\_WG pulse sequence used to measure relaxation dispersion profiles. The sample is moved from the high-field magnetic center of the spectrometer to a chosen position at low field and back during the period  $\tau_{HL-LF}$  ( $\sim 55$ - $210$  ms) and  $\tau_{LF-HF}$  ( $\sim 60$ - $260$  ms), respectively. The variable relaxation delay at low magnetic field is labelled  $T_{rel}$ , the stabilization delay after the sample returns to the high-field position is labelled  $\tau_{st} = 100$  ms. Narrow and wide rectangles represent  $\pi/2$  and  $\pi$  proton pulses, respectively; and the semicircular present 1.5 ms selective pulse for  $\pi/2$  water signal. Striped rectangles stand for continuous-wave pre-saturation irradiation applied for durations  $d_1=4$  s and  $d_{10}=10$  ms with a low amplitude (75 Hz). The phase cycles were :  $\phi_1=(x, -x)$ ;  $\phi_2=8(x), 8(-x)$ ;  $\phi_3=2(x), 2(-x), 2(y), 2(-y)$ ;  $\phi_4=2(y), 2(-y), 2(-x), 2(x)$ ;  $\phi_5=2(-y), 2(y), 2(x), 2(-x)$ ;  $\phi_{acq}=(x, -x, -x, x, y, -y, -y, y, -x, x, x, -x, -y, y, y, -y)$ . The amplitudes of the 1ms pulsed field gradients  $G_1$ ,  $G_2$  and  $G_3$  applied along z-axis were 8.5, 4 and 20 G/cm, respectively.

### 5.3 Model sample

Feasibility studies were carried out using a model sample. This sample is designed to simulate the interaction of small molecules and macromolecules with three molecules typically found in blood plasma or serum, containing 0.5 mM of alanine, 50 mM of 3-(Trimethylsilyl)propionic-2,2,3,3-d<sub>4</sub> (TSP), used as a chemical shift reference, and 0.5 mM of bovine serum albumin (BSA), as a substitute for human serum albumin.

We use alanine as a reference. Alanine is a metabolite naturally present in many biological fluids, such as blood serum, for which no interaction with macromolecules was detected<sup>71</sup>. TSP is often introduced as an internal standard in metabolomics studies. Its interaction with macromolecules, in particular albumin, is known.<sup>64-65</sup> Serum albumin is the most abundant protein in blood serum and plasma and it is known to bind many different molecules<sup>72</sup>, hence we used BSA as reference macromolecule (MW = 66.5 kDa), which also grants a sample viscosity similar to blood plasma. We employed highly-purified and fat-free BSA in order to avoid interference or competition arising from small impurities. Alanine and BSA were used at their physiological concentration<sup>73-74</sup>. All compounds were dissolved in phosphate buffer (pH = 7.4) containing 35 mM of Na<sub>2</sub>HPO<sub>4</sub>, 10% of D<sub>2</sub>O and 0.0125% of NaN<sub>3</sub> as bacteriostatic agent.

### 5.4 Temperature control

To ensure relaxation rates would be measured at same temperature for each low field with the shuttle system, special protocol was used. A temperature calibration sample contains deuterated acetic acid buffer (50 mM) and 7% D<sub>2</sub>O. It was used to perform blank relaxation experiment to reproduce real experimental conditions, including shuttle transfers, relaxation at the low field position, etc. Then the temperature was calibrated using deuterium chemical shift difference between acetic acid and water (OD) signals. We ensured that NMR measurements were performed at (298.0 ± 0.2) K for all delays at all fields.

### 5.5 Data analysis

The collected Bruker data was processed using NMRPipe<sup>75</sup>. Signal integrals were obtained by summing up intensities in defined chemical shift ranges, corrected by the baseline intensities of the two ends. In case where the baseline is not linear, a polynomial function was

used. The errors of the integrals were estimated based on the random noise in regions of the spectra free of any signal.

The relaxation rates at each field were obtained by a fitting of the decay of the signal integrals to a mono-exponential function using the program Octave<sup>76</sup>. Fitted relaxation rates were then used to characterize interactions between small molecules and macromolecules. In order to account for potential sources of uncertainty other than spectral noise, errors in relaxation rates errors were considered to be at least 2%. Models mentioned in the theoretical section were used in Markov chain Monte Carlo analysis<sup>77</sup> and employed to find quantitative distributions of the parameters of the models.

## 6 Results and discussion

### 6.1 Plotting and qualitative analysis

Longitudinal relaxation rates of the model sample were measured at 19 magnetic fields covering 3 orders of magnitude (14.1, 10.0, 5.0, 2.5, 1.25, 0.625, 0.33, 0.23, 0.16, 0.12, 0.10, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.015 T). Each relaxation decay was characterized by 7 relaxation delays (between 0.001 and 1 s depending on the field). Methyl proton signals of alanine and TSP are found at 1.4 and -0.1 ppm respectively (*Fig.II.10*). Relaxation dispersions profiles for the methyl group signals of alanine and TSP were obtained. Their NMRD profiles showed very different features (*Fig.II.10*).

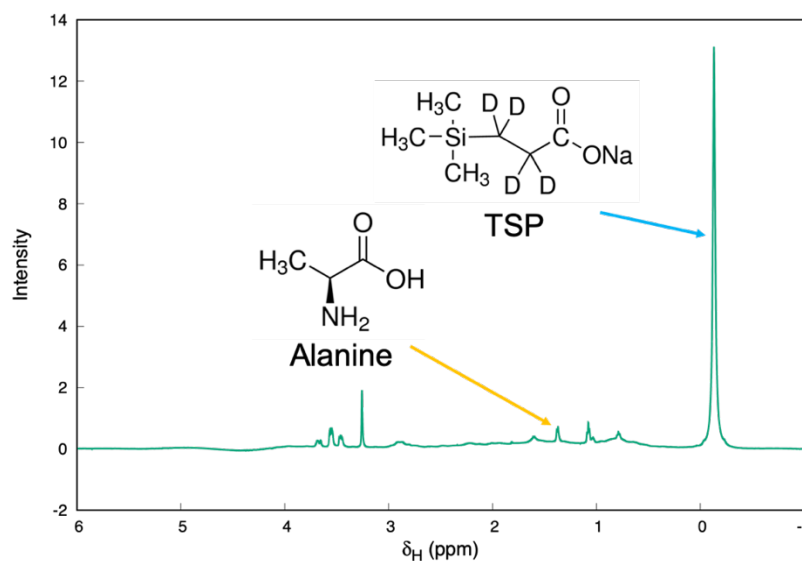
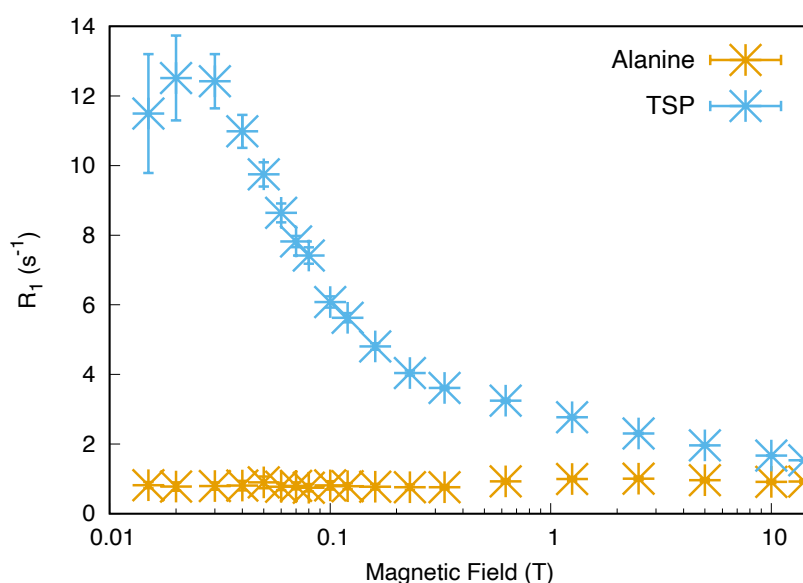


Figure II.10 1D NOESY <sup>1</sup>H spectrum of model sample: 0.5 mM of alanine (1.4 ppm), 50 mM of TSP (-0.1 ppm) and 0.5 mM of BSA (broad signals just above the baseline).

The nuclear magnetic relaxation dispersion profile of the methyl group of alanine (at  $\sim 1.4$  ppm) is shown in *Fig. II.11*. The longitudinal relaxation rate of this methyl group has no noticeable dependence over the magnetic field. The relaxation rate is close to  $0.85\text{ s}^{-1}$  over the entire range of magnetic fields. This absence of dependence of the longitudinal relaxation rate shows that the methyl protons of alanine are in the “extreme narrowing” regime, as expected in the absence of binding to macromolecules.

On the other hand, TSP manifested strong evidence of binding to macromolecules. As we decrease the magnetic field, the relaxation rate of the trimethylsilyl group ( $-0.1$  ppm) first increases from  $1.5\text{ s}^{-1}$  at  $14.1\text{ T}$  to  $4\text{ s}^{-1}$  around  $0.2\text{ T}$ . Below  $0.2\text{ T}$ , the longitudinal relaxation rate increases faster after at field range up to a plateau at  $0.03\text{ T}$  and below with a rate  $\sim 12\text{ s}^{-1}$ . The two inflection points around  $3\text{ T}$  and  $0.07\text{ T}$ . If we make the approximation that the inflection point of the NMRD is also the reflection point of the spectral density, we can estimate the corresponding correlation time to be  $1/\omega$ . The two motions inflected directly by the NMRD are estimated to be  $1.25\text{ ns}$  and  $53.6\text{ ns}$ . We may tentatively assign these two correlation times to the overall correlation times of TSP and of the complex of TSP and BSA. A quantitative analysis is discussed in the next section.



*Figure II.11 Relaxation dispersion profiles of alanine (yellow) and TSP (azure) in our model sample.*

## 6.2 Quantitative analysis of TSP relaxation dispersion

Both the rigid model (Equation [2.7]) and model-free with internal motions (Equation [2.12]) were used to fit the experimental relaxation results. The constant  $B$  ( $= 25.975 \times 10^9 \text{s}^{-2}$ ) was calculated based on the distance between protons in the methyl group (1.79 Å). The fitting results are shown in Fig.II.12. The fitting from the rigid model (green curve) is far off from the experimental results. When internal rotations were included, the quality of the fit was greatly improved as shown by the red curve.

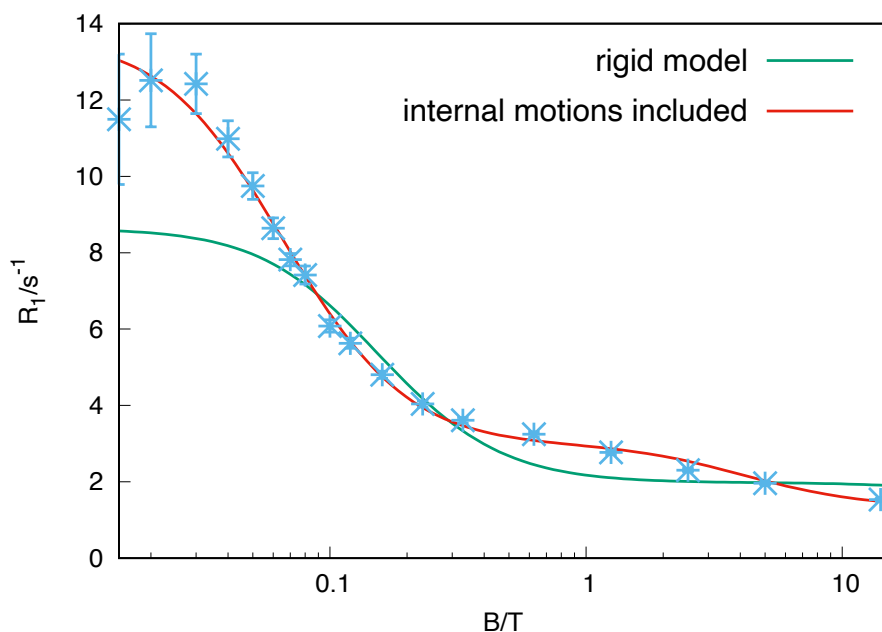


Figure II.12 Relaxation dispersion profile of the methyl groups in TSP in the model sample. The green fitting curve is obtained from the rigid model (Equation [2.9]), and the red one is resulted from the model where local internal motions are considered (Equation [2.14]).

Including internal motions model free for TSP molecule, we got a curve of same good quality at both high-field range and low-field range than the rigid model. Thanks to the optimized model, the interaction system was better defined (Table II-1). In its free form, the global rotational correlation time  $\tau_{global}$  of TSP is  $0.51 \pm 0.04$  ns. The internal rotation time of trimethylsilyl group  $\tau_{tM}$  is very short and poorly defined  $22.1 \pm 18$  ps. The error of  $\tau_{tM}$  was relatively large. But it was not key parameter to characterize the interaction. The rotational correlation time of the TSPs-BSA complex  $\tau_{bound}$  is  $35.2 \pm 4.0$  ns, which is in the range expected (MW = 66.5 kDa). Constant  $A$  ( $= (288.1 \pm 105.1) \times 10^9 \text{s}^{-2}$ ), correspond to an averaged  $r_{env}$  of 1.66~1.88Å. The binding pop ( $= 0.11 \pm 0.22$  %) were poorly defined

individually because they are strongly correlated with constant A (Fig.II.13), therefore cannot be fitted separately.

Table II-1 Fitting parameters of the TSP NMRD.

Fitted parameter	Value	Unit
A	$(288.075 \pm 105.1) \times 10^9$	s <sup>-2</sup>
$\tau_{global}$	$0.51 \pm 0.04$	ns
$\tau_{tM}$	$22.1 \pm 18$	ps
$\tau_{bound}$	$35.2 \pm 4.0$	ns
pop	$0.11 \pm 0.22$	%

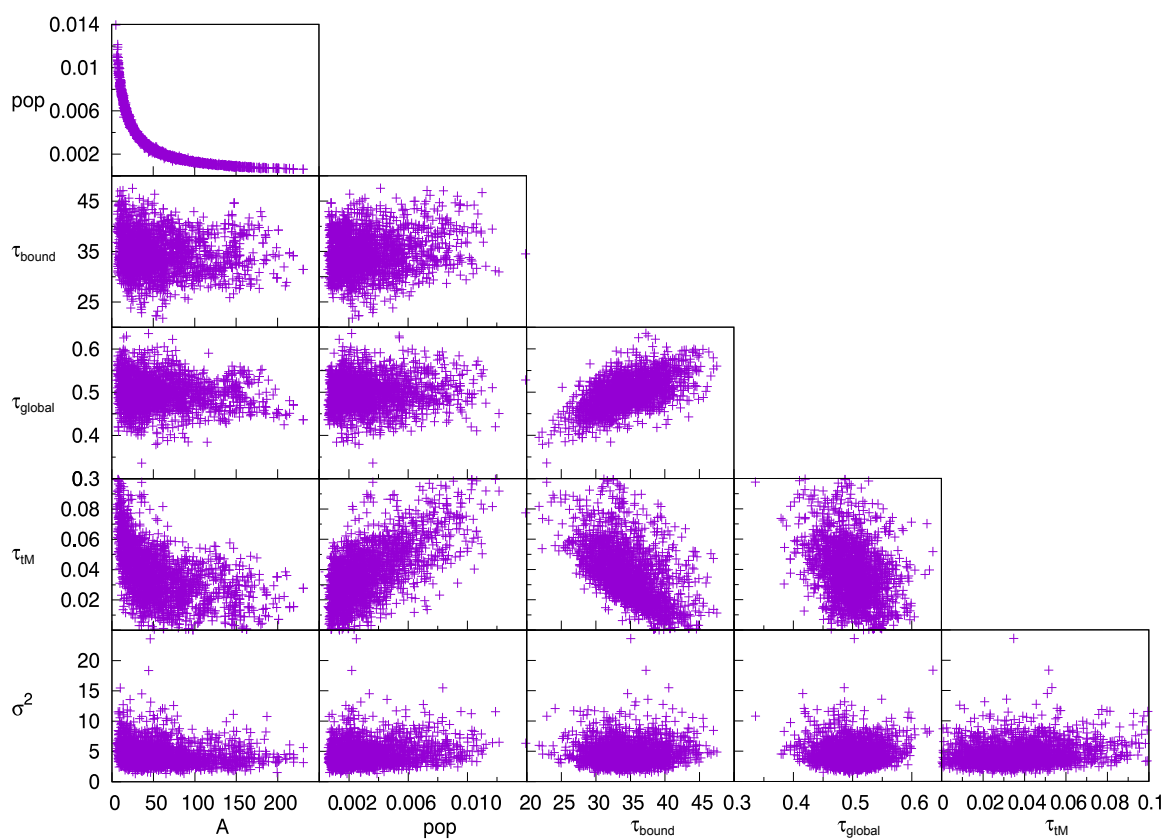
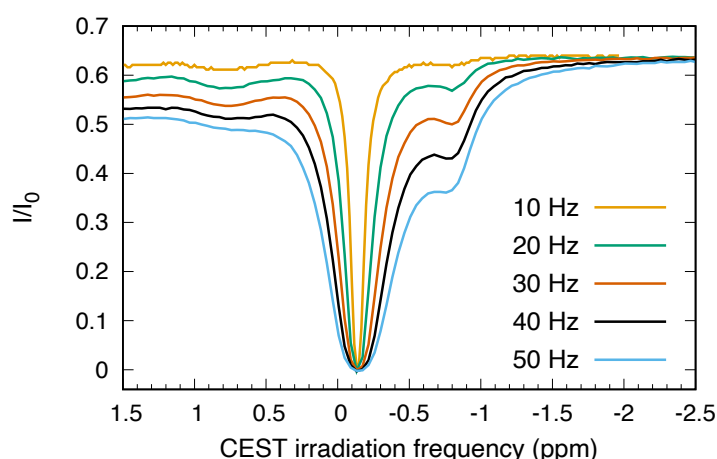


Figure II.13 Correlation plot of constant A, pop,  $\tau_{bound}$ ,  $\tau_{global}$ ,  $\tau_{tM}$  and the variance.

### 6.3 Binding of TSP to BSA is confirmed by chemical exchange saturation transfer

In order to confirm the interaction between TSP and BSA and obtaining a better estimation of the population of the bound form of TSP with BSA, chemical exchange saturation transfer (CEST) experiment<sup>25, 29-30, 78</sup> were carried out on the same sample. The power of CEST is to reveal the presence of minor states in slow exchange with a major state even if the signal of the minor state is too small and too broad to be visible in a spectrum. Weak  $B_1$  amplitudes (10-50 Hz) were used to irradiate selectively one by one all the regions of the proton spectrum. The irradiated signals are saturated by this irradiation. In the presence of chemical exchange, the saturation is transferred to the other form by slow chemical exchange. This other state can be a major state readily visible in the spectrum and the irradiated state may be a minor state, invisible in the spectrum. In this case, the invisible minor state can then be monitored through the intensity evolution of the major-state. The analysis of the CEST profile can provide information on the chemical shift of the invisible state, its population and the kinetics of exchange.

The CEST profiles obtained are shown in *Fig.II.14*. The curve represents the intensity of the signal of the methyl groups of TSP versus the position of irradiation (in ppm). The intensities were normalized by the intensity of the same peak obtained without the irradiation block. The large signal at -0.1 ppm and the small signal at -0.8 ppm correspond to the major state (free) and minor state (bound) of TSP. Only the right half of the profiles were used in the fit, for reasons that are given below. Fitting these profiles with a simple two-state exchange model gave an estimation of the population of the bound form of about 1.2%.



*Figure II.14 CEST profiles of the TSP signal in the controlled sample. Each curve represents evolution of TSP intensities after 300 ms of selective irradiation at amplitude  $B_1$  (10-50 Hz) as indicated with different colors in the figure. All intensities were normalized with the reference TSP intensities where the irradiation was set far from the spectral window (offset -11000 Hz).*



Our ability to extract an accurate estimation of interacting populations from these CEST experiments is limited here. First, the binding mechanism and sites of TSP on BSA are unknown. We used a model with two simple exchanging states of TSP to fit the data. Therefore, this result can only be considered as a rough estimate and not an accurate determination of the bound population. Second, as discussed previously in the paper of *Bouvignies et al.*<sup>79</sup>, the main problem with <sup>1</sup>H CEST is the saturation transfer to proximal protons via NOE effects. Here, cross-relaxation also leads to significant saturation transfer even when neither the free or the bound states are irradiated. As observed in the profiles, the intensities obtained when chemical shifts above 0.5 ppm are irradiated are smaller than when irradiating at chemical shifts below -1.5 ppm (where there is no signal). This is due to saturation transfer from the aliphatic protons of BSA to TSP from extensive cross-relaxation within the BSA-TSP complex. We used only half of the profiles with chemical shift below -0.1 ppm to fit the binding population to limit the effect of spin diffusion on the profile. Although this observation further compromises the quantitative analysis of the CEST profiles, it is clearly another evidence for the interaction of TSP and BSA.

## 7 Conclusion

We developed a new approach to identify binding of a small molecule to a small molecule using high-resolution NMR relaxometry. We were able to identify the interaction between the NMR standard molecule TSP and bovine serum albumin in a test sample. This interaction was confirmed by chemical exchange saturation transfer. By measuring proton NMRD over three orders of magnitude of magnetic field, the relaxation of TSP in its free and bound form was characterized. Importantly, this experiment provides an accurate estimate of the size of the complex. However, the bound population of TSP could not be determined precisely. An important feature of this approach is the ability to identify the interaction of a small molecule as long as its signal can be resolved in a one-dimensional spectrum. Thus, this approach could be used in complex fluids, which will be investigated in the following chapter.

## Chapter III Investigation of metabolite-protein interactions in biofluids by high-resolution NMR relaxometry

Metabolites are small chemical compounds issued by metabolism. Extensively existing, metabolites have multitude functions that are essential to life activities. Not only metabolites are substrates and products of enzymatic reactions, but also, they are cofactors and regulators of large numbers of biochemical processes. Although there exist different biological and chemical methods, not many allow to analyze metabolite-protein interactions in complex medium with high-throughput. In this chapter, one new approach of investigating metabolites and their interaction with macromolecules in biological samples by high-resolution NMR relaxometry will be presented.

## 1 Introduction

### 1.1 Metabolomics

#### 1.1.1 General introduction

Metabolites are intermediate and end products of metabolism, usually small molecules with a molecular weight under 1500 Da. The metabolome is the ensemble of metabolites present in a biological sample, including amino acids, lipids, carbohydrates and other small organic or inorganic molecules. Metabolomics<sup>80-84</sup> is the characterization and quantification of the full metabolome in biological samples. It is one of the -omics disciplines in system biology.

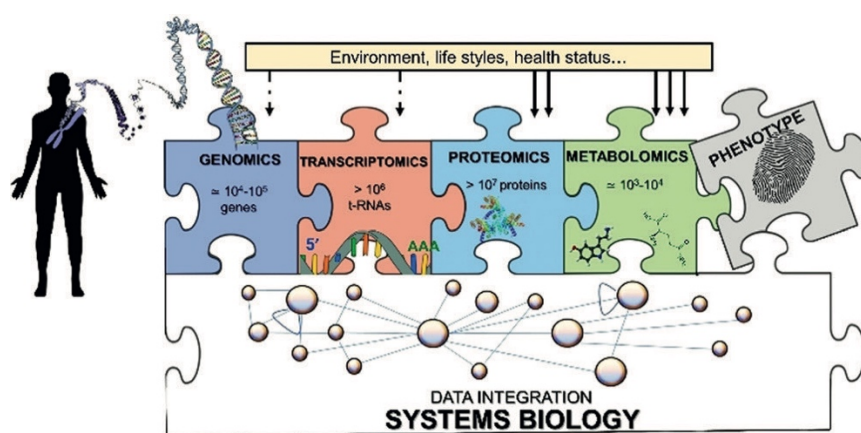


Figure III.1 The flow of information in systems biology proceeds from the genome to the transcriptome, the proteome and finally to the metabolome. They are variable during an individual lifespan, and all concur to the phenotype.<sup>85</sup>

Compared to other upstream -omics science, metabolomics has two unique qualities (Fig. III. 1): sensitivity and concision. First, metabolome is very sensitive to environmental factors. While genome stays globally unchanged during one's lifetime, the metabolome reflects stress level, nutrition state and living environment rapidly and directly. It provides rich information that allows to evaluate the health condition of an individual. Second, metabolomics contains information that is more condensed. There are about 3000 endogenous metabolites in humans. The number of biomolecules involved in transcriptome and proteome are about a thousand times larger. These features quickly make metabolomics a popular probe in many domains of application, from agriculture to food science. Most excitingly, this relatively young discipline is now generating insights into diseases such as cancers, heart diseases and kidney diseases.<sup>86</sup> It is also contributing to drug development and personalized medicine.<sup>87-88</sup>

Metabolomics works with solid samples, like a piece of tissue, or, more commonly, liquid sample, as biofluid, fermentation or tissue extraction. Different sample types need different analytical strategies. In this thesis, metabolomics of human biofluid is the main topic we will discuss.

### *1.1.2 Analytical techniques*

Metabolomics is a science that relies on accurate analytical technologies. Mass spectrometry (MS) and NMR spectroscopy<sup>85,89-95</sup> are the two major tools used in metabolomics. They give two different approaches to metabolomics. Both techniques provide rich types of information that are complementary for metabolite profiling. Because of the difference in their basic principles and properties, each technique has its pros and cons.

MS is a powerful technique. It is very sensitive: the detection limit can be down to the picomole level. Many different types of MS analysis can be performed. It is possible to couple MS with preparation and separation methods, as extractions, chemical modifications and chromatography, depending on physical and chemical properties of molecules to be detected. Though, elaborate preparation is a double-edged sword. During sample processing, there is a loss of information and a higher chance to induce errors. Ionization is another critical point of MS. Due to matrix effects, the relative intensities of mass spectra do not reflect accurately concentration ratios.

In spite of its lower sensitivity, NMR is a unique tool to investigate and characterize complex samples. NMR requires much simpler sample preparation than MS. Usually, only one simple dilution is needed for a liquid sample. This straightforward preparation comes with many benefits. It is less time consuming, there are fewer chances to induce human error. It allows to preserve the complete information of the sample and to work in conditions very close to the real physiological state. For most abundant metabolites, quantification by NMR is more accurate and reproducible than MS.

### *1.1.3 NMR strategies for metabolomics*

In general, there are four steps to carry out a metabolomics study by NMR (*Fig. III. 2*): sampling, NMR experiment performance, NMR data processing and statistical analysis. Every step should be made delicately following the standardized protocol<sup>96</sup>, for quality results. Depending on the context of the study, metabolomics adapts itself with different strategies. Here, we focus on the analysis of human biological liquid samples by NMR.

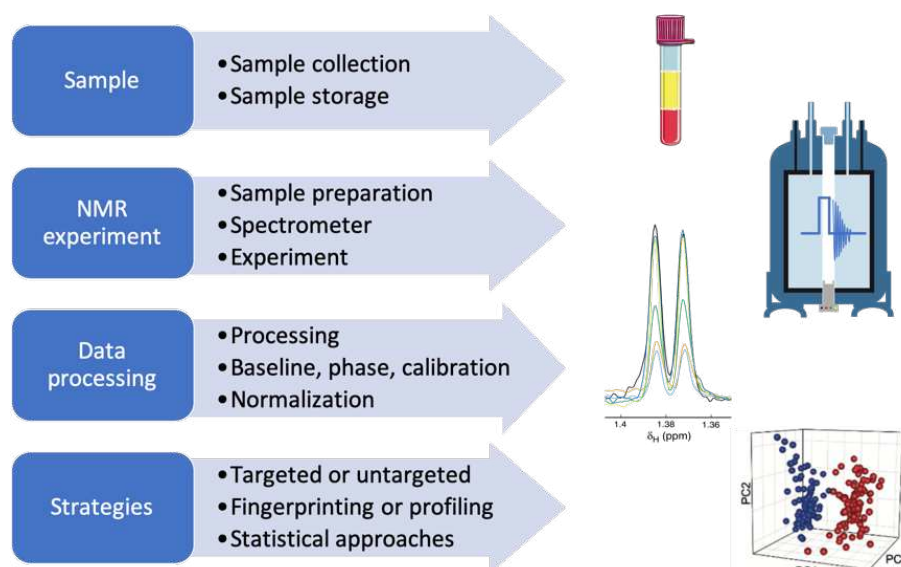


Figure III.2 Workflow of metabolomics study by NMR. Figure is based on, and the figure of principal component analysis is adapted from the review of Vignoli et al.<sup>85</sup>

Blood plasma or serum and urine are the most common subjects for human metabolomic studies. Metabolome is very sensitive to the change of environmental factors. From the patient condition, sample collection, transportation and storage, till analytical phase, the quality control of the sample is critical to the final results.

## 1.2 Metabolite-protein interactions

Metabolites affect cellular activities in intricate ways. Not only metabolites are substrates and products of enzymatic reactions, but they are also cofactors and regulators of large numbers of proteins<sup>97</sup>. Metabolites bind to macromolecules, usually proteins, and alter their functions in signaling, regulation and other cellular processes. Metabolite-protein interaction is one important subject in an emerging domain, interactomics. By contrast to protein-protein and DNA-protein interactions, the understanding of MPIs is very limited, for lack of systematic analytical methods. To answer the need of a full picture of the biological network,<sup>39, 98</sup> there are new approaches in the progress of development, using advanced preparation and detection technologies.

MPIs are often studied in indirect ways. Biological methods of investigating MPI usually involve chemical modification of the metabolite or tagging the protein. This limits the subject to interaction between a unique protein and metabolites. Besides, MPI are of large range of natures, catalytical, allosteric, and driven by various effects: hydrophobic effect, hydrogen

bounds, electrostatic interactions. The lifetime of the complex can be very variable. Of all MPIs, weak interactions are hard to preserve if the method requires extensive preparation steps.

### 1.3 Available techniques

Different strategies to investigate MPIs have been reviewed, based on the detection method (radioactivity, fluorescence, MS, NMR), metabolite type (hormones, lipids, drugs) or studied system (*in vitro*, *in vivo*, *in silico*).<sup>39-41</sup> These methods detect mostly simple and specific binding between an individual protein and its ligand involved in enzymatic reactions. Protein microarray coupled with mass spectroscopy (MS) has been adapted for MPI studies. Being one of the popular techniques, it allows analysis of multiple bindings in parallel, yet needs highly purified sample and specific arraying method.<sup>99</sup> Several other available approaches have the capacity of detecting multiple interactions in more complex environments.

Differential radical capillary action of ligand assay (DRaCALA) developed by *K. G. Roelofs et al.* allows analyzing qualitatively protein-ligands interactions in whole cell lysate (*Fig. III. 3*).<sup>100</sup> By immobilizing during a few seconds proteins and the radiolabeled ligands on nitrocellulose membrane, bound metabolites were separated from free ones. It was also possible to determine kinetics according to the migration distances of the metabolites, if the protein was purified. This method is simple and rapid. The drawback is the use of radioactive agents. It is also limited, if the ligand is naturally present in the cell lysate.

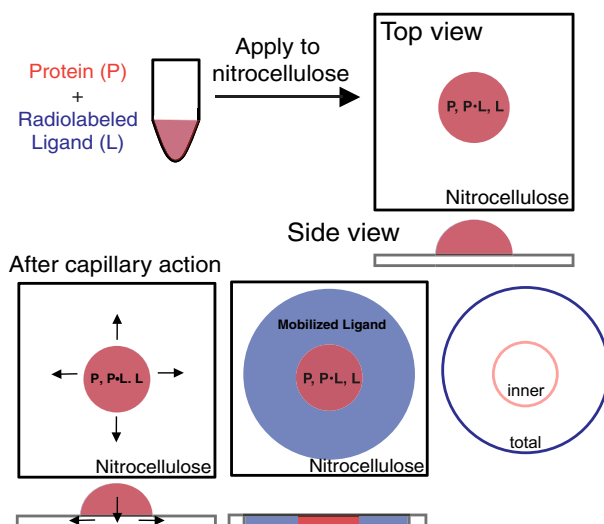


Figure III.3 Schematic representation of DRaCALA assay on application of protein-ligand mixture onto nitrocellulose and capillary action. Protein (P), ligand (L), and protein–ligand complex (P•L) distribution during the assay is shown.<sup>100</sup>

Large-scale investigation of metabolite-protein interactions was performed by *Li et al* (Fig. III. 4).<sup>101</sup> The *in vivo* method copurifies small metabolites bound to the targeted protein fused with IgG tag in yeast cells lysate.<sup>101</sup> The metabolites were further extracted and analyzed by liquid chromatography (LC) and atmospheric pressure chemical ionization (APCI)-MS. Due to loss during the extraction, only hydrophobic molecules were detectable.

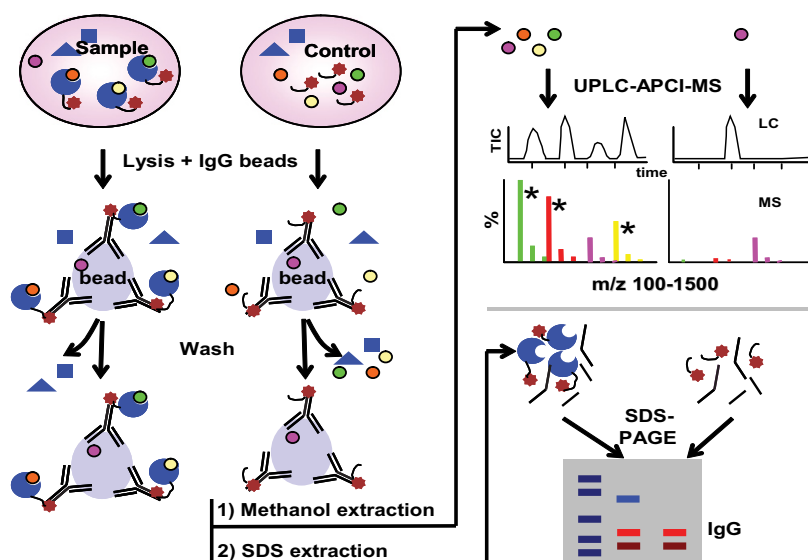


Figure III.4 Flowchart for the identification of small metabolites bound to proteins.<sup>101</sup> The target protein is tagged with an Ig-binding domain and expressed in yeast. The cell lysate is treated with a solid phase to isolate the recombinant proteins. The metabolites bound to the protein is extract by methanol to be analyzed by MS. The proteins are extracted by SDS and analyzed on gel. Native yeast extractions are used to control the results.

More recently, *Piazza et al.* employed limited proteolysis (LiP) and MS to identify and map protein-metabolite interactions, named Lip-SMap (Fig. III. 5).<sup>102</sup> It is based on shielding of protein binding sites by metabolites. With the presence or absence of small molecules, the conformation of bound proteins changed and were fragmented differently by LiP. Multitudinous interactions of broad range of affinities and putative binding sites were identified, among which many are novel. The high-throughput method is efficient and needs no specific modification either on the ligands nor on the proteins. Yet, this approach has limitations. First, not all interactions are detectable: it is not efficient, if no proteinase K (PK) cleavage site presents on the binding site, in the case of low abundance protein, or in the presence of a weak interaction. Second, the binding sites are only putative. Protein conformation changes are not



straightforward. Its connection with metabolite binding can be direct or tangential, for example in the case of allostery.

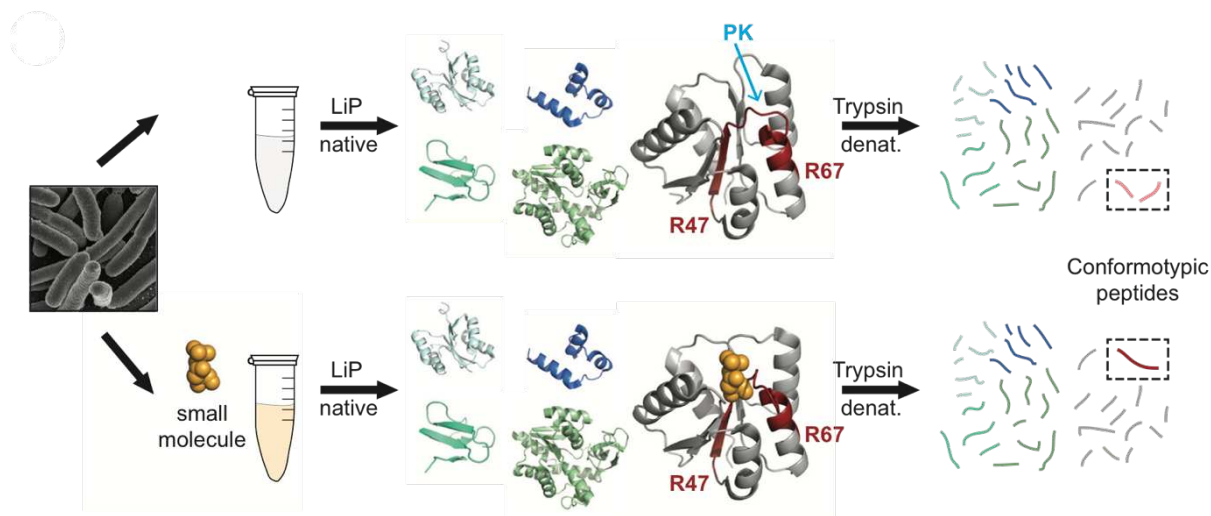


Figure III.5 Whole-cell lysates extracted under native lysis conditions are treated or not with a metabolite. Metabolite binding to a protein alters local proteolytic susceptibility. A limited proteolysis step is performed with proteinase K (PK) under native conditions, followed by complete digestion with trypsin under denaturing conditions to generate MS-measurable peptides. In the example, structurally informative peptides are produced for the protein FixJ bound to aspartyl phosphate (PDB: 1DBW) and its ligand-free form (PDB: 1D5W). Peptides specific for the bound and unbound conformations (conformotypic peptides) are depicted in red.<sup>102</sup>

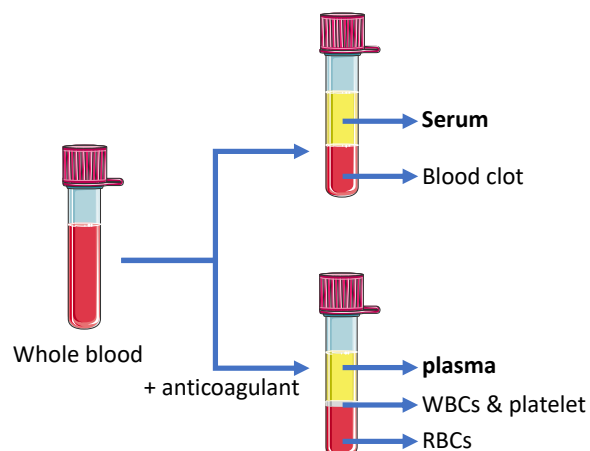
Despite recent advances in methodologies, quantitative and systematic MPI investigation remains challenging. NMR spectroscopy is one of the most accurate and advanced analytical technologies. Its power of investigating complex mixture *in situ* is well known. However, its application to the field of metabolite-interactomics is limited. We explored the potential of high-resolution NMR relaxometry to investigate typical biological samples of metabolomics.

## 2 Human blood sample

We have carried our investigation on human blood plasma and serum. It is a light-yellow aqueous solution obtained from centrifuged whole blood. It contains proteins (mainly serum albumin, globulin and lipoproteins, note that plasma also contains coagulation proteins), peptides, metabolites (such as lipids and amino acids) and electrolytes. Blood serum is very similar to blood plasma (Fig. III. 6), except that there are no anticlotting factors introduced



during the preparation. We carried preliminary measurements on blood plasma, a second series of samples of blood serum was used to carry a more systematic investigation.



*Figure III.6 Schematic presentation of blood serum and plasma*

We choose to work with human blood samples for three reasons: (i) blood plasma and serum contain biomarkers with great interests in clinic. It is also popular subjects in metabolomics studies, with rich databank. (ii) Unlike urine, blood sample is self-buffered and has more stable composition, making long-durance-experiment possible. (iii) Blood sample is abundant in protein, therefore greater chance of discovering of metabolite-protein interactions. *Fig. III.7* summarizes size and concentration distribution of all proteins that are commonly found in blood plasma and serum<sup>103-111</sup>. The protein concentrations are widely distributed from  $10^{-10}$  to  $10^{-4}$  M, and most of which are on the level of  $10^{-6}$  M or above. The sizes of the proteins cover from several hundred Dalton to thousands of kDa, corresponding to correlation time from ns to ms.

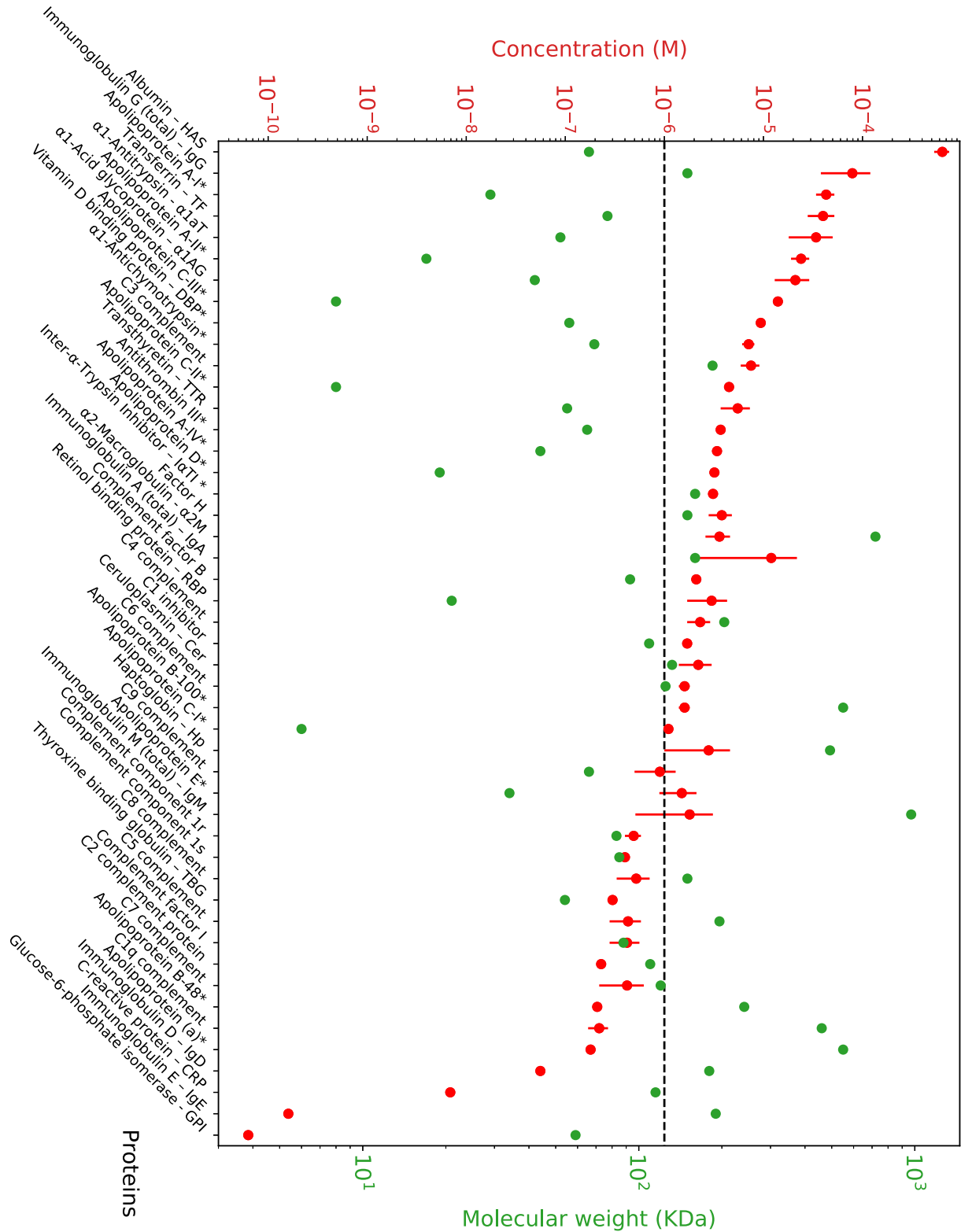


Figure III.7 Concentrations (red) and molecular size (green) of common proteins in human blood serum. Concentrations of asterisked proteins are from blood plasma. Figure is prepared by Simone Pisano.

Experiments were carried out with human blood plasma or serum provided by our collaborators from CERM in Florence. They were collected, prepared and stored in  $-80^{\circ}\text{C}$  in standard metabolomic protocols. In case where samples were diluted, the used buffer was the standard phosphate buffer (pH 7.4) including 70 mM of  $\text{Na}_2\text{HPO}_4$ , 0.025% of  $\text{NaN}_3$  and 10% of  $\text{D}_2\text{O}$ . Data was collected on these blood samples at 298 K using Stelar FFC relaxometer and the Bruker high-resolution relaxometry NMR presented in the previous chapter.

### 3 Results

#### 3.1 FFC experiment on human blood plasma

In order to evaluate the ability of relaxometry to probe the distribution of sizes of macromolecules in blood plasma, we carried our water relaxation measurements in three plasma samples by FFC relaxometry. The NMRD profiles of plasma samples (*Fig III.8*) from different patients were very similar, except for small variation at very low field magnitude range ( $<0.002\text{T}$ ). We analyzed these NMR profiles with a fitting model<sup>58, 112</sup> with three different correlation times on the order of 1  $\mu\text{s}$ , 100 ns, and 10 ns. The results of the fitted parameters are summarized in *Table III-1*. These results qualitatively agree with the size distribution shown in *Fig III.7*. These experiments were performed in CERM in Florence and the data fitting was provided by Pr. Parigi.

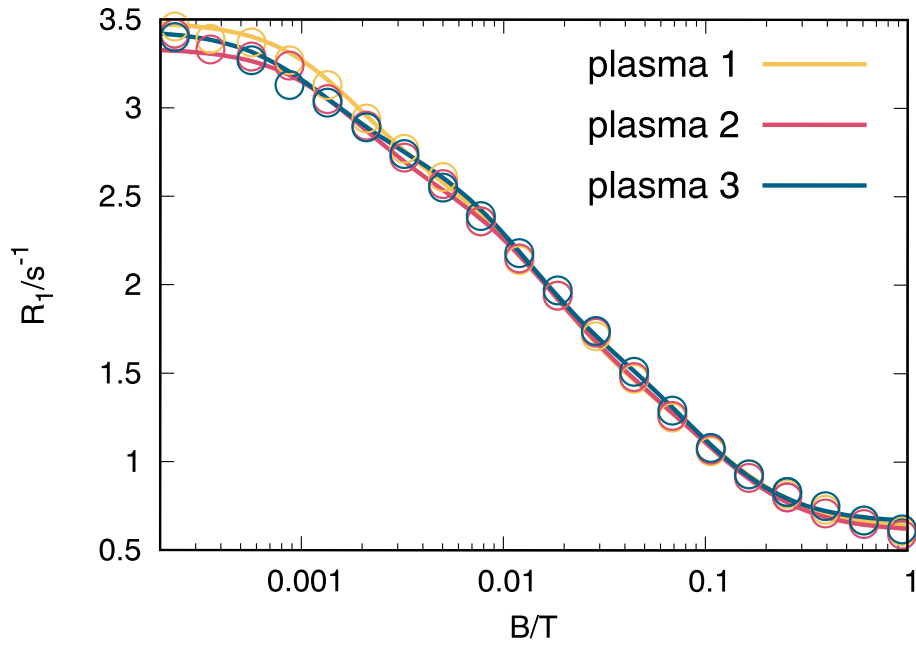


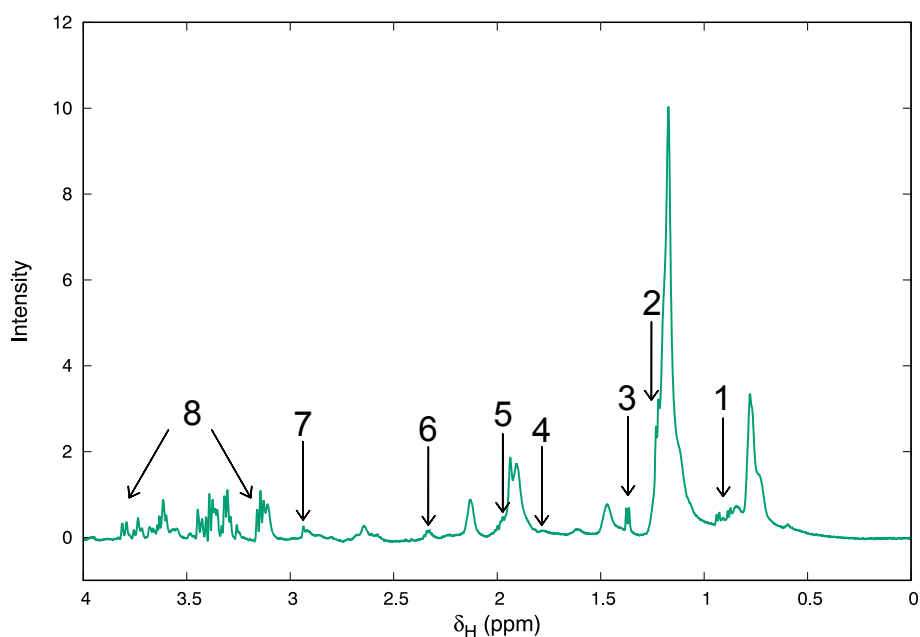
Figure III.8  $^1\text{H}$  NMRD of three plasma sample and the fit curves of the experimental results. The equation used for the fit was  $R_1 = (0.3 + A) + B \left[ c_1 \left( \frac{0.8\tau_1}{1+4\omega^2\tau_1^2} + \frac{0.2\tau_1}{1+\omega^2\tau_1^2} \right) + c_2 \left( \frac{0.8\tau_2}{1+4\omega^2\tau_2^2} + \frac{0.2\tau_2}{1+\omega^2\tau_2^2} \right) + c_3 \left( \frac{0.8\tau_3}{1+4\omega^2\tau_3^2} + \frac{0.2\tau_3}{1+\omega^2\tau_3^2} \right) \right]$ , where  $0.3+A$  is the sum of the intrinsic relaxation of water and relaxation rate arising from interactions with components other than macromolecule,  $B$  is a constant that includes effective squared order parameters and averaged squared dipolar interaction energies,  $c_1$ ,  $c_2$ , and  $c_3$  ( $c_1+c_2+c_3 = 1$ ) are the proportions of macromolecules with correlation times  $\tau_1$ ,  $\tau_2$  and  $\tau_3$ , respectively.

Table III-1 Fitted values of all parameters from the NMRD.

Parameter	Plasma 1	Plasma 2	Plasma 3
A (s <sup>-1</sup> )	0.33 ± 0.02	0.31 ± 0.02	0.36 ± 0.02
B (10 <sup>6</sup> s <sup>-2</sup> )	54 ± 5	5.4 ± 5	49 ± 4
c <sub>1</sub>	0.016 ± 0.003	0.014 ± 0.002	0.009 ± 0.002
τ <sub>1</sub> (ns)	1100 ± 100	1100 ± 100	1600 ± 200
c <sub>2</sub>	0.18 ± 0.02	0.17 ± 0.03	0.14 ± 0.02
τ <sub>2</sub> (ns)	120 ± 20	120 ± 20	160 ± 20
c <sub>3</sub>	0.80	0.84	0.85
τ <sub>3</sub> (ns)	17 ± 3	18 ± 2	22 ± 3

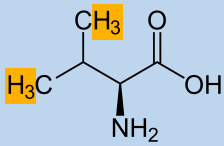
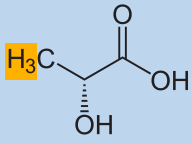
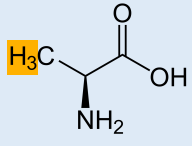
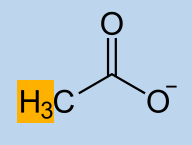
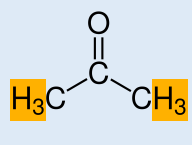
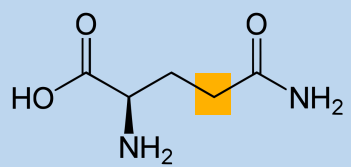
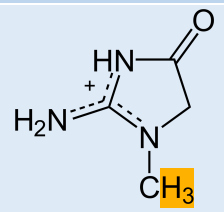
3.2 High-resolution relaxometry experiment of serum samples

High-resolution relaxometry experiments on blood serum were measured in the same way as for the model sample. Data were collected from blood serum samples collected from 3 healthy individuals. About twenty different metabolites were detected, of which at least 8 small molecules (*Fig III.9* and *Table III-2*) has relatively good signal-to-noise to obtain relaxation dispersion profiles over the entire range of magnetic fields accessible. Relaxation data of 4 common metabolites are presented in *Fig III. 10*. The relaxation rates from the 3 different patients were reasonably close one to another with few systematic differences.



*Figure III.9*  $^1\text{H}$  spectrum of blood serum. The numbered signals were 1 valine, 2 lactate, 3 alanine, 4 acetate, 5 acetone, 6 glutamines, 7 creatinine and 8 glucose.

Table III-2 Signal used for NMRD analysis of some main metabolites in serum

Number	Molecule	Group	Multiplet	Chemical shift (ppm)
1	Valine		d	0.89
			d	0.95
2	Lactate		d	1.24
3	Alanine		d	1.38
4	Acetate		s	1.83
5	Acetone		s	1.95
6	Glutamine		q	2.37
7	Creatinine		s	2.94
8	Glucose	See details in the session for glucose	m	3-4

The NMRD profile of lactate showed signs of interaction with macromolecules in serum (Fig. IV. 10 A). The relaxation profiles produced similar results between serum samples. Importantly, we can observe a plateau for relaxation rates at low fields, indicating that a full dispersion was obtained. Thus, the analysis of the NMRD profile of lactate provided the overall

rotation correlation time of the complex, that is, an estimation of the size of the complex. The overall tumbling time of the free lactate was  $50 \pm 10$  ps, and that of the complex was  $33.2 \pm 2.3$  ns, corresponding to a molecular weight around 65 kDa. The interaction of lactate with serum albumin has been reported previously<sup>71, 113-114</sup>. The rotational correlation fitted from our NMRD was in good agreement with the size of HSA (66.5 kDa). The measured relaxation rates of protons in metabolites were very similar for three different serum samples, which reflect similar compositions of these serum in healthy subjects, as expected.

The NMRD profile of creatinine also indicates evidence of binding (*Fig. IV. 10 C*) with macromolecules. By contrast to the case of lactate, we can observe some systematic differences between serum 3 and the other two serums in the NMRD profile of creatinine. Unfortunately, the data could not be interpreted quantitatively due to low precision and only partial determination of the NMRD profile (the maximum relaxation rate at low field is not reached). Nevertheless, it can be concluded that creatinine is probably interacting with large macromolecules in serum, larger than serum albumin, such as macroglobulin, C-reactive proteins or lipoproteins. In order to evaluate the size of the complex, it would be necessary to reach lower magnetic fields.

Glutamine and alanine (*Fig. IV. 10 B and D*) had relatively flat NMRD profile. They did not show significant signs of binding. The analysis of the NMRD profiles estimated the correlation time for tumbling  $\tau_c$  around 0.08-0.10 ns. These results were close to our expectation for small molecules in serum, and in good agreement with the literature<sup>115</sup>.

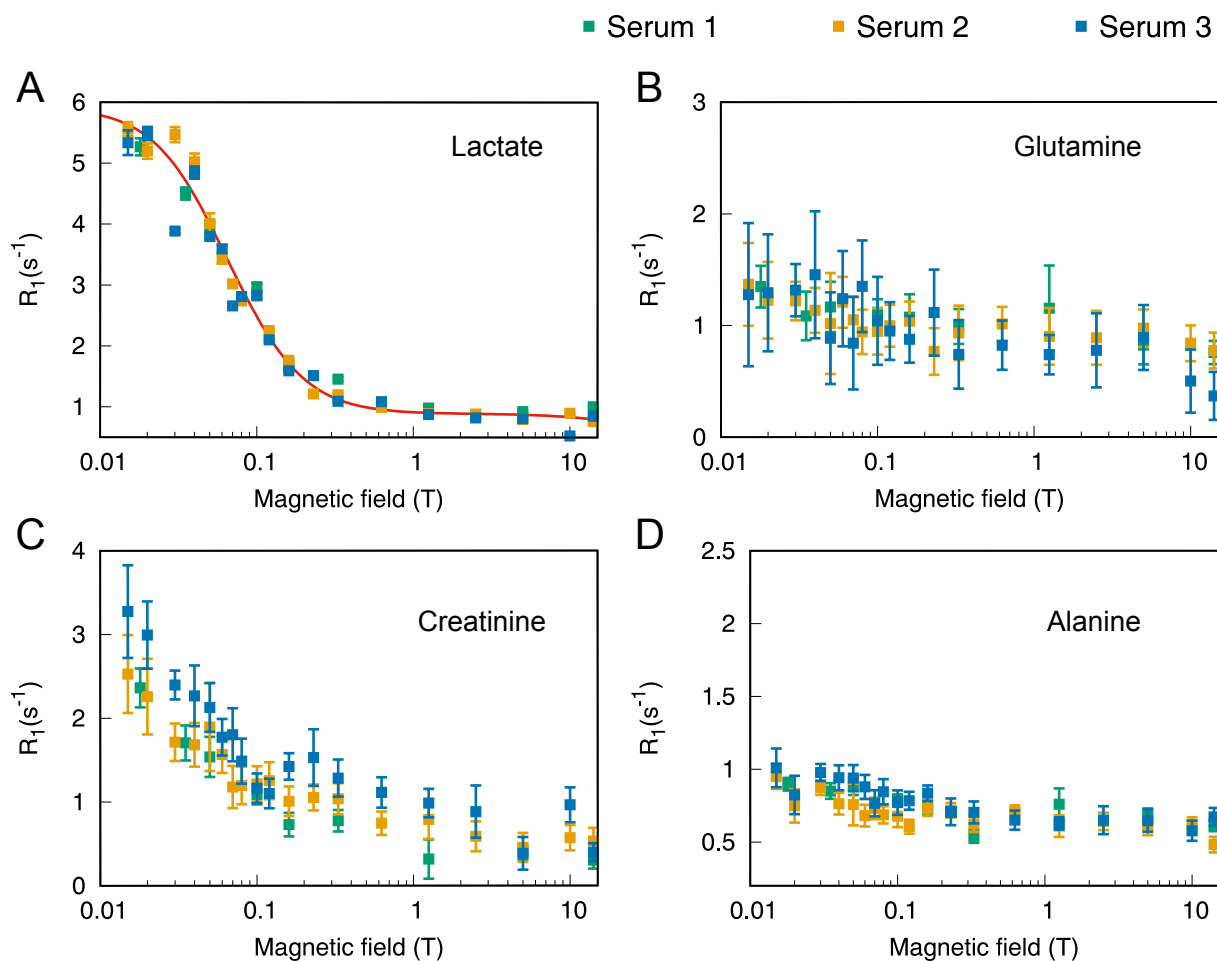


Figure III.10 NMRD of metabolites collected from three individuals, presented with dots in different colors. The profiles of lactate (A) and creatinine (C) showed dependence on the magnetic field, e.g. signs of interactions with macromolecules, while the profiles of glutamine (B) and alanine (D) did not reveal any significant effect of binding to macromolecules in the explored range of magnitude of magnetic field.

### 3.3 Glucose

Glucose has six different observable protons and two anomers ( $\alpha$  and  $\beta$ ) in solution. Their signals are mainly located between 3 and 4 ppm. The relaxation profiles of most of these signals vary with the magnetic field between 0.6 and 1.6  $s^{-1}$  as H2 and H5 shown in Fig. III. 11. This could be interpreted as relaxation dispersion due to interactions with macromolecules. The NMRD profile of H6 is an exception. The dispersion is relatively flat, with relaxation rate values around 1.6  $s^{-1}$  with relaxation rates possibly slightly lower at the lowest magnetic fields. Two hypotheses could explain this observation. The first hypothesis is the interaction with a macromolecule, resulting in the increase of the relaxation rates at low field for ring protons. If the C5-C6 bond were allowed to rotate freely in the complex, the effect of binding would be much attenuated for H6 protons. Yet, in this case, the relaxation rate of H6 protons would



slightly increase at the lowest fields. The second hypothesis is the effect of strong couplings between these protons at low magnetic fields. Strong scalar couplings would lead to coherent transfer of longitudinal polarization at low field, efficiently mixing proton longitudinal polarizations throughout their decays. One would then observe convergence of relaxation rates to a single value at low field, with an increase of the relaxation rates of the protons on the ring and a decrease of the relaxation rates of H6. The fact that all protons share similar effective relaxation rates at fields lower than 0.05 T supports this hypothesis. We have measured the relaxation rates of glucose in a solution without any macromolecule (data not shown) and a similar effect was observed, which definitely supports the latter hypothesis.

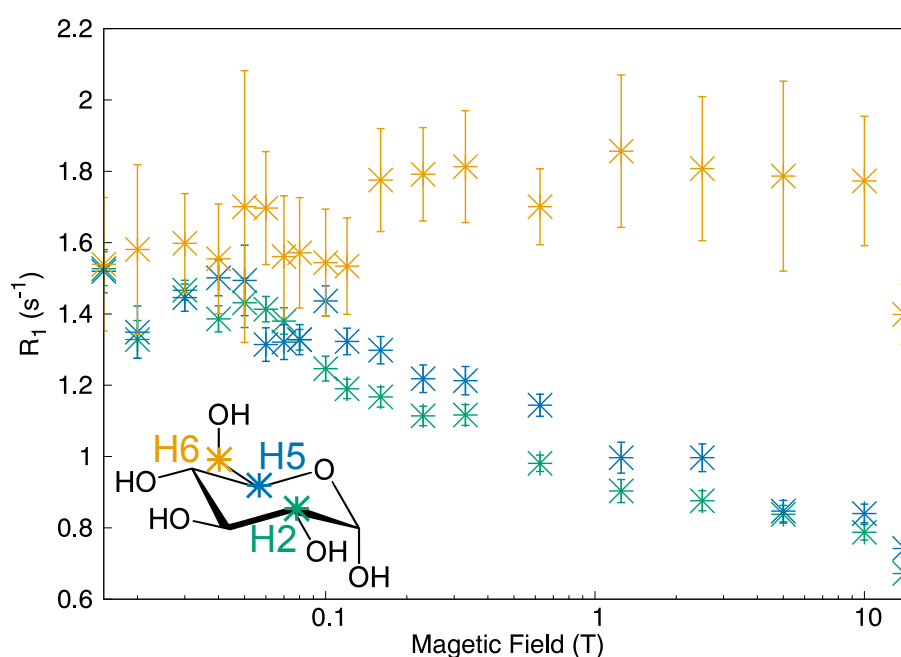


Figure III.11 NMRD of different protons of glucose in serum sample. The carbon atoms bound to H2, H5, and H6 are labeled with different colors and the NMRD profiles of these protons are presented in accordant colors.

### 3.4 Binding competition

As aforementioned, TSP is a molecule used as chemical shift reference in metabolomics studies and we showed its binding to serum albumin. In analogous fashion, lactate is also binding to a macromolecule of similar size of human serum albumin (HSA), we therefore decided to investigate if the introduction of TSP could affect MPIs between lactate and HSA. We prepared two blood serum samples with and without TSP (30 mM), respectively. Comparison of lactate and creatinine NMRD profiles are shown in Fig. III. 12A and 12B. In presence of TSP, relaxation rates of lactate were much lower compared to the case without TSP,

but the inflection point remained at the same field value. The  $\tau_c$  of the complex stayed around 33 ns. Therefore, lactate is in competition with TSP in serum, which suggests that they bind to the same hydrophobic pockets of HSA (and possibly other proteins) or act as each other's allosteric regulator. Regarding creatinine (Fig. III 8B), NMRD profiles showed that there is no binding competition with TSP, which is in accordance to what reported by *R. Barrilero et al.*<sup>65</sup>

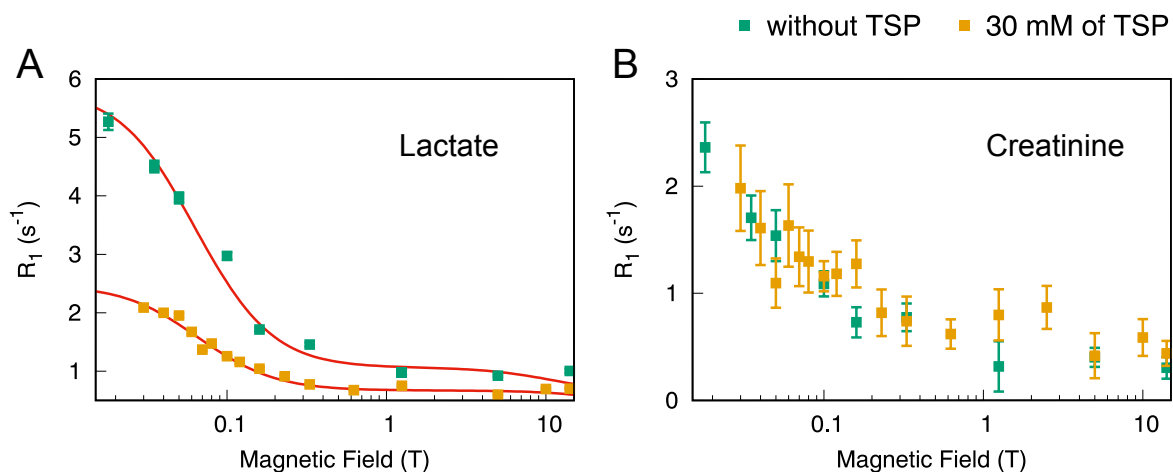


Figure III.12 NMRD lactate and creatine in the same serum sample with and without the presence of TSP.

#### 4 Discussion and conclusion

We developed a new approach to investigate MPIs with high-resolution NMR relaxometry. It is based on the measurement and analysis of relaxation rates over several orders of magnitude of magnetic field (from 0.015 to 14 T) paired with high-resolution detection. This method benefits from several advantages as compared to others that are currently in use. First, there is no separation step during sample preparation. Once the serum is obtained from blood by traditional procedure, the sample is directly sealed into a sample tube with minimum manipulation. Second, relaxometry is able to probe weak transient interactions, which are extremely difficult to detect by existing methods. Third, the analysis is both qualitative and quantitative. Currently, the range of magnetic field that we can probe allows to quantify the rotational correlation time of a complex as slow as ~50 to 100 nanoseconds. The determination of this correlation time provides an estimate of the size of complexes, hence of the size of proteins which interact with a given metabolite. Given the limited number of proteins concentrated enough to be detected by our method, a correlation time with a precision of about 10% restricts the number of potential proteins to 1 to 8 (*Fig. III. 13A*). We were also able to

identify competitive binding between small molecules. This could prove to be a useful approach to perform fragment-based drug design directly in a complex but most relevant environment: human blood.

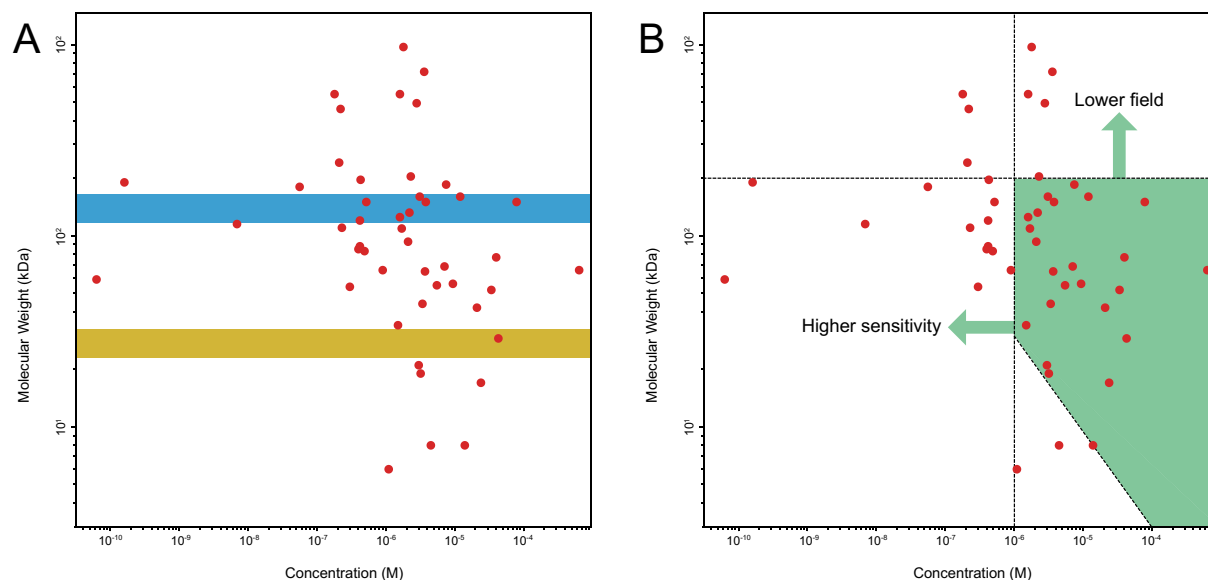


Figure III.13 Molecular size dispersion on function of concentrations of common proteins in human blood serum, same as in Fig. III. 7. A With 10% precision, the number of proteins corresponding to a fitted correlation time are from 1 (yellow area) to 8 (salmon area). B The proteins inside the green area are potentially detectable by the current high-resolution relaxometer. A higher sensitivity will allow detecting metabolites and proteins of lower concentration. Lower field range can help to reveal the full NMRD profile of larger complex.

The main restrictions of the technique are the sensitivity of the measurement and the range of accessible low fields (Fig. III. 13B). Metabolites are present in blood with a very broad range of concentrations. The samples studied here are human blood serum, in which the concentration of metabolites and macromolecules are often very low as compared to cell lysates. In routine metabolomics studies by NMR, micromolar level metabolites are detected. Here, the sensitivity of our system is very low (about 10% of the sensitivity of a normal room-temperature inverse probe), *i.e.* around 100  $\mu\text{M}$  for metabolites and 1  $\mu\text{M}$  for the complex if we consider 1% binding population. In addition, significant losses of polarization arise during the transfers between high and low magnetic fields, especially in the case of transient binding to macromolecules. It is difficult to extract precise parameters for most of the small molecules in a sample of blood serum. Improvements on sample volume and a more sensitive probe could dramatically improve the signal to noise, probe metabolites with lower concentrations and shorten the experiment time. Finally, as shown by FFC relaxometry and the literature, blood

contains proteins with large sizes that would lead to dispersion at magnetic fields lower than those that can be reached with the current sample shuttle. We are currently working on all these limitations with our collaborators from Bruker.

In spite of these limitations, this work can be considered as a proof of concept that demonstrates the enormous potential of high-resolution NMR relaxometry, which can surely become a crucial analytical tool for complex fluids mixtures such as human blood serum.



## Chapter IV Probing the intrinsic disorder of the human protein XLF

NMR is one of the most important techniques used in structural biology. Its capacity of observing macromolecules in solution in near physiological condition gives it many advantages compared to X-ray crystallography and cryogenic electron microscopy. Especially, NMR has the unique power of investigating disordered protein at atomic resolution.

In this chapter, we attempt to understand functional dynamics of a factor involved in DNA repairing mechanism, the human protein XRCC4-like factor (XLF). It is predicted that the C-terminal region of XLF is intrinsically disordered. In this chapter, we will first introduce XLF and present our work, producing isotopically labeled samples and performing a first NMR analysis of the disordered region of the protein.

## 1 General introduction

### 1.1 Protein structure and dynamics

Proteins are polymers of amino acids linked by peptide bonds. The organization of the polypeptide backbones and sidechains gives protein its unique conformation in space and function. For structured proteins, there are four orders of organization:

1. Natural proteins are constructed from 20 amino acids. The primary structure of a protein is the unique order of the consisting amino acids.
2. The secondary structure corresponds to regular conformation of the peptide backbone stabilized by weak interactions (e.g. hydrogen bonds) between atoms of the backbone. The most common secondary structures are the  $\alpha$  helix and the  $\beta$  sheet.
3. Tertiary structure is the comprehensive three-dimensional structure of a single polypeptide chain. It is stabilized by hydrogen bonds, hydrophobic interactions, ionic bond and disulfide bridge between sidechain atoms.
4. In case of proteins formed with several polypeptide chains, there exists quaternary structures. The same types of interaction are implicated in binding between subunits than in the tertiary structure.

A specific spatial conformation is essential to protein functions, making the investigation of the structure function relationship in proteins the foundation of the field of structural biology. However, proteins are dynamic ensembles of conformations. The growing knowledge on intrinsically disordered proteins (IDPs) or regions (IDRs) is showing that long regions of proteins can be functional in the absence of a stable 3D structure. As one of the main analytical tools of structure biology, NMR has the unique power of providing information about conformational properties and dynamics at atomic level in disordered proteins, compared to crystallography and cryo-electron microscopy. We intended to get a better understanding of protein dynamics and functions through example of the protein XLF, which is predicted to have a C-terminal IDR.

### 1.2 Intrinsically Disordered Proteins

An IDP or an IDR is a protein or a protein segment lacking a stable 3D structure folded around a hydrophobic core. IDPs and IDRs challenged the dominant “structure-function” relationship dogma<sup>116-117</sup>. The classical concept implies that protein sequence determines the structure and further defines the function. For a long time, IDRs were considered simply as

linkers between structured core domains. However, despite their self-insufficient folding, IDPs and IDRs are proved to be functional and play diverse roles in living systems.

IDPs and IDRs are especially large abundant in eukaryotes as compared to procaryotes.<sup>118-119</sup> It is predicted that about 40% of human protein-coding genes contain disordered segments of 30 residues or more in length.<sup>120</sup> IDPs and IDRs are enriched in hydrophilic and charged residues: Ala, Arg, Gly, Gln, Ser, Pro, Glu and Lys, called “disorder-promoting” amino acids.<sup>121</sup> Missing of hydrophobic core, the proteins lack tertiary structure and are more mobile and dynamic. Unlike folded proteins, IDRs don't have a clear global minimum for conformational energy, leading to significant random fluctuations of atomic position and backbone Ramachandran angles over time<sup>118</sup>. The high plasticity gives the protein in particular the adaptability to bind a wide range of partners.

The known biological functions of IDPs and IDRs have been reviewed.<sup>116-118, 122</sup> They have great ability of binding or interacting with other macromolecules such as DNA, protein or complex of molecules. Many IDPs and IDRs can undergo disorder-to-order transition with the presence of the target molecules. These proteins often have central role as effectors or assemblers.

The discovery of functional significance of IDPs in the last thirty years was led by advanced characterization technologies, especially by NMR. X-ray crystallography is inappropriate for IDRs for multiple reasons (1) crystallizing proteins with long disordered regions is very challenging, (2) in most cases, disordered regions mobile in the crystal do not provide any electron density and (3) if a fragment on an IDR is rigid in the crystal and provides electron density, this conformation will not be representative of the conformational ensemble present in native conditions. While NMR is able to follow dynamics and conformation properties of IDPs at the atomic level, which is the key to understanding the physics and chemistry underlying their function.

## 2 Biological role of XLF in DNA repair

### 2.1 DNA damages and repair

Stability of DNA is essential to living organisms, for that it is the repository of genetic information. However, it is constantly exposed to damages due to environmental and internal factors, and results with different lesions such as base pair mismatches and backbone breaks. If unrepaired, the damages may lead to mutation, apoptosis or cancer. Many ingenious repair mechanisms have been evolved to repair or tolerate diverse types of damages at different



cellular stages. *Fig.IV.1* shows examples of damaging resources, lesion types and repair mechanisms.

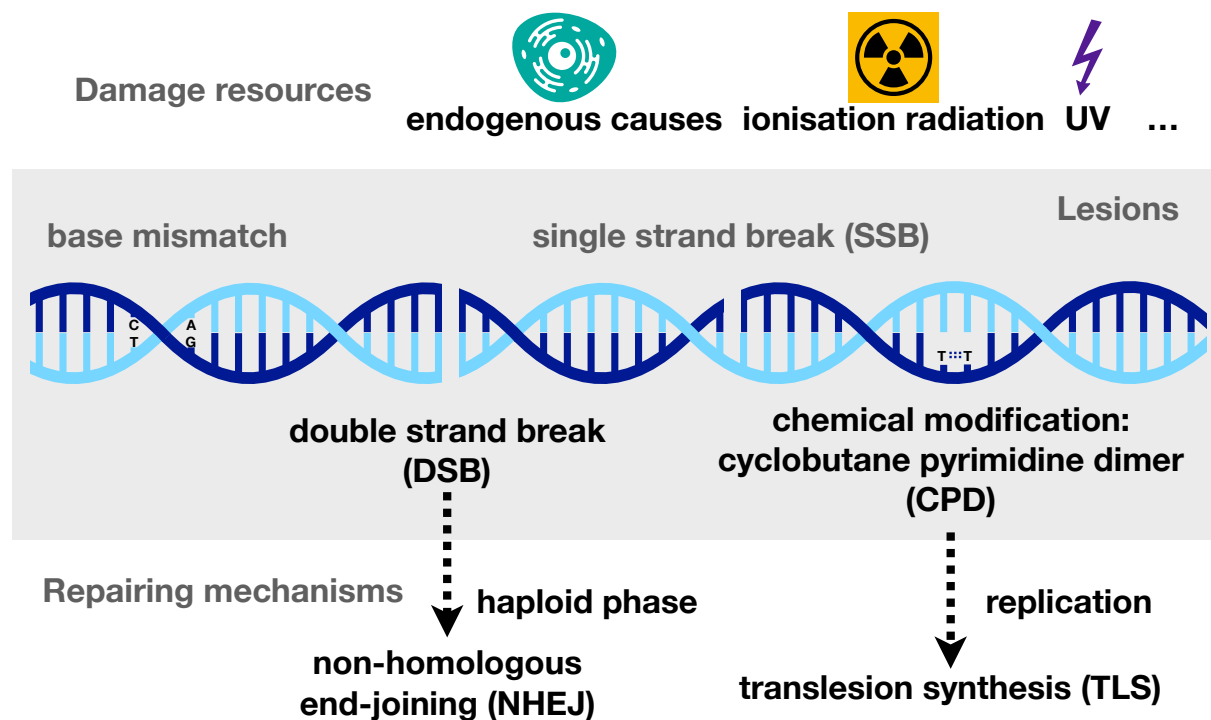


Figure IV.1 Examples of DNA damages and repair mechanisms<sup>123</sup>.

## 2.2 Non-homologous end-joining

DNA Double-strand breaks (DSB) are the most toxic of DNA lesions. It may be consequence from repair failure of single-strand breaks, reactive oxygen species, ionizing radiation. DNA double strand breaks also occur under controlled conditions in physiological processes such as V(D)J recombination and immunoglobulin heavy chain class switch recombination, as well as during DNA modifications by the CRISPR-Cas9 system<sup>124</sup>. These resources cause different types of DSBs. Accidental DSBs often result in incompatible ends of various chemical and topological patterns on the DNA molecules<sup>125</sup> (*Fig.IV.2*). Two repair mechanisms exist in eukaryotes to fix DSBs. The first is homology-directed repair: homologous recombination, which is the most common form, based on the use of the undamaged DNA template of the sister chromatid. It is exclusive during the period where homologous sister chromatids are available. The second mechanism, non-homologous end-joining (NHEJ), is the

pathway of repairing DSBs without homology template, which may occur during all phases of the cell cycle. The choice between two pathways is balanced by complex regulatory systems.

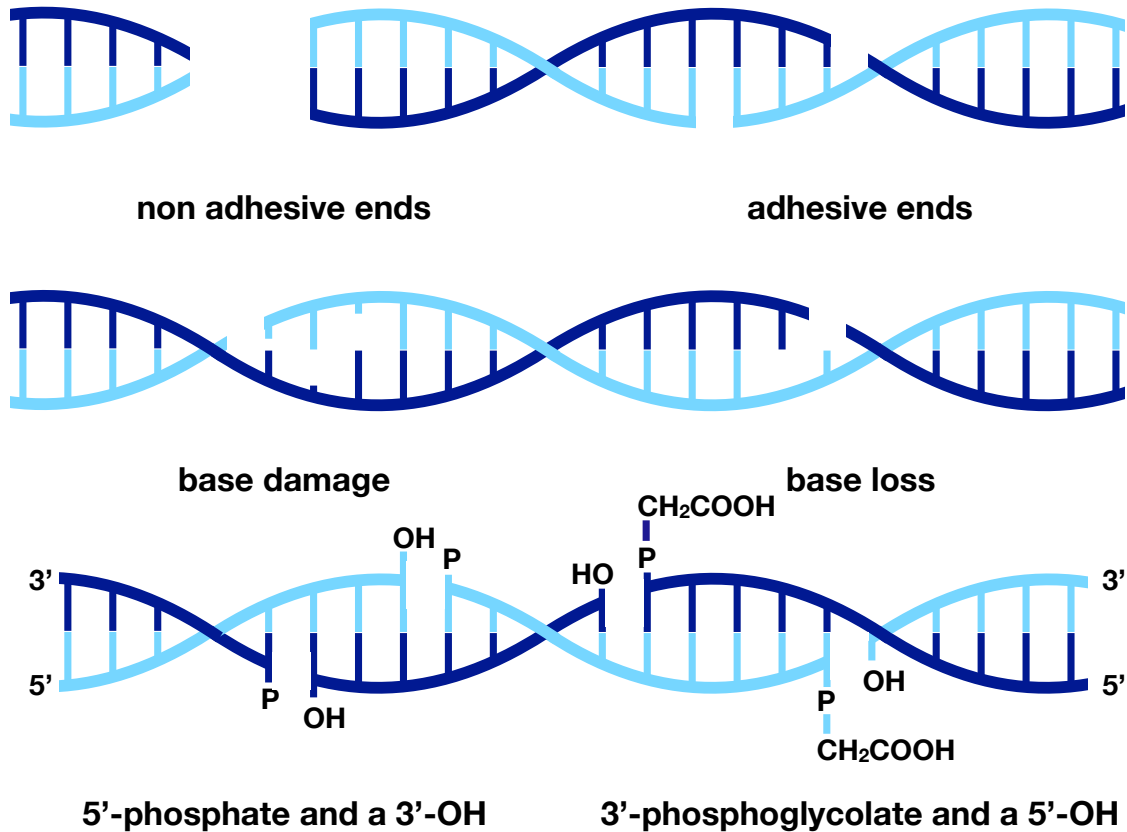


Figure IV.2 Different types of DSBs and properties of DNA molecules extremities<sup>125</sup>.

Flexibility is the key word to characterize NHEJ apparatus.

On the functional level, this pathway has convoluted machinery with high adaptability to a diverse range of DSB DNA substrates<sup>126</sup> as shown in *Fig.IV.2*. In human cells, NHEJ involves a large protein assembly. Depending on the DNA substrates, the endings are resected, prolonged or directly ligated. The process initiates with the protection of damaged DNA ends by the Ku70-Ku80 heterodimer as toolbelt protein. It is followed by the recruitment of DNA-PKcs and the nuclease Artemis. Artemis-DNA-PKcs complex is then phosphorylated by the activated kinase function of DNA-PKcs. Next, DNA polymerase  $\mu$  or  $\lambda$  are involved to fill in new DNA. In the end, ligase IV-X-ray repair cross-complementing protein (XRCC4)-XRCC-like factors (XLF) complex finish the final joining of the two DNA ends.<sup>126-128</sup> A generalized procedure is summarized graphically in *Fig.IV.3*. This sequential description is challenged by

several important discoveries, such as the role of XRCC4 and XLF in bridging broken fragments of DNA<sup>129</sup>.

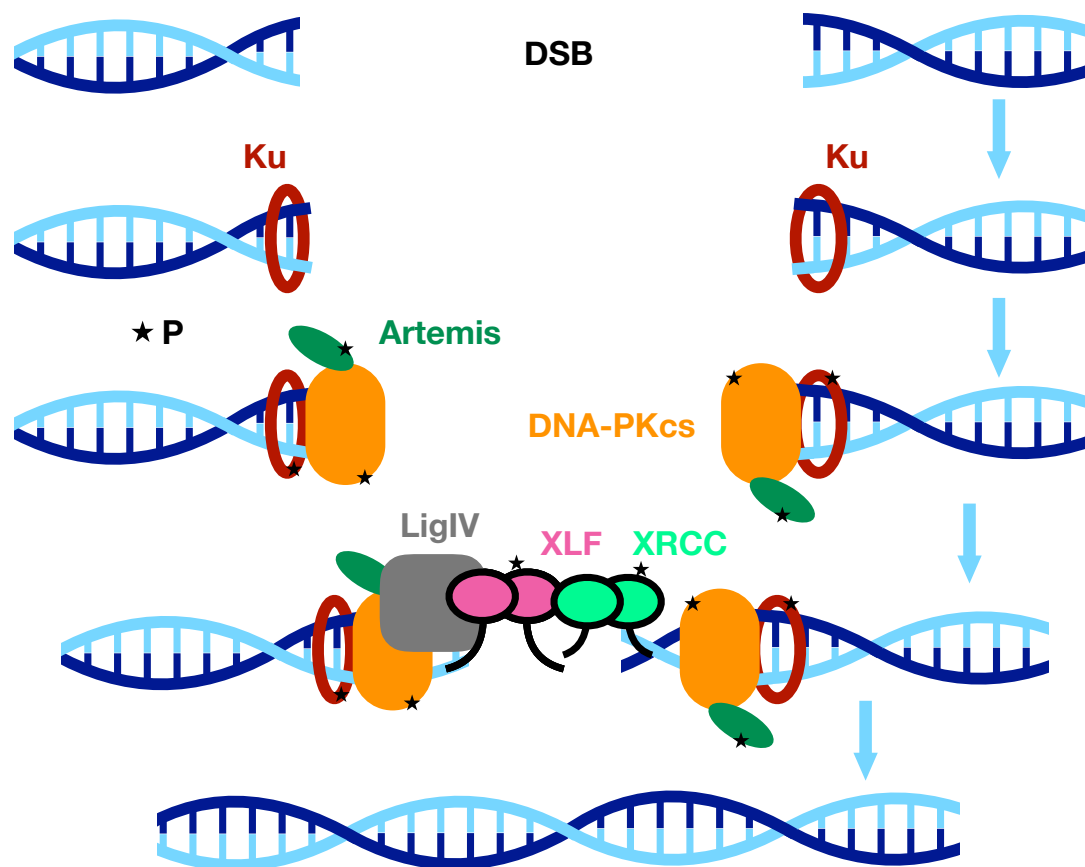


Figure IV.3 Schematic presentation of human NHEJ repairing process and the involved main factors.

On the structural level, disordered regions are encountered in most actors of NHEJ.<sup>130</sup> Plastic regions of NHEJ factors are listed in *Table.IV-1*. Their conformational suppleness has a deep link to the function compatibility. From the table, it is evident that the potential functions of the flexible regions are particularly centralized to interactions and bindings between different elements involved in the pathway. have long disordered regions. Yet, the exact role of disorder is an unexplored aspect of the mechanism.

Table IV-1 Summary of flexible regions presented in main NHEJ factors

Protein	Flexible region	Potentially related function
Ku70	N-terminus	Interaction with histones <sup>130</sup>
	Linker between core and SAP	Binding to DNA <sup>131-132</sup>
Ku80	Linker between core and CTR	Facilitation protein-protein interaction <sup>133</sup>
	C-terminus	Binding to DNA-PKcs <sup>134-135</sup>
DNA-PKcs	Multiple disordered regions	Regulation of the activity by phosphorylation <sup>136</sup> , DNA binding <sup>137</sup>
Artemis	C-terminus	Interaction with DNA-PKcs <sup>138</sup> and Ligase-IV <sup>139</sup>
XRCC4	C-terminus	Interaction with Ligase-IV, DNA binding <sup>140</sup>
XLF	C-terminus	Interaction with Ku <sup>141</sup> , DNA binding <sup>142</sup>

### 2.3 XLF

In this project we focus on one of the NHEJ factor, human XLF. It was discovered in 2006<sup>143-144</sup>, also known under the name of Cernunnos. It has been reported as a mutation site in human immunodeficiency with developmental abnormalities<sup>144</sup>. The ligase IV-XRCC4-XLF complex intervenes on the final step of the NHEJ repairing process, and is in charge of the final joining of DSB ends. XLF enhances the affinity of ligase IV-XRCC4 to noncompatible DNA ends<sup>145</sup> and interact directly with XRCC4 at the level of their N-terminuses<sup>143</sup>. Yet, the underlying mechanisms of the function of are not well understood. In particular, little is known on the conformational properties of its disordered C-terminus.

Human XLF has 299 residues. Crystallographic structures of the structured domain of XLF (residue 1-223) alone and in different complexes have been reported. XLF forms a stable homodimer via the coiled-coil region (PDB 2QM4<sup>146</sup> and 2R9A<sup>147</sup>). XLF shares structural similarity with XRCC4, despite their low sequence homology. The structured part of XLF has

a globular N-terminal head domain and  $\alpha$ -helical tail in crystal. Structured domains of XLF-XRCC4 complex has also been resolved using crystallography under different conditions (PDB 3SR2<sup>148</sup>, 3Q4F<sup>149</sup>, 3RWR<sup>142</sup>, 3W03<sup>150</sup>). Similar observation was made in all structures that XLF and XRCC4 homodimers extended to filament by binding their head domains. The disordered residues were not included in the protein constructions used in these studies, to reduce the flexibility of the protein and facilitate crystallization. Yet, experimental results indicate clearly that the unsolved disordered region of XLF (residue 224-299) contains crucial part of the DNA binding interface<sup>142</sup>. The structure of a complex of Ku-DNA and a short peptide consisting of the last (19 or 13) C-terminal residues of XLF has been solved recently<sup>141</sup>. Phosphomimetic studies have shown that phosphorylation of the C-terminal disordered regions of XLF and XRCC4 accelerates the dissociation of XLF-XRCC4 filaments from DNA.<sup>129</sup> Without the C-terminus of XLF, the binding ability to DNA of XRCC4-XLF complex decrease dramatically.<sup>142</sup>

For this reason, we aim to get better understanding to the molecular mechanisms involving the disordered C-terminal region of XLF, and its interaction with DNAs. High resolution NMR is a tool of choice to investigate this disordered region in the context of the full length human XLF.

### 3 Sample preparation

#### 3.1 General strategy

The construct of the fusion protein, corresponding to full length (residue 1-299) *Homo sapiens* XLF and a polyhistidine-tag on the C-terminus (*Fig.IV.4*). The preparation start with transformation of pET-duet1 plasmid containing this construct into competent *E. coli* BL21 Gold™(DE3)pLysS. The fusion protein was expressed in the host *E. coli* in <sup>13</sup>C and <sup>15</sup>N labeled M9 medium. The protein is purified with a Histrap column followed by and gel filtration.

A



B

10	20	30	40	50	60
MEELEQGLLM	QPWAWLQLAE	NSLLAKVFIT	KQGYALLVSD	LQQVWHEQVD	TSVVVSQRAKE
70	80	90	100	110	120
LNKRLTAPPA	AFLCHLDNLL	RPLLKDAHP	SEATFSCDCV	ADALILRVRS	ELSGLPFYWN
130	140	150	160	170	180
FHCMLASPSL	VSQHLIRPLM	GMSLALQCQV	RELATLLHMK	DLEIQDYQES	GATLIRDRLK
190	200	210	220	230	240
TEPFEENSFL	EQFMIEKLPE	ACSIGDGKPF	VMNLQDLYMA	VTTQEVQVGQ	KHQGAGDPHT
250	260	270	280	290	300
SNSASLQIGID	SQCVNQPEQL	VSSAPTLSAP	EKESTGTSGP	LQRPQLSKVK	RKKPRGLFSH

**HHHHH**

Figure IV.4 Construct of recombinant human XLF (A) and its sequence (B). The globular domain, IDR and the tag polyhistidine are in grey, blue and red, respectively.

### 3.2 Transformation

*E. coli* BL21 Gold™(DE3)pLysS competent cells were transformed with the pET-duet1 plasmid containing the recombinant XLF coding sequence. 1 µL of plasmid solution was mixed with 100 µL of cells in a prechilled 14-ml BD Falcon polypropylene round-bottom tube. The reaction was incubated on ice for 20 min, then heat shocked in a 42°C-water bath for 25 s and immediately replaced on ice. 800 µL of preheated LB medium was added to the transformed cells after 2 min, and were incubated at 225 rpm, 37°C for 1 hour. 100 µL of the reaction were spread onto a LB agar plate containing 34 µg/mL chloramphenicol and 100 µg/mL ampicillin for overnight incubation. The same concentrations of antibiotics were used all along the expression process of the recombinant XLF. Single-colonies were amplified in LB at 37°C the next day. Stocks of the cells were prepared by adding 15% glycerol, and were aliquoted, snap frozen and kept at -80°C.

### 3.3 Expression

The expression protocol of <sup>13</sup>C and <sup>15</sup>N labeled XLF was based on the rapid and efficient method developed by Marley *et al.*<sup>151</sup> Schematic summary of the protocol is presented in Fig.IV.5.

1. This method started with reliving 250 µL of cell stock in 5 mL LB at 37°C under agitation at 225 rpm for about 4 hours.

2. Enough quantity (around 1 mL) of this culture was transferred into 50 mL of fresh LB to start an overnight preculture from OD<sub>600</sub> 0.02.
3. The next morning, the preculture is used to infect 2 L of fresh LB to start from OD<sub>600</sub> 0.05.
4. When the density of culture reached OD<sub>600</sub> 0.6, the bacteria was centrifuged at 3000 g for 15 min at room temperature. The cell pellets were resuspended in 500 mL of M9 medium (without glucose and ammonium).
5. After 40 min of incubation, <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C labeled glucose were added to the medium.
6. After another 40 min of incubation, the temperature of the incubator was decreased to 18°C, and 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) was introduced to the culture to induce the expression of the culture overnight.
7. The next morning, the cells were harvested by centrifuging for 15 min at 3000 g, 4°C. The pellet was resuspended in 15 mL phosphate-buffered saline (PBS) buffer (Na<sub>2</sub>HPO<sub>4</sub> 20 mM, NaCl 100 mM, pH 8), and was frozen at -80°C.

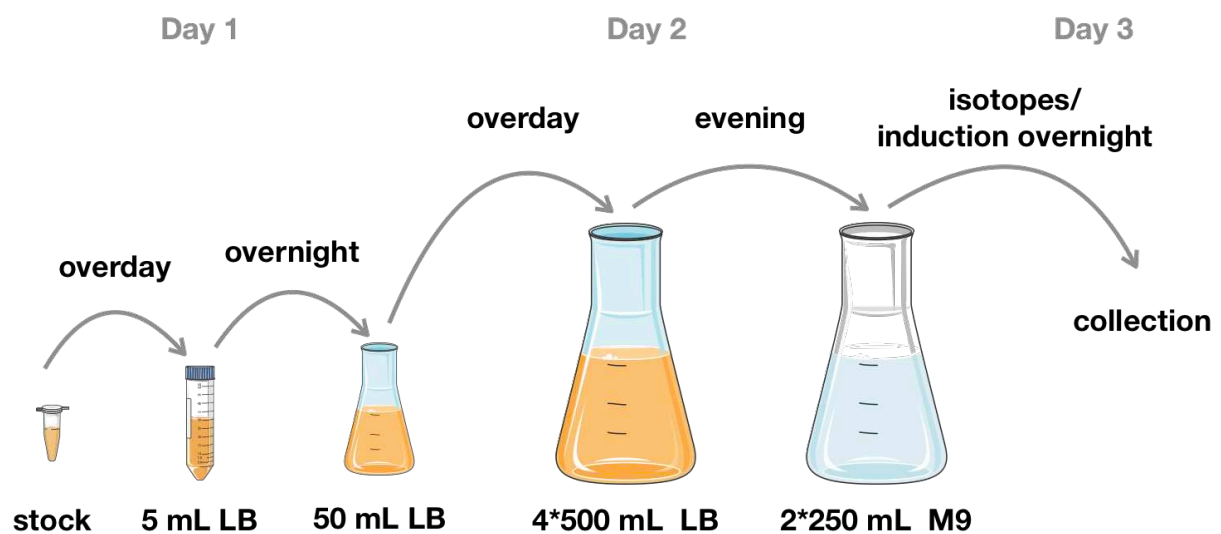


Figure IV.5 Protein expression <sup>13</sup>C/<sup>15</sup>N labelled protein in *E. coli*.

### 3.4 Purification

Two chromatography columns were used to extract XLF from the cells.

#### 1. Bacteria lysis

Into one pellet of bacterial culture, 20 mL lysis buffer (Tris 40 mM pH 8, NaCl 1.6 M, EDTA 2 mM, DTT 4 mM, imidazole 20 mM, glycerol 20%), antiprotease pepstatin A (PEP),

leupeptin (LEU), phenylmethylsulfonyl fluoride (PMSF), 50 mg lysozyme and 1 mL of Triton X100 were added. The mixture was thawed on ice for one hour, then lysed by 10 sonication cycles of 15 s interleaved with 1 min breaks. The lysate was centrifuged at 30000 g for 30 min. The supernatant was filtered with 0.2 $\mu$ m pore membrane. The sample was always kept at 4 °C or on ice to avoid protein degradation.

#### 2. Ni-NTA – His-tag affinity chromatography

A 5ml HisTrap excel GE was pre-equilibrated with buffer A (Tris 20 mM pH 8, NaCl 0.8 M, EDTA 1 mM, DTT 2 mM, imidazole 10 mM, glycerol 20%) at 2 mL/min. The clarified sample was loaded into the column at 1.5 mL/min. The column was washed by 25 mL buffer A then 25 mL 5% buffer B (same as A with 500 mM imidazole) at 2 mL/min. The target protein was eluted by 35 mL of 100% buffer B.

#### 3. Gel Filtration Superdex 75 – size exclusion chromatography

The 120 mL gel filtration (GF) column was pre-equilibrated with GF buffer (20 mM Bis-Tris pH 6.5, 150 mM KCl, 1 mM EDTA, 2 mM DTT). Fraction collected from the HisTrap column were concentrated in Amicon® tube with a 10 kDa size cut off into a sample of 2 mL, during which process the buffer was changed gradually into GF buffer. The sample is injected into the column. Using UV detection, protein fractions were collected. The fractions containing XLF were concentrated. The sample was stored at -20°C.

## 4 NMR analysis

NMR experiments were performed on an 800 MHz wide-bore Bruker Avance III HD equipped with a TXI triple-gradient room-temperature probe at 293K. A series of triple-resonance experiments were carried out: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, CC(CO)HN and H(CCO)NH for backbone and sidechain assignment. A series of <sup>15</sup>N relaxation experiments were also carried out in order to evaluate the dynamics of XLF C-terminal region: longitudinal and transverse relaxation rates as well as steady-state <sup>1</sup>H-<sup>15</sup>N nuclear Overhauser effects (NOE's) were measured, also at 800 MHz. The relaxation delays for the longitudinal relaxation rate were: 0.02, 0.20, 0.44, 0.80, 1.20 and 1.80 s; those for the transverse relaxation rate were: 0.004, 0.02, 0.04, 0.06, 0.08 and 0.1 s. The recovery and saturation delays were 12 and 8 s, respectively. The saturation on proton channel was carried out with a train of 725 echo distanced by 11 ms intervals. The final sample of XLF used for the NMR experiment was at 400  $\mu$ M, in 20 mM Bis-Tris (pH 6.5), 150 mM KCl, 1 mM EDTA, 2 mM of DTT and 5% of D<sub>2</sub>O. All experiments were processed using NMRPipe<sup>75</sup> and the peak



intensities were obtained from Sparky<sup>152</sup>. The longitudinal  $R_1$  and transverse  $R_2$  relaxation rates were obtained from a fit to single exponential decays with Python<sup>153</sup>.

## 5 Results and discussions

All non-proline residues 232 to 293 corresponding to the disordered region of XLF were assigned (*Fig.IV.6-7*). Residues 295 to 305, including the Ku binding site, were difficult to assign because of their low intensities in all spectra (see below). The folded domains of the protein were not assigned: the 60-kDa dimer are too large to be investigated without deuteration.

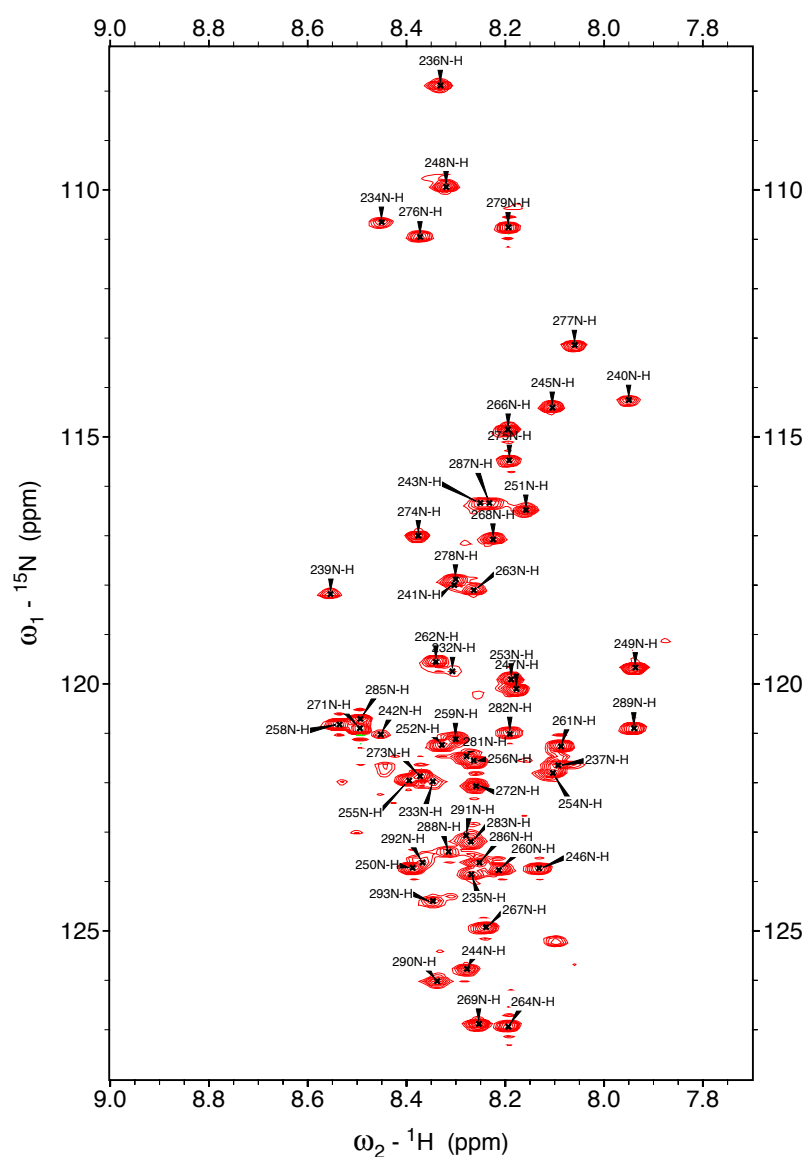


Figure IV.6  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of XLF. Assignments of IDR residues are indicated.

240                      250                      260                      270                      280  
 KHQGAGDPHT SNSASLQIGD SQCVNQPEQL VSSAPTLSAP EKESTGTSGP  
 290                      300  
 LQRPQLSKVK RKKPRGLFSH HHHHH

Figure IV.7 IDR of XLF with the His-tag at the C-terminus. Unassigned residues are shown in grey and assigned residues in green.

Based on chemical shifts of  $C\alpha$  and  $C\beta$ , secondary structure propensities (SSP)<sup>154</sup> were calculated (Fig.IV.8). It can be seen that the C-terminal IDR has limited local conformational preferences. Between residues 240 and 290, there might be short transient  $\alpha$ -helices formed.

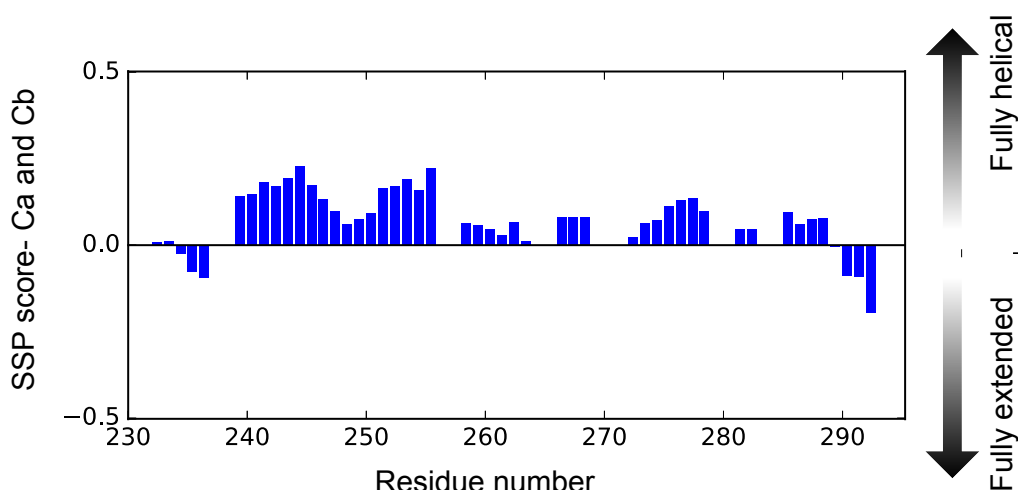


Figure IV.8 SSP of the assigned residues of XLF, based on CA and CB chemical shifts. SSP scores are 1 for an alpha helix and -1 for a beta sheet.

The NOE ratios are in good agreement with the SSP profile. NOE ratios were slightly above 0, similar to results obtained with other IDRs at 800 MHz and similar temperatures<sup>155-157</sup> showed this disordered region had little conformational restraint (Fig.IV.9). Higher NOE values near residue 230 are likely due to the fact that these residues are within one persistence length of the folded helical domain. Slight increases around residues 250 and 290 may indicate some constraint of the conformational space.

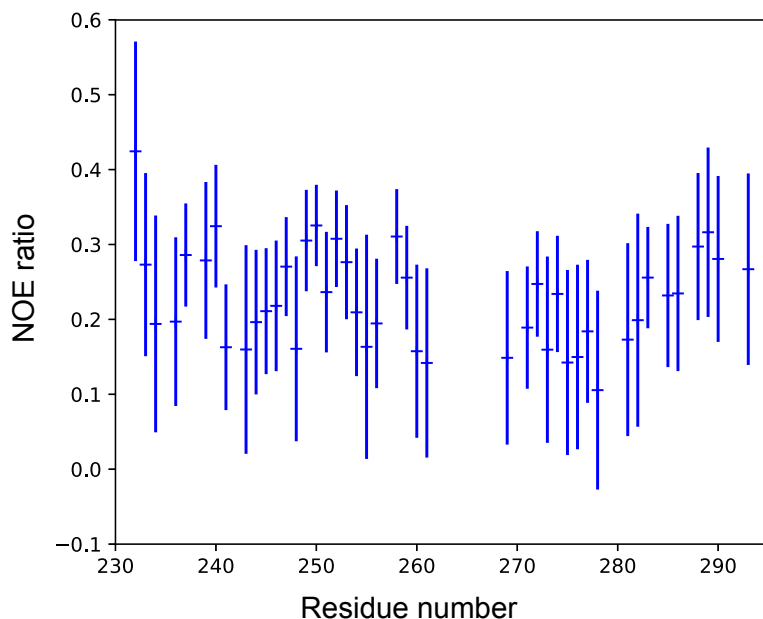


Figure IV.9  $^{15}\text{N}$ - $^1\text{H}$  Nuclear Overhauser effects at 800 MHz and 293 K for the region 232-293 of recombinant XLF.

Longitudinal,  $R_1$ , and transverse,  $R_2$ , relaxation rates were measured with 6 relaxation delays each. The relaxation decays of individual residues were of good quality.  $R_1$  values are uniform, ranging between 1.2 and 1.5  $\text{s}^{-1}$  for all residues (Fig.IV.10 A). This is somewhat often the case for disordered proteins at this field<sup>155-157</sup>.

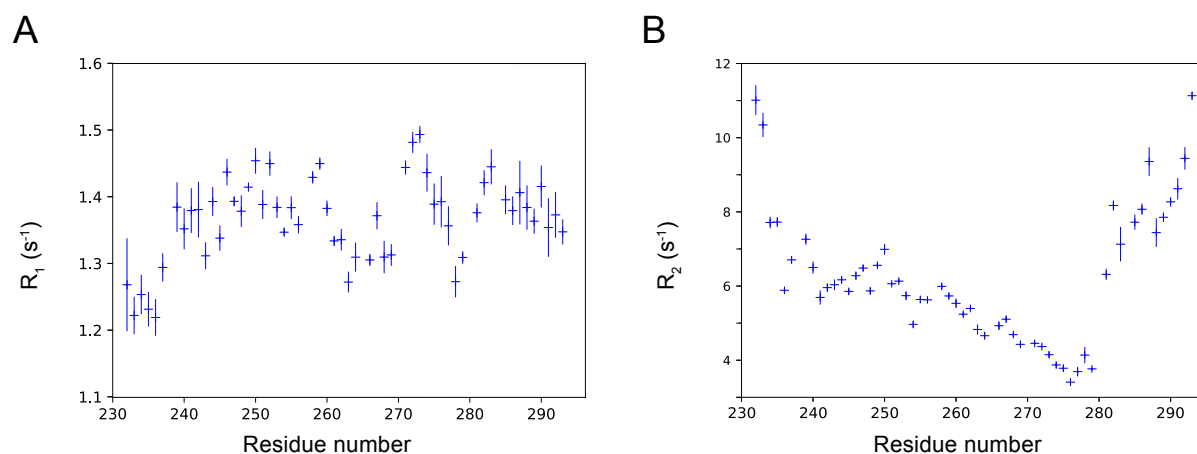


Figure IV.10 Longitudinal relaxation rates  $R_1$  profile(A) and transverse relaxation rate  $R_2$  profiles (B) of the region 232-293 of recombinant XLF, with  $R_2$ -CPMG profile (blue cross) and  $R_2$ -singleEcho profile (green cross). The relaxation delays were from 0.02 s to 1.8 s for  $R_1$  0.004 s to 0.1 s for  $R_2$ -CPMG.

Transverse relaxation rates  $R_2$  (Fig.IV.10 b) vary much more along the sequence with values between 3 and 14  $s^{-1}$  obtained from CPMG experiment. The lowest  $R_2$  values, observed in the region 235-280 with a slow decay as the residues get further from the structured domain, indicate this region is highly dynamic within full-length XLF.  $R_2$  of the C-terminal residues 280-293 are much higher and similar to those for the region 232-235 that connects the structured domain. This may imply that the C-terminus is binding transiently to the structured globular part of the protein with fast (sub-ms) exchange between free and bound forms. We cannot exclude that this effect is artefactual and caused by the positively charged His-tag. This binding and subsequent fast transverse relaxation likely explains the difficulty to obtain signals of residues beyond residue 293 in triple-resonance experiments.

With this construct, we have assigned most of the disordered C-terminus of full-length XLF. Chemical shift analysis and relaxation rates confirm the disordered nature of the C-terminus. Transverse relaxation rates clearly suggest an interaction between the C-terminus of the disordered region and the folded domain of XLF. It is possible that this interaction is artefactual and driven by the interaction of the His<sub>6</sub> tag with the negatively charged patches on XLF (Fig.IV.11). It is necessary to work with a construct with no His-tag (including a cleavage site) in order to confirm the presence of a native interaction between the C-terminus of the protein and the folded domains or disrupt an artefactual interaction.

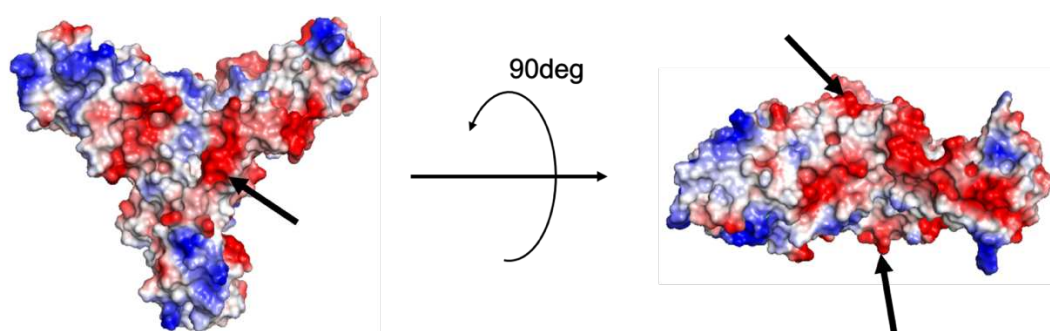


Figure IV.11 Vacuum electrostatics of the folded region of XLF. Arrows point to the C-terminus end of the folded domain.

A new XLF without the C-terminal His-tag was prepared by Duc Duy Vu more recently in our group, thanks to which, the last residues on the C terminus were assigned. Series of relaxation experiments were carried out, as well as a titration with DNA. We present some of these last results to complete this first conformational dynamics analysis of the IDR of XLF.

To determine the dissociation constant  $K_D$ ,  $^{13}\text{C}/^{15}\text{N}$  labelled XLF was titrated by DNA (30 base pairs). Seven concentration ratios  $[\text{DNA}]/[\text{XLF}]$  were used with from 0 to 4. *Fig. IV.12* shows the evolution of the  $^1\text{H}-^{15}\text{N}$  signals on the HSQC spectra during titration of XLF protein by DNA. Chemical shift perturbations (CSP) profile of the disordered region is shown in *Fig. IV.13*. Significant CSP can be observed from residue 282 to 299, indicating a fast exchange of binding with DNA. Residues 289 to 295 is a highly conserved cluster in different species.<sup>147</sup> The region 288-295 (KVKRKKPR) includes 4 lysine residues and 2 arginine residues, which are well known to interact with DNA through the insertion of their positively charged side-chain groups into the minor groove of DNA<sup>158</sup>. This region displays the largest CSPs, showing that it is the binding site with DNA. Note however that these changes of chemical shift are smaller than those expected upon folding of the interacting motif. Thus, the interaction of XLF and DNA is likely mediated by interactions of sidechains while the backbone retains a significant degree of disorder. The last 5 residues on the C-terminus, which is the Ku binding site, show smaller CSPs, indicating that the Ku binding site is not directly involved in the interaction with DNA and available for interaction with Ku.

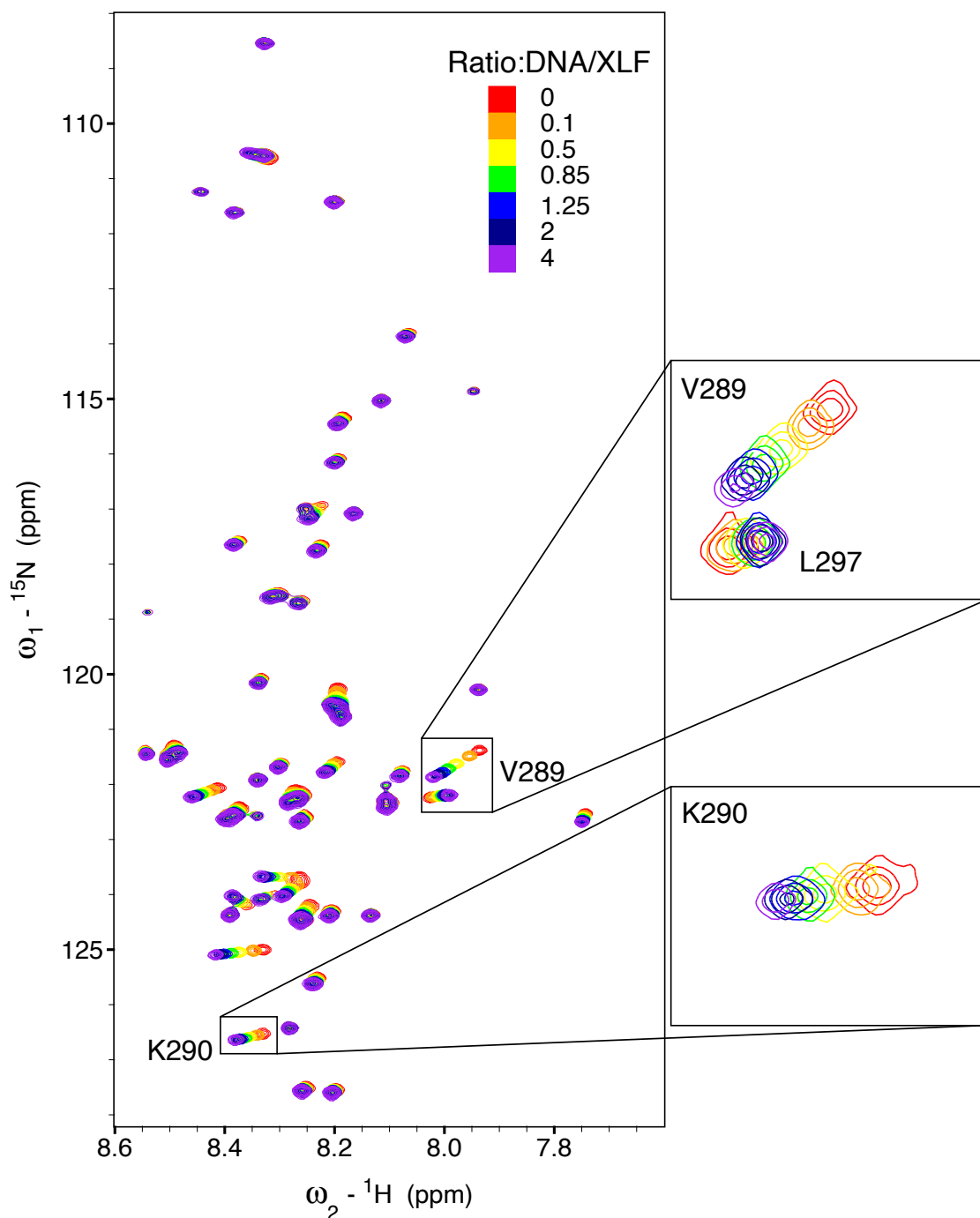


Figure IV.12  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{13}\text{C}/^{15}\text{N}$  labelled XLF titrated with DNA (30 base pairs). The spectra obtained with different  $[\text{DNA}]/[\text{XLF}]$  ratios are presented in different colors. Inserts show the regions of residues V289 K290 and L297, where substantial changes of chemical shifts can be observed. The double strand DNA sequences are : 5'-GATCCCTCTAGATATCGGGCCCTC GATCCG -3' and 5'- CGGATCGAGGGCCCGATATCTAGAGGGATC-3'.

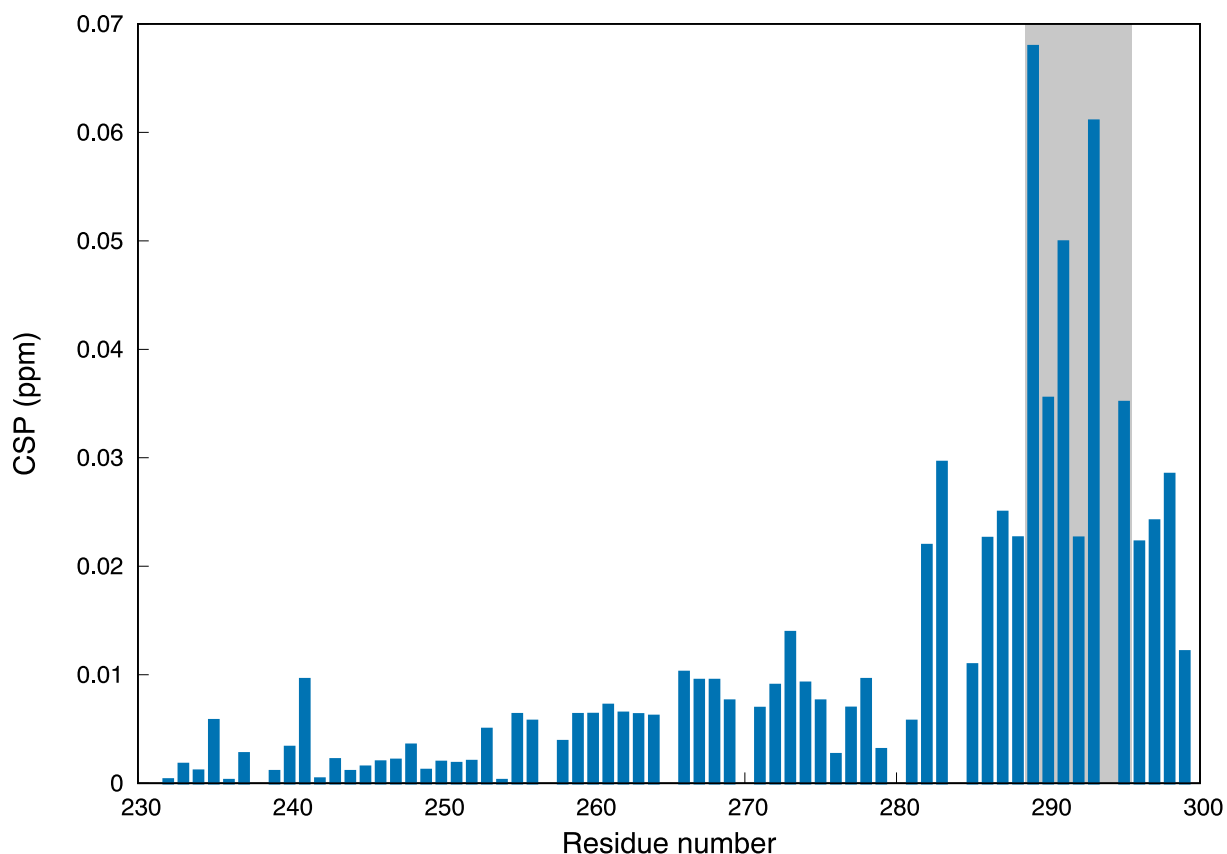


Figure IV.13 CSP in the IDR of XLF almost saturated by DNA.  $\Delta_{CSP} = \frac{1}{2} \sqrt{(\Delta_H^2 + 0.01\Delta_N^2)}$ .

The lineshape of residues V289 and K290 were fitted and the CSPs were used to characterize the dissociation constant of the complex  $K_D$  in TITAN<sup>159</sup> (Fig.IV.14). Their profiles show that the protein was near saturation in the end of the titration. The fitted  $K_D$  values were  $14.5 \pm 0.4 \mu\text{M}$  and  $11.6 \pm 0.8 \mu\text{M}$  for V289 and K290, respectively. Note that this rather weak binding of a single XLF dimer with a rather short double-stranded DNA would lead to extremely strong interactions under cooperative binding in the context of XLF-XRCC4 filaments.

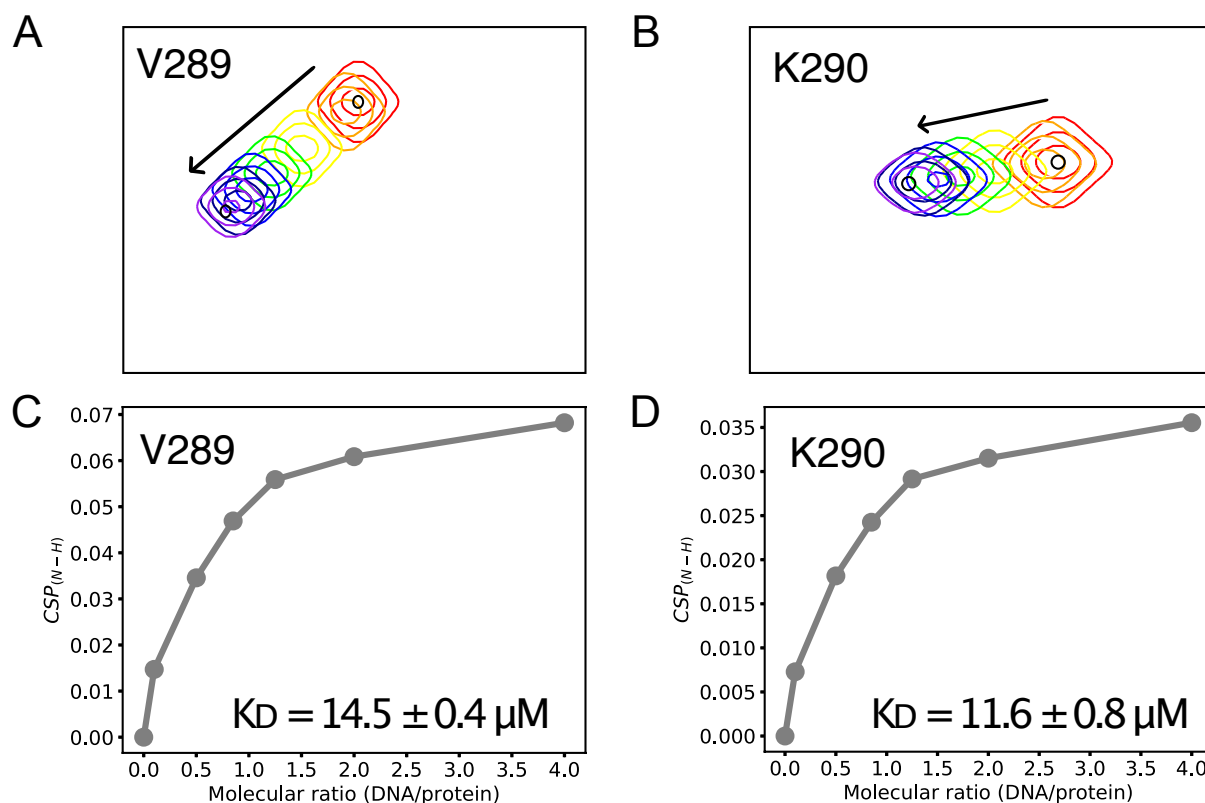


Figure IV.14 A and B are the fitted signals from the experimental spectra of residues V289 and K290, respectively. Their titration profiles and results are shown in C and D.

## 6 Conclusion

At the current stage, we can make some conclusions on human XLF based on the first NMR results. We have assigned the C-terminal region of XLF, this assignment confirms it is an IDR. The relaxation analysis has identified an artefactual interaction of the C-terminal His-tag of our first construct, likely with the folded domains of XLF. This has led us to design a second construct, with a cleavable His tag at the N-terminus, which has been prepared by Duc Duy Vu. We have investigated the interaction of the IDR of XLF with double-stranded DNA. The IDR of XLF and DNA interact through the highly positively charged 288-295 region. The dissociation constant in the micromolar range corresponds to a weak interaction of individual XLF dimers with DNA but explain the tight interaction observed for XLF-XRCC4 filaments and long double-stranded DNA.





## Chapter V DNA polymerase $\eta$ : protein expression and isotope labeling

NMR is pushing its limitation in investigation of large biomolecule systems, thanks to employment of advanced instruments, novel pulse sequences and the development of new sample labeling techniques.

Our objective here was to combine modern protein isotope labeling and advanced NMR methodology to study conformational dynamics of DNA polymerase  $\eta$ , which is a large folded protein involved in translesion DNA synthesis. In the frame of this thesis, protein expression and isotope labeling optimization will be presented and discussed.

## 1 Translesion synthesis and polymerase $\eta$

Many DNA repair mechanisms correspond to different types of DNA damages, as presented in the previous chapter. In case where lesions remain unrepaired in the S-phase, DNA damage tolerance pathways help cells to survive. Translesion DNA synthesis (TLS) polymerases is the last support to overcome blockage of replication.

Polymerase  $\eta$  (Pol $\eta$ ) is a multi-domain protein that belongs to the largest family of all TLS polymerases, the Y-family. Pol $\eta$  is the member known for accurate replication through ultraviolet-induced cyclobutane pyrimidine dimer (CPD) in eukaryotic cells<sup>160</sup>. Patients with nonfunctional Pol $\eta$  suffer from an increased incidence of sunlight-induced skin cancers (the variant form of *xeroderma pigmentosum*)<sup>161-162</sup>.

The Y-family has common structure topology with other families of polymerases: a right hand with palm, thumb and fingers domains. These domains are highly conserved throughout the Y-family. The C-terminus of Y-family polymerases called polymerase associated domain (PAD) or little finger domain, is a distinctive feature of the Y family. This domain also determines the biochemical properties of each polymerase member.<sup>163-164</sup> Single molecules fluorescence resonance energy transfer (smFRET) studies on another member of the DNA polymerase family, polymerase I, has shown that it is in conformational equilibrium between an open and a closed form, and the equilibrium is modified once the DNA is bound to the protein.<sup>165</sup> Pre-steady-state kinetics have suggested that conformational changes of Pol $\eta$  are a key step in the catalytic cycle.<sup>166</sup> Structures of the open *apo* form<sup>167</sup> and closed *holo* form (with DNA templet)<sup>162</sup> of Pol $\eta$  have been resolved by crystallography (*Fig. V.1*). The PAD (green) domain moves towards to the thumb domain (yellow) when binding with the DNA template.

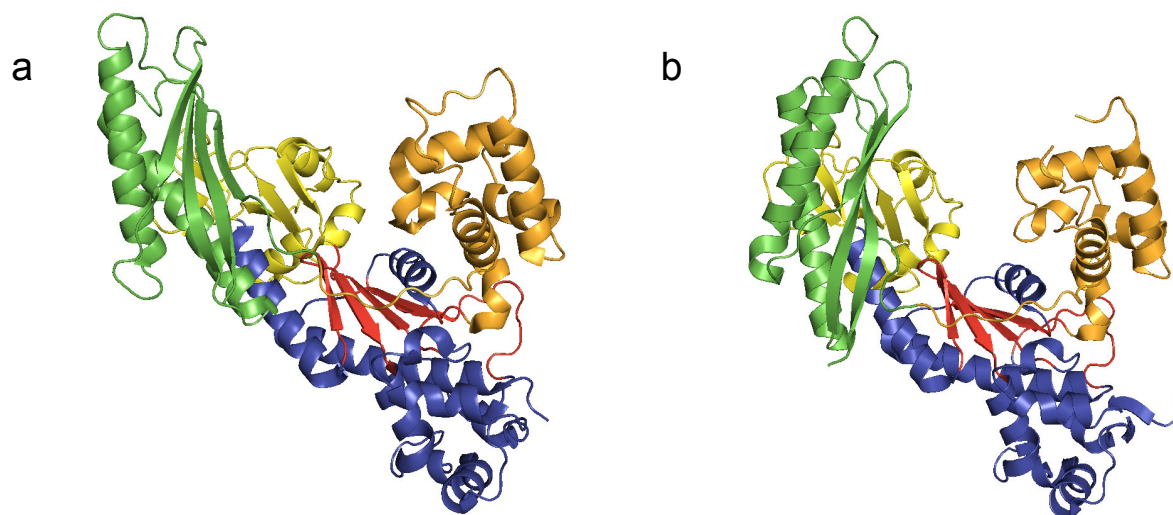


Figure V.1 Crystal structures of the apo (PDB: 1JIH) (a) and holo (PDB: 3MFH, ternary complex with ATP and undamaged DNA template) (b) forms of *S. cerevisiae* Pol $\eta$ . The palm, thumb, fingers and PAD domain are labelled respectively in blue, orange yellow and green. The red beta sheets are a palm subdomain, essential for metal ion binding. The PAD domain moves relatively to the other domains of the protein by about 1.3 nm when DNA template is bound.

The aim of the project was to characterize domain and internal motions in *S. cerevisiae* Pol $\eta$  (513 residues, 58 kDa). Recently developed NMR methods give the possibility to study fluctuations of conformations and transitions between different conformations that take place on timescales ranging from the picosecond to the second<sup>25, 62, 168</sup>. Instead of uniformly  $^{13}\text{C}$  labeling, only methyl group on side chains of methyl-bearing amino acids (Ala, Leu, Val, Ile, Thr, Met) are selectively labeled with  $^{13}\text{CH}_3$  in a deuterated background. These selectively labeled methyl groups serve as probes to motions of the protein and improve the spectral resolution for the large-size system.

The specific labeling requires a protein production in  $\text{D}_2\text{O}$ . This initial step is essential to the project. The sample preparation optimization will be presented and discussed in this chapter.

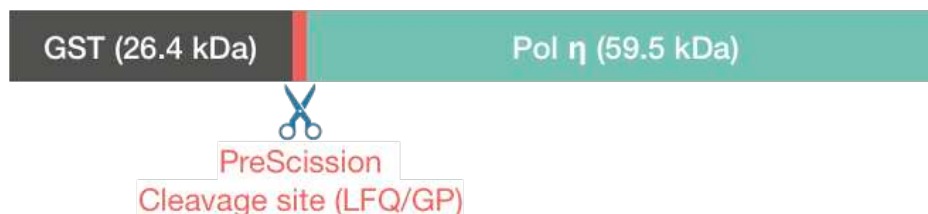
## 2 Protein expression and labelling

### 2.1 General strategy

The construct of the fusion protein corresponds to residue 1-513 of *S. cerevisiae* Pol $\eta$ , a cleavage site and Glutathione S-transferase (GST) (Fig. V.2). Plasmid pGEX-6p-3 containing this construct was transformed into competent *E. coli* BL21 Star<sup>TM</sup>(DE3)pLysS. The fusion protein was expressed in the host *E. coli*, and then extracted for purification. The fusion protein

(MW 85.9 kDa) was first bound to GST resin. Then, the target protein was cleaved by PreScission protease recognition site from the GST tag. Followed by additional chromatography steps, Pol $\eta$  (MW 59.5 kDa) was purified.

**A**



**B**

10	20	30	40	50	60	70	80	90	100
GPLGSPNSRV	DMSKFTWKEL	IQLGSPSKAY	ESSLACIAHI	DMNAFFAQVE	MSPILGYWKI	KGLVQPTRL	LEYLEEKYEE	HLYERDEGDK	WRNKKFELGL
110	120	130	140	150	160	170	180	190	200
EFPNLPYYID	GDVKLTQSM	IIRYIADKHN	MLGGCPKERA	EISMLEGAVL	DIRYGVSRIA	YSKDFETLKV	DFLSKLPPEML	KMFEDRLCHK	TYLNGDHVTH
210	220	230	240	250	260	270	280	290	300
PDFMLYDALD	VVLYMDPMCL	DAFPKLVCFK	KRIEAIPOID	KYLKSSKYIA	WPLQGWQATF	GGGDHPPKSD	LEVLFQGPLG	SPNSRVDMSK	FTWKELIQLG
310	320	330	340	350	360	370	380	390	400
SPSKAYESSL	ACIAHIDMNA	FFAQVEQMRC	GLSKEDPVVC	VQWNSIIAVS	YAARKYGISR	MDTIQEALKK	CSNLIPIHTA	VFKKGEDFWQ	YHDGCGSWVQ
410	420	430	440	450	460	470	480	490	500
DPAKQISVED	HKVSLPEYRR	ESRKALKIFK	SACDLVERAS	IDEVFLDLGR	ICFNMLMFDN	EYELTGDCLK	KDALSNIREA	FIGGNYDINS	HLPLIPEKIK
510	520	530	540	550	560	570	580	590	600
SLKFEQDVFN	PEGRDLITDW	DDVILALGSQ	VCKGIRDSIK	DILGYTTSCG	LSSTKNVCKL	ASNYKKPDAQ	TIVKNDCLLD	FLDCGKFEIT	SFWTLGGVVG
610	620	630	640	650	660	670	680	690	700
KELIDVLDLP	HENSIKHIRE	TWPDNAGQLK	EFLDAKVQKS	DYDRSTSNID	PLKTADLAEK	LFKLSRGRYG	LPLSSRPVVK	SMMSKNLNRG	KSCNSIVDCI
710	720	730	740	750	760	770	780	790	800
SWLEVFCAEL	TSRIQDLEQE	YNKIVIPRTV	SISLKTksye	VYRKSgpVAY	KGINFQshel	LKVGikFvTD	LDIKGknksy	YPLTKLSMTI	TNFDIIDLQK

AAAS

Figure V.2 Construct of recombinant *S. cerevisiae* Pol $\eta$  (A) and its sequence (B). The GST-tag, the cleavage site and the protein are labelled in grey, red and green respectively.

To optimize the transformation and expression protocol, LB culture medium and H<sub>2</sub>O M9 medium were used. For the selective labeling, M9 media with different concentrations of D<sub>2</sub>O were used.

## 2.2 Transformation optimization and strain selection

### 2.2.1 Optimization of *E. coli* transformation and protein expression

To obtain Pol $\eta$  from *E. coli* strain, different transformation combinations and expression conditions were tested. Fig. V.3 describes the main optimized parameters in protein production steps. Two vectors carrying the recombinant protein coding sequence (pGEX-6p-3 and pET-

42a) and two competent BL21 (Gold<sup>TM</sup>(DE3)pLysS and Star<sup>TM</sup>(DE3)pLysS) were tested. Each plasmid was successfully transformed into two strains. Colonies of all four transformations were induced for protein expression in small volume of LB medium. Trials were realized under different conditions such as incubation durations, temperatures and IPTG concentrations. Only the combination of pGEX-6p-3 and BL21 Star(DE3)pLysS was able to produce the target protein.

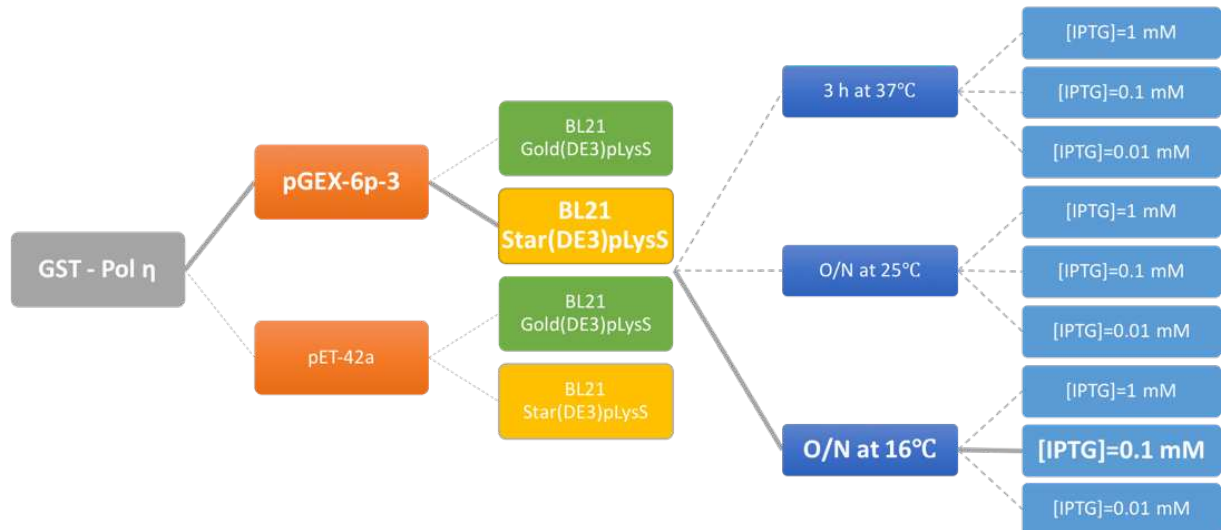


Figure V.3 Different conditions used for transformation and induced expression in LB culture medium. Combinations of two plasmids: pGEX-6p-3 and pET-42a and two strains: Gold<sup>TM</sup>(DE3)pLysS and Star<sup>TM</sup>(DE3)pLysS were tested for transformation. Obtained colonies are amplified in 5 mL of LB. Expression was induced by adding different concentrations of IPTG at 3 temperatures. The optimized protocol is connected with real gray line.

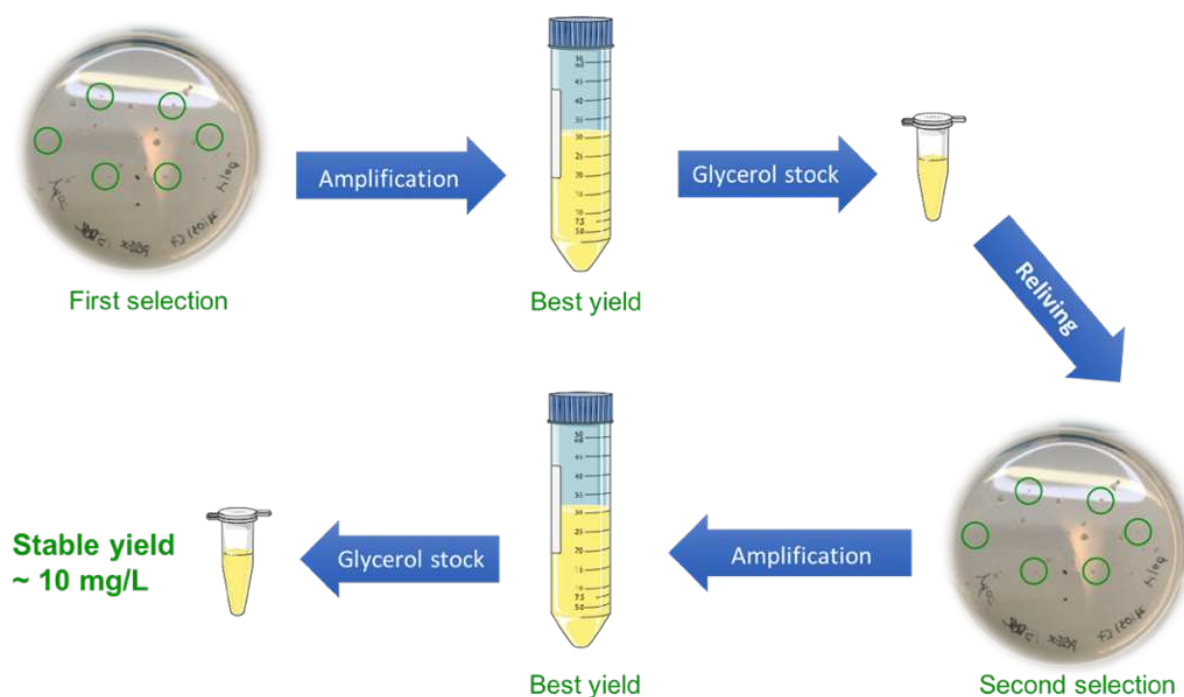
The mixture of one vial of Star<sup>TM</sup>(DE3)pLysS competent cells and 1  $\mu$ L of pGEX-6p-3 plasmid was incubated on ice for 30 min, then heat shocked in a 42°C-water bath for 30 s and immediately replaced on ice. 250  $\mu$ L of preheated LB medium were added, the transformed cells were shaken at 225 rpm, 37°C for 1 hour. 20  $\mu$ L of the reaction were spread onto a LB agar plate containing 34  $\mu$ g/mL chloramphenicol and 100  $\mu$ g/mL ampicillin for overnight incubation. The same concentrations of antibiotics were used all along the expression process of the recombinant Pol $\eta$ .

### 2.2.2 Double colony selection and glycerol stock

The combination of pGEX-6p-3 and BL21 Star(DE3)pLysS led to the expression of the target protein. Yet, the level of expression was not stable. It was observed that colonies of the transformed bacteria after an overnight incubation at 37°C were of different sizes (Fig.V.4),

and only small sized colonies gave good yield of protein production. This could be an indication of toxicity of the recombinant plasmid to the host bacterial cells. Moreover, a severe decrease of yield was also observed after freezing stocks at  $-80^{\circ}\text{C}$ .

To overcome the production instability, a double selection on LB agar dishes<sup>169</sup> was performed (*Fig.V.4*). Six small-sized-colonies were amplified and induced separately in fresh LB medium. The strain that led to the best yield was spread on a new dish, and went through a second selection of size and yield. The final champion colony was conserved in LB medium with 15% of glycerol at  $-80^{\circ}\text{C}$  afterwards.



*Figure V.4 Photo of the irregular colonies and workflow of double colony selection on LB agar.*

### 2.3 Protein expression in $\text{H}_2\text{O}$ M9

The following protocol was suitable for producing uniformly  $^{13}\text{C}$ - $^{15}\text{N}$  labeled Pol $\eta$  (*Fig.V.5*). It was also designed to for further adaptation of production in  $\text{D}_2\text{O}$ .

1. One bacterial stock was thawed and then relived in 5 mL LB medium agitating at 225 rpm,  $37^{\circ}\text{C}$ .
2. When the  $\text{OD}_{600}$  reached 0.6, the culture was centrifuged at 3000 g,  $37^{\circ}\text{C}$  for 10 min. The pellet was resuspended in 100 mL M9 medium to start an overnight-preculture from  $\text{OD}_{600}$  of 0.05.

3. The preculture was collected and centrifuged the next morning, to start 2 L of fresh M9 from OD<sub>600</sub> of 0.05.
4. When OD<sub>600</sub> reached 0.6 ~ 0.8, the temperature was decreased to 16°C and protein expression was induced by adding IPTG to a final concentration of 0.1 mM.
5. After 18 hours of incubation, the culture was centrifuged for 30 min at 3000 g, 4°C. The pellet was transferred into a falcon tube and kept at -80°C. Samples were collected before and after induction for SDS-PAGE analysis.

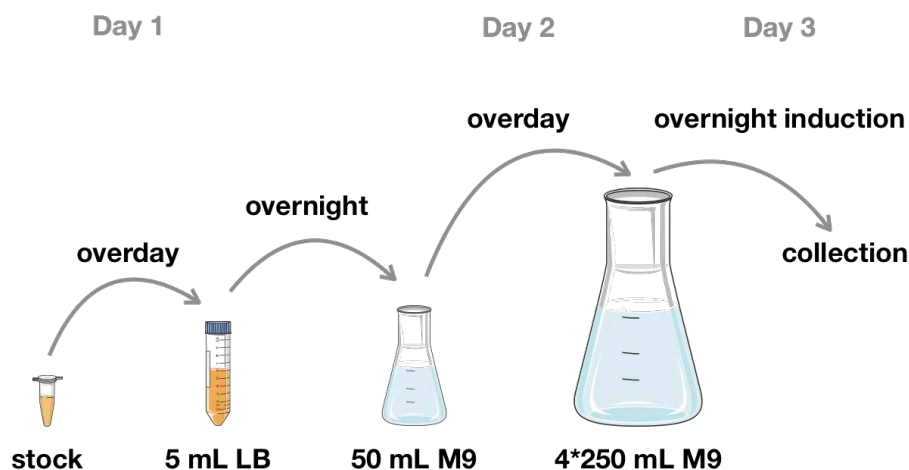


Figure V.5 *E. coli* adaptation and protein expression in M9.

#### 2.4 Protein expression in D<sub>2</sub>O M9

The production in D<sub>2</sub>O M9 is based on the protocol provided by the Practical School on Advanced Isotopic Labelling Methods for Biological NMR in Grenoble 2017. Compared to production protocol in H<sub>2</sub>O, progressive steps were added to train the bacteria to adapt the critical living environment in 100% D<sub>2</sub>O (Fig. V.6).

1. One bacterial stock was thawed and then relived in LB medium 225 rpm, 37°C.
2. When the OD<sub>600</sub> reached 0.6, enough quantity of culture was taken and centrifuged at 3000 g, 37°C for 10 min. The pellet was resuspended in 50 mL H<sub>2</sub>O M9 medium to start an overnight-preculture from OD<sub>600</sub> of 0.01, to avoid saturation of the medium.
3. The next morning, enough quantity of the preculture was collected, centrifuged and transferred to 50 mL of 50% D<sub>2</sub>O M9 medium, with a starting OD<sub>600</sub> of 0.05.
4. OD<sub>600</sub> should reach 2 ~ 2.5 in the evening the same day. The culture was collected, centrifuged and transferred to 100 mL of 100% D<sub>2</sub>O M9 medium, with a starting OD<sub>600</sub> of 0.05.



5. In the morning, OD<sub>600</sub> should reach 1 ~ 1.5. Enough quantity of culture was taken and centrifuged. This final culture had a starting OD<sub>600</sub> of 0.1.
6. When OD<sub>600</sub> reaches 0.6 ~ 0.8, the temperature was decreased to 16°C and recombinant protein expression was induced by adding IPTG to a final concentration of 0.1 mM.
7. After 18 hours of incubation, the culture was centrifuged for 30 min at 3000 g, 4°C. The pellet was transferred in a falcon tube and kept at -80°C. Samples for SDS-PAGE electrophoresis were collected before and after induction.

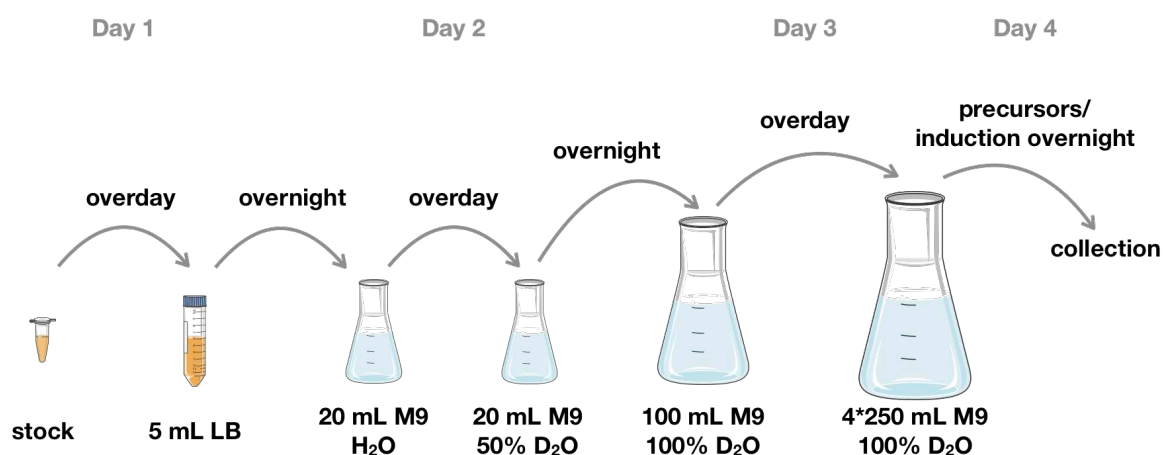


Figure V.6 *E. coli* adaptation and protein expression in D<sub>2</sub>O.

## 2.5 Protein purification

Three columns of different purification principles were employed in this protocol.

### 1. Bacteria lysis

One pellet of 1 L of bacterial culture was re-suspended in 30 mL of glutathione resin equilibration buffer (25 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1% Triton, 5% glycerol) to which a small amount of lysozyme was added. The pellet was gently mixed till well dispersed in the buffer, then lysed by 10 sonication cycles of 15 s interleaved with 1 min breaks. The homogenous lysate was centrifuged at 30 000 g for 30 min. The supernatant was filtered by membranes with successive pore diameters 1.2, 0.45 and 0.2 μm. Afterwards, the sample was clear and transparent. The sample was always kept at 4°C or on ice to avoid protein degradation. Besides, at the beginning of the lysis, the antiproteases PEP, LEA, PMSF were used for the same purpose.

### 2. Glutathione resin – GST tag affinity chromatography

The solution obtained from the previous step was mixed with 4 mL pre-equilibrated GST resin under agitation at 4°C. After 4 hours, the resin was washed by 50 mL equilibration

buffer and 50 mL PreScission protease cleavage buffer (50 mM Tris-HCl pH 7, 150 mM NaCl, 1 mM EDTA, 1 mM DTT). An overnight on-resin cleavage was performed by adding 1mg of PreScission protease to the re-suspended resin. The resin was packed into a glass column the next day. The flow through and 10 mL cleavage buffer used for washing the resin was collected for further purifications. The residue of GST-tag was eluted with a washing buffer (50 mM Tris-HCl pH 7, 150 mM NaCl, 25 mM glutathione).

### 3. SP Sepharose column – cation exchange chromatography

The fractions of interest from the last step were dialyzed in 1L of SP binding buffer (25 mM NaH<sub>2</sub>PO<sub>4</sub> pH 6.5, 50 mM KCl, 0.1 mM MgSO<sub>4</sub>, 1 mM EDTA, 1 mM DTT) at 4°C under agitation overnight, before being injected into the column. The 5 mL column was washed with the binding buffer till the UV value and conductivity stabilized. A gradient of 40 min from 0 to 90% elution buffer (25 mM NaH<sub>2</sub>PO<sub>4</sub> pH 6.5, 50 mM KCl, 0.1 mM MgSO<sub>4</sub>, 1 mM EDTA, 1 mM DTT) was used.

### 4. Gel Filtration Superdex 75 – size exclusion chromatography

Fractions collected from the SP column were concentrated and injected into the 120 mL gel filtration (GF) column. The column was eluted by GF buffer (25 mM NaH<sub>2</sub>PO<sub>4</sub> pH 6.5, 150 mM KCl, 0.1 mM MgSO<sub>4</sub>, 1 mM DTT). Collected fractions were concentrated up to 300  $\mu$ M. This buffer was optimized with a scanning technique as presented in the following paragraph.

## 2.6 Buffer optimization by differential scanning fluorimetry

Maintaining the sample stability at high protein concentration during a long acquisition time can be challenging. In general, the stability of the purified protein is influenced by the temperature and the chemical environment. Choosing a right buffer for the recombinant protein can help to keep the protein well folded and extend the lifetime of the sample. One way of evaluating protein stability is through the measurement of the melting temperature ( $T_m$ ), at which half of the protein in the sample is unfolded. For that, differential scanning fluorimetry (DSF) tests were used to optimize the buffer for NMR experiment<sup>170-171</sup>. DSF is a method that allows to measure efficiently the melting temperature of protein in different buffers. The experiment was performed on a real-time quantitative reverse transcription instrument as a fluorimeter, with a 96-well PCR-plate and using a hydrophobic dye as an indicator.

The used fluorophore for Pol $\eta$  buffer screening is SYPRO Orange. SYPRO Orange signal is quenched by water. It gains back the fluorescence once binding to hydrophobic pocket

of the protein. When the protein is unfolded, more hydrophobic regions are exposed, and the fluorescence signal increases.

To find a suitable NMR buffer for Pol $\eta$ , two rounds of DSF tests are carried out using 96-well microplates. For the first plate, we verified the concentration of protein and dye to obtain sufficient signal to noise ratio. We also tested four common NMR buffers (NH<sub>4</sub>OAc, MES, Tris-Bis and NaH<sub>2</sub>PO<sub>4</sub>) at different concentrations, covering pH 4.5-6.5, two most used salt (NaCl and KCl) and two additives (EDTA and MgCl<sub>2</sub>). The design of the second screening microplate was a refinement based on result from the first one. More additives were tested. The screened conditions are presented in *Table V-1* and *V-2*.

*Table V-1 DSF screening conditions on first microplate*

	1	2	3	4	5	6	7	8	9	10	11	12
A	water	SYPRO Orange 1-5X, Protein 0.01-0.1%										
B	NH <sub>4</sub> OAc pH 4.5-5.5				MES pH 5.5-6.5			Tis-Bis pH 6-6.5			NaH <sub>2</sub> PO <sub>4</sub> pH 6.5	
C	B + 50 mM NaCl											
D	B + 150 mM NaCl											
E	B + 50 mM KCl											
F	B + 150 mM KCl											
G	C + 1 mM EDTA											
H	C + 10 mM MgCl <sub>2</sub>											

*Table V-2 DSF screening conditions on second microplate*

	1	2	3	4	5	6	7	8	9	10	11	12
A	50 mM Tis-Bis pH 6-6.5				25 mM NaH <sub>2</sub> PO <sub>4</sub> pH 6.5-7				50 mM NaH <sub>2</sub> PO <sub>4</sub> pH 6.5-7			
B	A + 150-250 mM NaCl											
C	B + 0.1-1 mM MgCl <sub>2</sub>											
D	B + 0.1-1 mM MnCl <sub>2</sub>											
E	B + 0.1-1 mM MgSO <sub>4</sub>											
F	B + 1-5% mM glycerol											
G	B + 0.1-1 mM EDTA											
H	B + 0.1-1 mM ATP											

The obtained raw data was the absolute fluorescence as a function of temperature. To facilitate the identification of  $T_m$  and comparisons between different conditions, the results are presented as unfolded protein fraction versus temperature.

### 3 Results and discussion

#### 3.1 Protein expression

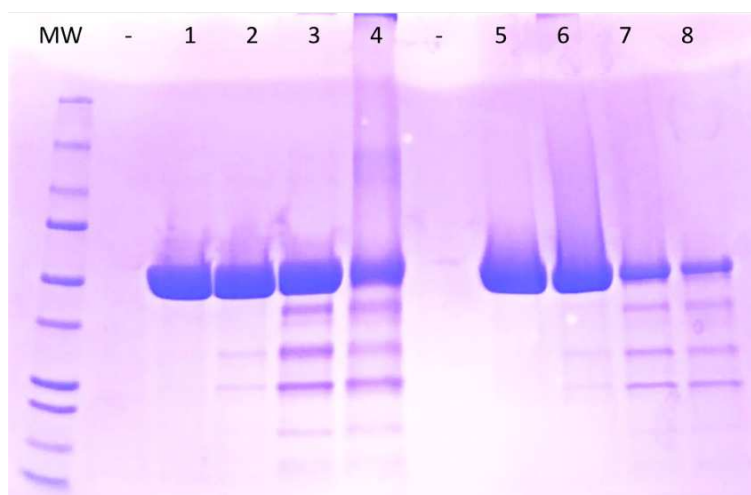
On the basis of a series of tests, it was found that only small colonies were efficient at producing recombinant Pol $\eta$ . Thanks to the double selection on agar plate, the final glycerol bacterial cells stocks gave a stable and satisfactory yields of 10 mg of fusion protein per liter of H<sub>2</sub>O M9 medium. As for the production in D<sub>2</sub>O, the yield decreased dramatically. Only 0.5 mg of Pol $\eta$  was obtained from 1 L of D<sub>2</sub>O M9 medium. The low yield might be due to several aspects.

1. Toxicity of the transferred plasmid. The size of transformed colonies was nonuniform, and only small-size-colonies produced Pol $\eta$ . The generation time of the strain was long in H<sub>2</sub>O M9 (1.5 hours) compare to usual case (30-45 min). It might be that the replication of the plasmid or expression leaking of Pol $\eta$  slowed down the growth of the bacteria. Optimization of the coding sequence or use of more competent strains could be tested.
2. Critical surviving condition in D<sub>2</sub>O. Culture tests were made in small volume of D<sub>2</sub>O (5 mL). The culture had a reasonable generation time (around 2 hours) and showed nice expression band on the SDS-PAGE. As for large volume of D<sub>2</sub>O culture, the generation time was more than 4 hours, and the band of recombinant Pol $\eta$  was not visible. Residual H<sub>2</sub>O (HOD) was higher in small-scale test than in large-scale culture, which might be crucial for the wellbeing of this transformed *E. coli* strain. For this, double colony selection on D<sub>2</sub>O agar plates may help to find strain that is more resistant to D<sub>2</sub>O and improve the yield.
3. Instability of ampicillin. Ampicillin-resistant *E. coli* express  $\beta$ -lactamase to neutralize the antibiotic. In the case of long culture in D<sub>2</sub>O, ampicillin might be totally neutralized. This less selective medium could lead to a decrease of the yield. To solve this problem, it is possible to use carbenicillin, which is equivalent to ampicillin but has better resistance to  $\beta$ -lactamase.

### 3.2 Protein purification and stability

The first challenge came from what was first thought to be “impurities” apparent on SDS-PAGE. Although UV detected chromatography indicated nice symmetric, low molecular weight bands on SDS-PAGE, mass spectrometry analysis indicated that the “impurity bands” corresponded to fragments of Pol $\eta$ . Many trials to eliminate possible contaminants and sources of degradations during the purification process were attempted (use of different anti-proteases, tag cleavage on resin and off resin, rinsing by ion exchange chromatography, different buffers). Yet none of them led to an improvement.

An unconventional hypothesis was made: the degradation may be caused by the heating process during sample preparation for the SDS-PAGE. To verify this hypothesis, purified samples were mixed with SDS-PAGE loading buffer, containing or without 20% of  $\beta$ -mercaptoethanol (BME). After leaving the samples at room temperature, 60°C for 10 min, 95°C for 10 min and 1 h, they were analyzed on an SDS-PAGE (*Fig. V.7*). Samples that stayed at room temperature showed one band on the gel, while the ones that received heating showed various degrees of degradation. The presence of BME could not prevent or slow down the degradation. Normally, the heating should not be sufficient to break covalent bonds. However, this unexpected phenomenon was observed in the case of Pol $\eta$ . The origin of this unusual instability is still unknown, but this should not influence any future experiments at room temperature.



*Figure V.7 SDS-PAGE of the stability test under different conditions. Lanes 1 to 4 were samples mixed with loading buffer without BME; lanes 5 to 8 were samples mixed with loading buffer containing 20% of BME. Lanes 1 and 5: without heating; lanes 2 and 6: heated at 60°C for 10 min; lanes 3 and 7: heated at 95°C for 10 min; lanes 4 and 8: heated at 95°C for 60 min.*

After resolving the problem, SDS-PAGE proved that the purification of the protein was efficient (Fig.V.8). The GST tag was removed from the recombinant protein with the help glutathione resin chromatography. The ion exchange column helped to decontaminate from binding DNA. Final filtration by size exclusive chromatography purified furtherly the final product.

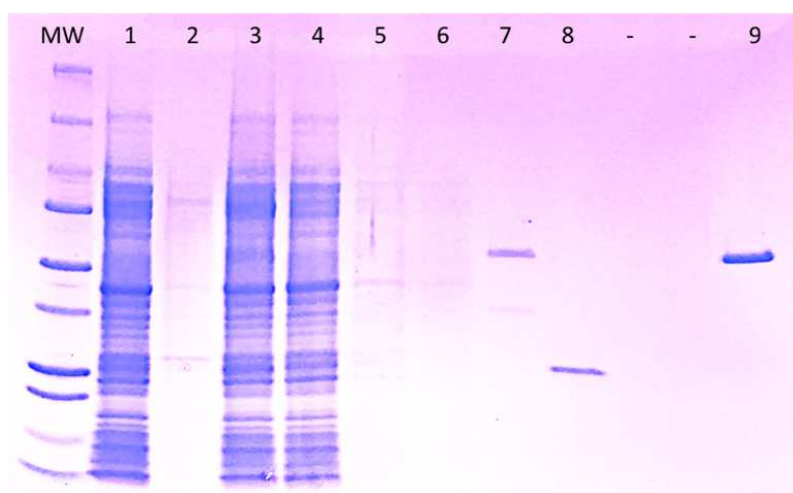


Figure V.8 Electrophoresis SDS-PAGE of the purification process. Lane 1 to 8 are samples collected during the glutathione resin purification. 1: supernatant fluid of the bacterial lysate; 2: solid fraction of the lysate; 3: whole sample after fixation to resin; 4: flow through after over; 5: wash with equilibration buffer; 6: wash with cleavage buffer; 7: flow through after digestion by PreScission, the heavier band was Pol $\eta$ ; 8: wash with elution buffer, the band was cleaved GST-tag. Lane 9 was the final product after all purification steps, containing only Pol $\eta$ .

### 3.3 Buffer optimization

The optimal concentration of SYPRO Orange and protein were 5X and 0.05% respectively. From the two DSF tests, it can be concluded that Pol $\eta$  preferred neutral pH (6.5-7) in phosphate buffer than Tris-Bis. Sodium salt and potassium salt stabilized Pol $\eta$ , and the effect had a positive correlation with the concentration. As additives, 0.1 mM MgSO<sub>4</sub>, 0.1 mM ATP and 5% glycerol improved the stability. However, ATP and glycerol were not used in the buffer, for that ATP might induce conformation changes and glycerol increase the viscosity of the buffer, which is undesirable for NMR experiments. DTT didn't help increasing T<sub>m</sub> but was introduced to the buffer to keep cysteine residues in their reduced form.

The final buffer is composed by 25 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 6.5), 150 mM KCl, 0.1 mM MgSO<sub>4</sub>, 1 mM DTT. Thanks to DSF, the melting temperature of the protein increased by 7.5 °C

(Fig.V.9) compared to the initially used buffer. This optimized buffer is used in the final purification step by SEC.

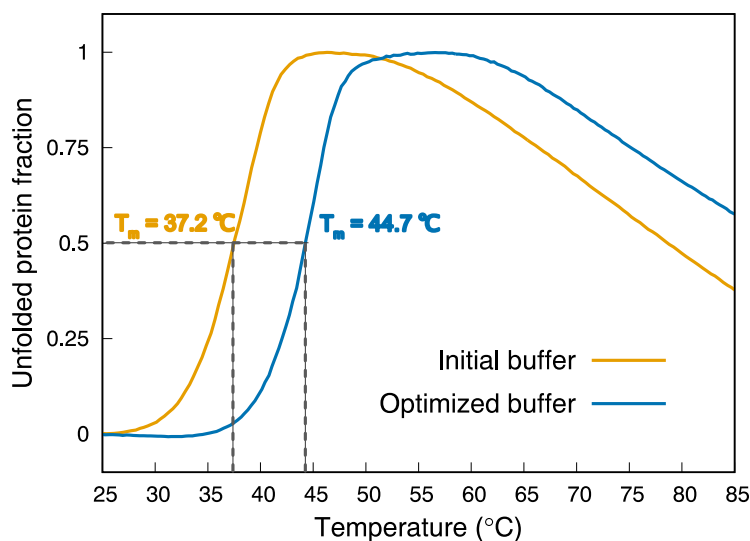


Figure V.9 Melting curves of Pol $\eta$  in two different buffers. The  $T_m$  was 37.2 °C in initial buffer (20 mM Bis-Tris pH 6.5, 50 mM NaCl, 1 mM EDTA and 1 mM DTT) and 44.7 °C in the optimized buffer (25 mM NaHPO<sub>4</sub> pH 6.5, 150 mM KCl, 0.1 mM MgSO<sub>4</sub> and 1 mM DTT).

#### 3.4 Stability challenge in D<sub>2</sub>O

In spite of the optimization of sample buffer conditions, the stability of the final Pol $\eta$  sample was still challenging. Once D<sub>2</sub>O was introduced into the final concentrated sample for NMR experiment, severe precipitation was observed immediately. This may be explained by the alteration of the hydrophobic effect and hydrogen bonds in D<sub>2</sub>O, which could reduce the stability of a folded domain of Pol $\eta$  and increase intermolecular interactions. Increasing progressively D<sub>2</sub>O concentration avoids brutal environmental shock to the protein. However, the D<sub>2</sub>O tolerance level of Pol $\eta$  was no more than 2%. The sample became opaque after 5 min after a progressive introduction of 2% D<sub>2</sub>O.

## 4 Conclusion

As any project in its early stage, sample preparation is the essential step to be established. Protocol of transformation, expression and purification of recombinant pol $\eta$  from *S. Cerevisiae* in H<sub>2</sub>O M9 minimum medium has been optimized. The yield is good, around 10 mg of pure protein was obtained per liter of M9 H<sub>2</sub>O medium. A differential scanning fluorimeter has been used to improve the stability of the protein in different NMR buffers. Despite of many efforts

were made, the decrease of the yield in D<sub>2</sub>O culture medium and the low stability of the protein in D<sub>2</sub>O have prevented us to use high-resolution NMR efficiently.





## Bibliography

1. Bloch, F., Nuclear Induction. *Physical Review* **1946**, *70* (7-8), 460-474.
2. Bloembergen, N.; Purcell, E. M.; Pound, R. V., Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Physical Review* **1948**, *73* (7), 679-712.
3. Levitt, M. H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Wiley: 2001.
4. Keeler, J., *Understanding NMR Spectroscopy*. Wiley: 2011.
5. Solomon, I., Relaxation Processes in a System of Two Spins. *Physical Review* **1955**, *99* (2), 559-565.
6. Tugarinov, V.; Sprangers, R.; Kay, L. E., Probing Side-Chain Dynamics in the Proteasome by Relaxation Violated Coherence Transfer NMR Spectroscopy. *Journal of the American Chemical Society* **2007**, *129* (6), 1743-1750.
7. Sekhar, A.; Rumpfolt, J. A. O.; Broom, H. R.; Doyle, C. M.; Bouvignies, G.; Meiering, E. M.; Kay, L. E., Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *eLife* **2015**, *4*, e07296.
8. Rosenzweig, R.; Sekhar, A.; Nagesh, J.; Kay, L. E., Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *eLife* **2017**, *6*, e28030.
9. Mas, G.; Guan, J.-Y.; Crublet, E.; Debled, E. C.; Moriscot, C.; Gans, P.; Schoehn, G.; Macek, P.; Schanda, P.; Boisbouvier, J., Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci Adv* **2018**, *4* (9), eaau4196-  
eaau4196.
10. Wangness, R. K.; Bloch, F., The Dynamical Theory of Nuclear Induction. *Physical Review* **1953**, *89* (4), 728-739.
11. Redfield, A. G., The Theory of Relaxation Processes. In *Advances in Magnetic and Optical Resonance*, Waugh, J. S., Ed. Academic Press: 1965; Vol. 1, pp 1-32.
12. Abragam, A., *The principles of nuclear magnetism*. Oxford: Clarendon press: 1961.
13. Goldman, M., Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. *Journal of Magnetic Resonance (1969)* **1984**, *60* (3), 437-452.
14. Korzhnev, D. M.; Billeter, M.; Arseniev, A. S.; Orekhov, V. Y., NMR studies of Brownian tumbling and internal motions in proteins. *Progress in Nuclear ...* **2001**, *38* (3).
15. Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skelton, N. J., CHAPTER 5 - RELAXATION AND DYNAMIC PROCESSES. In *Protein NMR Spectroscopy (Second Edition)*, Academic Press: Burlington, 2007; pp 333-404.
16. Nicholas, M. P.; Eryilmaz, E.; Ferrage, F.; Cowburn, D.; Ghose, R., Nuclear spin relaxation in isotropic and anisotropic media. *Progress in nuclear magnetic resonance spectroscopy* **2010**, *57* (2), 111-158.
17. Ivanov, K.; Yurkovskaya, A.; Vieth, H.-M., High resolution NMR study of T1 magnetic relaxation dispersion. I. Theoretical considerations of relaxation of scalar coupled spins at arbitrary magnetic field. *The Journal of Chemical Physics* **2008**, *129* (23), 234513.
18. Korchak, S.; Ivanov, K.; Yurkovskaya, A.; Vieth, H.-M., High resolution NMR study of T1 magnetic relaxation dispersion. II. Influence of spin-spin couplings on the longitudinal spin relaxation dispersion in multispin systems. *The Journal of Chemical Physics* **2010**, *133* (19), 194502.

19. Werbelow, L. G.; Grant, D. M., Intramolecular Dipolar Relaxation in Multispin Systems. In *Advances in Magnetic and Optical Resonance*, Waugh, J. S., Ed. Academic Press: 1977; Vol. 9, pp 189-299.
20. McConnell, H. M., Reaction Rates by Nuclear Magnetic Resonance. *The Journal of Chemical Physics* **1958**, *28* (3), 430-431.
21. Carr, H. Y.; Purcell, E. M., Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review* **1954**, *94* (3), 630-638.
22. Meiboom, S.; Gill, D., Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **1958**, *29* (8), 688-691.
23. Palmer, A. G., NMR Characterization of the Dynamics of Biomacromolecules. *Chemical Reviews* **2004**, *104* (8), 3623-3640.
24. Korzhnev, D. M.; Kay, L. E., Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Accounts of Chemical Research* **2008**, *41* (3), 442-451.
25. Vallurupalli, P.; Bouvignies, G.; Kay, L. E., Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *Journal of the American Chemical Society* **2012**, *134* (19), 8148-8161.
26. Forsén, S.; Hoffman, R. A., Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. *The Journal of Chemical Physics* **1963**, *39* (11), 2892-2901.
27. Gupta, R. K.; Redfield, A. G., Double nuclear magnetic resonance observation of electron exchange between ferri- and ferrocytochrome c. *Science* **1970**, *169* (3951), 1204-1206.
28. Cayley, P.; Albrand, J.; Feeney, J.; Roberts, G.; Piper, E.; Burgen, A., Nuclear magnetic resonance studies of the binding of trimethoprim to dihydrofolate reductase. *Biochemistry* **1979**, *18* (18), 3886-3895.
29. Ward, K.; Aletras, A.; Balaban, R. S., A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *Journal of magnetic resonance* **2000**, *143* (1), 79-87.
30. Van Zijl, P. C.; Yadav, N. N., Chemical exchange saturation transfer (CEST): what is in a name and what isn't? *Magnetic resonance in medicine* **2011**, *65* (4), 927-948.
31. Bouvignies, G.; Vallurupalli, P.; Kay, L. E., Visualizing Side Chains of Invisible Protein Conformers by Solution NMR. *Journal of Molecular Biology* **2014**, *426* (3), 763-774.
32. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531-1534.
33. Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W., NMR-Based Screening of Proteins Containing 13 C-Labeled Methyl Groups. *Journal of the American Chemical Society* **2000**, *122* (33), 7898-7904.
34. Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E., Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* **2003**, *41* (10), 843-852.
35. Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E., Cross-correlated relaxation enhanced 1H-13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society* **2003**, *125* (34), 10420-10428.
36. Tugarinov, V.; Kay, L. E., An isotope labeling strategy for methyl TROSY spectroscopy. *Journal of biomolecular NMR* **2004**, *28* (2), 165-172.
37. Tugarinov, V.; Sprangers, R.; Kay, L. E., Line narrowing in methyl-TROSY using zero-quantum 1H-13C NMR spectroscopy. *Journal of the American Chemical Society* **2004**, *126* (15), 4921-4925.

38. Bolik-Coulon, N.; Cousin, S. F.; Kadeřávek, P.; Dumez, J.-N.; Ferrage, F., Understanding the methyl-TROSY effect over a wide range of magnetic fields. *The Journal of Chemical Physics* **2019**, *150* (22), 224202.
39. Li, X.; Wang, X.; Snyder, M., Systematic investigation of protein–small molecule interactions. *IUBMB Life* **2013**, *65* (1), 2-8.
40. Matsuda, R.; Bi, C.; Anguizola, J.; Sobansky, M.; Rodriguez, E.; Badilla, J.; Zheng, X.; Hage, B.; Hage, D. S., Studies of metabolite–protein interactions: A review. *Journal of Chromatography B* **2014**, *966*, 48-58.
41. Yang, G.; Li, X.; Snyder, M., Investigating metabolite–protein interactions: An overview of available techniques. *Methods* **2012**, *57* (4), 459-466.
42. Maity, S.; Gundampati, R.; Kumar, T., NMR Methods to Characterize Protein-Ligand Interactions. *Natural Product Communications* **2019**, *14* (5).
43. Chen, A.; chemistry, S.-M. J., Peer Reviewed: Affinity NMR. *Analytical chemistry* **1999**.
44. Johnson, C. S., Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1999**, *34* (3-4), 203-256.
45. Gonnella, N.; Lin, M.; Shapiro, M. J.; Wareing, J. R.; Zhang, X., Isotope-Filtered Affinity NMR. *Journal of Magnetic Resonance* **1998**, *131* (2), 336-338.
46. London, R. E., Theoretical Analysis of the Inter-Ligand Overhauser Effect: A New Approach for Mapping Structural Relationships of Macromolecular Ligands. *Journal of Magnetic Resonance* **1999**, *141* (2), 301-311.
47. Post, C., Exchange-transferred NOE spectroscopy and bound ligand structure determination. *Current Opinion in Structural Biology* **2003**, *13* (5), 581-588.
48. Chen, A.; of the Society, S.-M. J., NOE pumping: a novel NMR technique for identification of compounds with binding affinity to macromolecules. *Journal of the American Chemical Society* **1998**.
49. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angewandte Chemie International Edition* **1999**, *38* (12), 1784-1788.
50. Vogtherr, M.; Peters, T., Application of NMR Based Binding Assays to Identify Key Hydroxy Groups for Intermolecular Recognition. *Journal of the American Chemical Society* **2000**, *122* (25), 6093-6099.
51. Meyer, B.; Peters, T., NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angewandte Chemie International Edition* **2003**, *42* (8), 864-890.
52. Viegas, A.; Manso, J. o.; Nobrega, F. L.; Cabrita, E. J., Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education* **2011**, *88* (7), 990-994.
53. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B., WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular NMR* **2001**, *21* (4), 349-359.
54. Anorado, E.; Galli, G.; Ferrante, G., Fast-field-cycling NMR: applications and instrumentation. *Applied Magnetic Resonance* **2001**, *20* (3), 365-404.
55. Kimmich, R.; Anorado, E., Field-cycling NMR relaxometry. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2004**, *44* (3-4), 257-320.
56. Steele, R. M.; Korb, J. P.; Ferrante, G.; Bubici, S., New applications and perspectives of fast field cycling NMR relaxometry. *Magnetic Resonance in Chemistry* **2016**, *54* (6), 502-509.

57. Madl, T.; Mulder, F. A., Small Paramagnetic Co-solute Molecules. In *Paramagnetism in Experimental Biomolecular NMR*, 2018; pp 283-309.
58. Luchinat, C.; Parigi, G., Collective Relaxation of Protein Protons at Very Low Magnetic Field: A New Window on Protein Dynamics and Aggregation. *Journal of the American Chemical Society* **2007**, *129* (5), 1055-1064.
59. Borsi, V.; Luchinat, C.; Parigi, G., Global and Local Mobility of Apocalmodulin Monitored through Fast-Field Cycling Relaxometry. *Biophysical Journal* **2009**, *97* (6), 1765-1771.
60. Parigi, G.; Rezaei-Ghaleh, N.; Giachetti, A.; Becker, S.; Fernandez, C.; Blackledge, M.; Griesinger, C.; Zweckstetter, M.; Luchinat, C., Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **2014**, *136* (46), 16201-16209.
61. Persson, E.; Halle, B., Nanosecond to Microsecond Protein Dynamics Probed by Magnetic Relaxation Dispersion of Buried Water Molecules. *Journal of the American Chemical Society* **2008**, *130* (5), 1774-1787.
62. Charlier, C.; Khan, S. N.; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F., Nanosecond Time Scale Motions in Proteins Revealed by High-Resolution NMR Relaxometry. *Journal of the American Chemical Society* **2013**, *135* (49), 18665-18672.
63. Cousin, S. F.; Bolik-Coulon, N.; Gu, Y.; Charlier, C.; Carlier, L.; Bruschiweiler-Li, L.; Marquardsen, T.; Tyburn, J.-M.; Bruschiweiler, R.; Ferrage, F., Time-resolved protein side-chain motions unraveled by high-resolution relaxometry and molecular dynamics simulations. *Journal of the American Chemical Society* **2018**.
64. Kriat, M.; Confort-Gouny, S.; Vion-Dury, J.; Sciaky, M.; Viout, P.; Cozzone, P. J., Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. *NMR in Biomedicine* **1992**, *5* (4), 179-184.
65. Barrilero, R.; Ramirez, N.; Vallvé, J.; Taverner, D.; Fuertes, R.; Amigó, N.; Correig, X., Unravelling and quantifying the "NMR-invisible" metabolites interacting with human serum albumin by binding competition and T2 relaxation-based decomposition analysis. *Journal of Proteome Research* **2017**.
66. Cousin, S. F.; Charlier, C.; Kadeřávek, P.; Marquardsen, T.; Tyburn, J.-M.; Bovier, P.-A.; Ulzega, S.; Speck, T.; Wilhelm, D.; Engelke, F.; Maas, W.; Sakellariou, D.; Bodenhausen, G.; Pelupessy, P.; Ferrage, F., High-resolution two-field nuclear magnetic resonance spectroscopy. *Physical Chemistry Chemical Physics* **2016**, *18* (48), 33187-33194.
67. Cousin, S. F.; Kadeřávek, P.; Haddou, B.; Charlier, C.; Marquardsen, T.; Tyburn, J.-M. M.; Bovier, P.-A. A.; Engelke, F.; Maas, W.; Bodenhausen, G.; Pelupessy, P.; Ferrage, F., Recovering Invisible Signals by Two-Field NMR Spectroscopy. *Angewandte Chemie (International ed. in English)* **2016**, *55* (34), 9886-9889.
68. Kadeřávek, P.; Strouk, L.; Cousin, S. F.; Charlier, C.; Bodenhausen, G.; Marquardsen, T.; Tyburn, J.-M.; Bovier, P.-A.; Engelke, F.; Maas, W.; Ferrage, F., Full Correlations across Broad NMR Spectra by Two-Field Total Correlation Spectroscopy. *ChemPhysChem* **2017**, *18* (19), 2772-2776.
69. McKay, R. T., How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts in Magnetic Resonance Part A* **2011**, *38A* (5), 197-220.
70. Weckwerth, W., *Metabolomics: methods and protocols*. Humana Press: 2007.


71. Bell, J. D.; Brown, J. C. C.; Kubal, G.; letters, S.-P. J., NMR-invisible lactate in blood plasma. *FEBS letters* **1988**.
72. Fasano, M.; Curry, S.; Terreno, E.; Galliano, M.; Fanali, G.; Narciso, P.; Notari, S.; Ascenzi, P., The extraordinary ligand binding properties of human serum albumin. *IUBMB life* **2005**, *57* (12), 787-796.
73. Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S., The Human Serum Metabolome. *PLoS ONE* **2011**, *6* (2).
74. Tatlidil, D.; Ucuncu, M.; Akdogan, Y., Physiological concentrations of albumin favor drug binding. *Physical Chemistry Chemical Physics* **2015**, *17* (35), 22678-22685.
75. Delaglio, F.; Grzesiek, S.; Bax, A.; Vuister, G. W.; Zhu, G.; Pfeifer, J., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995**, *6* (3), 277-293.
76. John W. Eaton, D. B., Søren Hauberg, Rik Wehbring, GNU Octave version 4.2.0 manual: a high-level interactive language for numerical computations. **2016**.
77. Haario, H.; Laine, M.; Mira, A.; Saksman, E., DRAM: Efficient adaptive MCMC. *Statistics and Computing* **2006**, *16* (4), 339-354.
78. Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M., Atomic-resolution dynamics on the surface of amyloid- $\beta$  protofibrils probed by solution NMR. *Nature* **2011**, *480* (7376), 268.
79. Bouvignies, G.; Kay, L. E., Measurement of Proton Chemical Shifts in Invisible States of Slowly Exchanging Protein Systems by Chemical Exchange Saturation Transfer. *The Journal of Physical Chemistry B* **2012**, *116* (49), 14311-14317.
80. Weckwerth, W., Metabolomics in systems biology. *Annual review of plant biology* **2003**, *54* (1), 669-689.
81. Idle, J. R.; Gonzalez, F. J., Metabolomics. *Cell metabolism* **2007**, *6* (5), 348-351.
82. Lindon, J. C.; Holmes, E.; Nicholson, J. K., Metabonomics in pharmaceutical R & D. *The FEBS journal* **2007**, *274* (5), 1140-1151.
83. Cabrera, W. L.; Knapp, J. S., *Metabolomics: metabolites, metabonomics, and analytical technologies*. Nova Science Publishers: 2011.
84. Patti, G. J.; Yanes, O.; Siuzdak, G., Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews Molecular cell biology* **2012**, *13* (4), 263.
85. Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C., High-Throughput Metabolomics by 1D NMR. *Angewandte Chemie International Edition* **2019**, *58* (4), 968-994.
86. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols* **2007**, *2* (11), 2692-2703.
87. Lindon, J. C.; Holmes, E.; Nicholson, J. K., Metabonomics Techniques and Applications to Pharmaceutical Research & Development. *Pharmaceutical Research* **2006**, *23* (6), 1075-1088.
88. Everett, J. R., Chapter 10 - Applications of Metabonomics in Clinical Pharmaceutical R&D. In *The Handbook of Metabonomics and Metabolomics*, Lindon, J. C.; Nicholson, J. K.; Holmes, E., Eds. Elsevier Science B.V.: Amsterdam, 2007; pp 279-287.

89. Yee, A.; Chang, X.; Pineda-Lucena, A.; Wu, B.; Semesi, A.; Le, B.; Ramelot, T.; Lee, G. M.; Bhattacharyya, S.; Gutierrez, P., An NMR approach to structural proteomics. *Proceedings of the National Academy of Sciences* **2002**, *99* (4), 1825-1830.
90. de Graaf, R. A.; Behar, K. L., Quantitative <sup>1</sup>H NMR spectroscopy of blood plasma metabolites. *Analytical chemistry* **2003**, *75* (9), 2100-2104.
91. Wishart, D. S., Quantitative metabolomics using NMR. *TrAC trends in analytical chemistry* **2008**, *27* (3), 228-237.
92. Shin, J.; Lee, W.; Lee, W., Structural proteomics by NMR spectroscopy. *Expert review of proteomics* **2008**, *5* (4), 589-601.
93. Gowda, G. N.; Raftery, D., Can NMR solve some significant challenges in metabolomics? *Journal of Magnetic Resonance* **2015**, *260*, 144-160.
94. Dona, A. C.; Kyriakides, M.; Scott, F.; Shephard, E. A.; Varshavi, D.; Veselkov, K.; Everett, J. R., A guide to the identification of metabolites in NMR-based metabolomics/metabolomics experiments. *Computational and structural biotechnology journal* **2016**, *14*, 135-153.
95. Nagana Gowda, G.; Raftery, D., Recent advances in NMR-based metabolomics. *Analytical chemistry* **2016**, *89* (1), 490-510.
96. Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P., NMR for sample quality assessment in metabolomics. *New Biotechnology* **2019**.
97. Johnson, C. H.; Ivanisevic, J.; Siuzdak, G., Metabolomics: beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology* **2016**, *17* (7), 451-459.
98. Yugi, K.; Kubota, H.; Hatano, A.; Kuroda, S., Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers. *Trends in Biotechnology* **2016**, *34* (4), 276-290.
99. Morozov, V. N.; Morozova, T.; Johnson, K. L.; Naylor, S., Parallel determination of multiple protein metabolite interactions using cell extract, protein microarrays and mass spectrometric detection. *Rapid Communications in Mass Spectrometry* **2003**, *17* (21), 2430-2438.
100. Roelofs, K. G.; Wang, J.; Sintim, H. O.; Lee, V. T., Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proceedings of the National Academy of Sciences* **2011**, *108* (37), 15528-15533.
101. Li, X.; Gianoulis, T. A.; Yip, K. Y.; Gerstein, M.; Snyder, M., Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **2010**, *143* (4), 639-650.
102. Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P., A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **2018**, *172*.
103. Takahashi, N.; Ortel, T. L.; Putnam, F. W., Single-chain structure of human ceruloplasmin: the complete amino acid sequence of the whole molecule. *Proceedings of the National Academy of Sciences* **1984**, *81* (2), 390-394.
104. Schwaeble, W.; Zwirner, J.; Schulz, T. F.; Linke, R. P.; Dierich, M. P.; Weiss, E. H., Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver. *European journal of immunology* **1987**, *17* (10), 1485-1489.
105. Refetoff, S., Inherited thyroxine-binding globulin abnormalities in man. *Endocrine Reviews* **1989**, *10* (3), 275-293.
106. Bhagavan, N. V., *Medical biochemistry*. Academic press: 2002.
107. Pepys, M. B.; Hirschfield, G. M., C-reactive protein: a critical update. *The Journal of clinical investigation* **2003**, *111* (12), 1805-1812.

108. Colombo, S.; Buclin, T.; Decosterd, L. In *Orosomuroid plasma concentration and genetic variants: effects on HIV protease inhibitor clearance and cellular accumulation*, 7th International Workshop on Clinical Pharmacology of HIV Therapy, 2006; pp 20-22.
109. Wobeto, V. P. d. A.; Zaccariotto, T. R.; Sonati, M. d. F., Polymorphism of human haptoglobin and its clinical importance. *Genetics and Molecular Biology* **2008**, *31* (3), 602-620.
110. Donato, L. J.; Jenkins, S. M.; Smith, C.; Katzmann, J. A.; Snyder, M. R., Reference and Interpretive Ranges for  $\alpha$ 1-Antitrypsin Quantitation by Phenotype in Adult and Pediatric Populations. *American journal of clinical pathology* **2012**, *138* (3), 398-405.
111. Soleimani, N.; Hosseinzadeh, M.; Habibagahi, Z., Value of serum glucose-6-phosphate isomerase in patients with rheumatoid arthritis and correlation with disease activity: A case-control study. *Journal of education and health promotion* **2019**, *8*.
112. Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Schlörb, C.; Schwalbe, H., NMR spectroscopic detection of protein protons and longitudinal relaxation rates between 0.01 and 50 MHz. *Angewandte Chemie International Edition* **2005**, *44* (15), 2223-2225.
113. Robertson, A.; Brodersen, R., Effect of lactate, pyruvate, acetone, acetoacetate, and  $\beta$ -hydroxybutyrate on albumin binding of bilirubin. *The Journal of pediatrics* **1983**, *102* (3), 433-438.
114. Chatham, J. C.; Forder, J. R., Lactic acid and protein interactions: implications for the NMR visibility of lactate in biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1999**, *1426* (1), 177-184.
115. Nicholson, J. K.; O'Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sönksen, P. H., Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochemical Journal* **1984**, *217* (2), 365-375.
116. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of intrinsically disordered regions and proteins. *Chemical reviews* **2014**, *114* (13), 6589-6631.
117. Wright, P. E.; Dyson, J. H., Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **2014**, *16* (1).
118. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z., Intrinsic disorder and protein function. *Biochemistry* **2002**, *41* (21), 6573-6582.
119. Basile, W.; Salvatore, M.; Bassot, C.; Elofsson, A., Why do eukaryotic proteins contain more intrinsically disordered regions? *bioRxiv* **2018**, 270694.
120. Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, K. A.; Gough, J., D2P2: database of disordered protein predictions. *Nucleic Acids Research* **2013**, *41* (D1).
121. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; of molecular ..., W.-R. M., Intrinsically disordered protein. *Journal of molecular ...* **2001**.
122. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **1999**, *293* (2), 321-331.
123. Chatterjee, N.; Walker, G. C., Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis* **2017**, *58* (5), 235-263.
124. Zhang, F.; Wen, Y.; Guo, X., CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics* **2014**, *23* (R1), R40-R46.
125. Schipler, A.; Iliakis, G., DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Research* **2013**, *41* (16), 7589-7605.



126. Chang, H. H. Y.; Pannunzio, N. R.; Adachi, N.; Lieber, M. R., Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* **2017**, *18* (8), 495-506.
127. Lieber, M. R., The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry* **2010**, *79* (1), 181-211.
128. Pannunzio, N. R.; Watanabe, G.; Lieber, M. R., Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry* **2018**, *293* (27), 10512-10523.
129. Brouwer, I.; Sitters, G.; Candelli, A.; Heerema, S. J.; Heller, I.; de, A. J.; Zhang, H.; Normanno, D.; Modesti, M.; Peterman, E. J. G.; Wuite, G. J. L., Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature* **2016**, *535* (7613), 566.
130. Williams, G. J.; Hammel, M.; Radhakrishnan, S.; Ramsden, D.; Lees-Miller, S. P.; Tainer, J. A., Structural insights into NHEJ: Building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. *DNA Repair* **2014**, *17*, 110-120.
131. Zhang, Z.; Zhu, L.; Lin, D.; Chen, F.; Chen, D. J.; Chen, Y., The three-dimensional structure of the C-terminal DNA-binding domain of human Ku70. *Journal of Biological Chemistry* **2001**, *276* (41), 38231-38236.
132. Hu, S.; Pluth, J. M.; Cucinotta, F. A., Putative binding modes of Ku70-SAP domain with double strand DNA: a molecular modeling study. *Journal of molecular modeling* **2012**, *18* (5), 2163-2174.
133. Hammel, M.; Yu, Y.; Mahaney, B. L.; Cai, B.; Ye, R.; Phipps, B. M.; Rambo, R. P.; Hura, G. L.; Pelikan, M.; So, S., Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *Journal of Biological Chemistry* **2010**, *285* (2), 1414-1423.
134. Gell, D.; Jackson, S. P., Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Research* **1999**, *27* (17), 3494-3502.
135. Falck, J.; Coates, J.; Jackson, S. P., Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **2005**, *434* (7033), 605-611.
136. Dobbs, T. A.; Tainer, J. A.; Lees-Miller, S. P., A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair* **2010**, *9* (12), 1307-1314.
137. Villarreal, S. A.; Stewart, P. L., CryoEM and image sorting for flexible protein/DNA complexes. *Journal of Structural Biology* **2014**, *187* (1), 76-83.
138. Ma, Y.; Pannicke, U.; Schwarz, K.; Lieber, M. R., Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell* **2002**, *108* (6), 781-794.
139. Malu, S.; De Ioannes, P.; Kozlov, M.; Greene, M.; Francis, D.; Hanna, M.; Pena, J.; Escalante, C. R.; Kurosawa, A.; Erdjument-Bromage, H.; Tempst, P.; Adachi, N.; Vezzone, P.; Villa, A.; Aggarwal, A. K.; Cortes, P., Artemis C-terminal region facilitates V(D)J recombination through its interactions with DNA Ligase IV and DNA-PKcs. *Journal of Experimental Medicine* **2012**, *209* (5), 955-963.
140. Mahaney, B. L.; Hammel, M.; Meek, K.; Tainer, J. A.; Lees-Miller, S. P., XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochemistry and Cell Biology* **2012**, *91* (1), 31-41.
141. Nemoz, C.; Ropars, V.; Frit, P.; Gontier, A.; Drevet, P.; Yu, J.; Guerois, R.; Pitois, A.; Comte, A.; Delteil, C.; Barboule, N.; Legrand, P.; Baconnais, S.; Yin, Y.; Tadi, S.; Barbet-Massin, E.; Berger, I.; Le Cam, E.; Modesti, M.; Rothenberg, E.; Calsou, P.; Charbonnier, J. B., XLF and

- APLF bind Ku80 at two remote sites to ensure DNA repair by non-homologous end joining. *Nat Struct Mol Biol* **2018**, *25* (10), 971-980.
142. Andres, S. N.; Vergnes, A.; Ristic, D.; Wyman, C.; Modesti, M.; Junop, M., A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Research* **2012**, *40* (4), 1868-1878.
143. Ahnesorg, P.; Smith, P.; Jackson, S. P., XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* **2006**, *124* (2), 301-313.
144. Buck, D.; Malivert, L.; de Chasseval, R.; Barraud, A.; Fondanèche, M.-C.; Sanal, O.; Plebani, A.; Stéphan, J.-L.; Hufnagel, M.; le Deist, F.; Fischer, A.; Durandy, A.; de Villartay, J.-P.; Revy, P., Cernunnos, a Novel Nonhomologous End-Joining Factor, Is Mutated in Human Immunodeficiency with Microcephaly. *Cell* **2006**, *124* (2), 287-299.
145. Tsai, C. J.; Kim, S. A.; Chu, G., Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proceedings of the National Academy of Sciences* **2007**, *104* (19), 7851-7856.
146. Li, Y.; Chirgadze, D. Y.; Bolanos-Garcia, V. M.; Sibanda, B. L.; Davies, O. R.; Ahnesorg, P.; Jackson, S. P.; Blundell, T. L., Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *The EMBO Journal* **2008**, *27* (1), 290-300.
147. Andres, S. N.; Modesti, M.; Tsai, C. J.; Chu, G.; Junop, M. S., Crystal Structure of Human XLF: A Twist in Nonhomologous DNA End-Joining. *Molecular Cell* **2007**, *28* (6), 1093-1101.
148. Hammel, M.; Rey, M.; Yu, Y.; Mani, R. S.; Classen, S.; Liu, M.; Pique, M. E.; Fang, S.; Mahaney, B. L.; Weinfeld, M.; Schriemer, D. C.; Lees-Miller, S. P.; Tainer, J. A., XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair\* . *The Journal of Biological Chemistry* **2011**, *286* (37), 32638-32650.
149. Ropars, V.; Drevet, P.; Legrand, P.; Baconnais, S.; Amram, J.; Faure, G.; Márquez, J. A.; Piétrement, O.; Guerois, R.; Callebaut, I.; Cam, E.; Revy, P.; de Villartay, J.-P.; Charbonnier, J.-B., Structural characterization of filaments formed by human Xrcc4–Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proceedings of the National Academy of Sciences* **2011**, *108* (31), 12663-12668.
150. Wu, Q.; Ochi, T.; Matak-Vinkovic, D.; Robinson, C. V.; Chirgadze, D. Y.; Blundell, T. L., Non-homologous end-joining partners in a helical dance: structural studies of XLF–XRCC4 interactions. *Biochemical Society Transactions* **2011**, *39* (5), 1387-1392.
151. Marley, J.; Lu, M.; Bracken, C., A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **2001**, *20* (1), 71-75.
152. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325-1327.
153. Van Rossum, G., Python tutorial, May 1995. *CWI Report CS-R9526* **1995**.
154. Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D., Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* **2006**, *15* (12), 2795-2804.
155. Khan, T.; Douglas, G. M.; Patel, P.; Nguyen Ba, A. N.; Moses, A. M., Polymorphism Analysis Reveals Reduced Negative Selection and Elevated Rate of Insertions and Deletions in Intrinsically Disordered Protein Regions. *Genome Biology and Evolution* **2015**, *7* (6), 1815-1826.
156. Gill, M. L.; Byrd, R. A.; Palmer, I. I. A. G., Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region. *Physical Chemistry Chemical Physics* **2016**, *18* (8), 5839-5849.
157. Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M., Identification of Dynamic Modes in an Intrinsically Disordered Protein Using

Temperature-Dependent NMR Relaxation. *Journal of the American Chemical Society* **2016**, *138* (19), 6240-6251.

158. Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B., The role of DNA shape in protein–DNA recognition. *Nature* **2009**, *461* (7268), 1248-1253.

159. Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J., Two-Dimensional NMR Lineshape Analysis. *Scientific Reports* **2016**, *6* (1), 24826.

160. Yoon, J.-H.; Prakash, L.; Prakash, S., Highly error-free role of DNA polymerase  $\eta$  in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. *Proceedings of the National Academy of Sciences* **2009**, *106* (43), 18219-18224.

161. Inui, H.; Oh, K.-S.; Nadem, C.; Ueda, T.; Khan, S. G.; Metin, A.; Gozukara, E.; Emmert, S.; Slor, H.; Busch, D. B.; Baker, C. C.; DiGiovanna, J. J.; Tamura, D.; Seitz, C. S.; Gratchev, A.; Wu, W.; Chung, K.; Chung, H.; Azizi, E.; Woodgate, R.; Schneider, T. D.; Kraemer, K. H., Xeroderma Pigmentosum-Variant Patients from America, Europe, and Asia. *Journal of Investigative Dermatology* **2008**, *128* (8), 2055-2068.

162. Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K., Structural basis for the suppression of skin cancers by DNA polymerase  $\eta$ . *Nature* **2010**, *465* (7301), 1039-1043.

163. Wilson, Ryan C.; Jackson, Meghan A.; Pata, Janice D., Y-Family Polymerase Conformation Is a Major Determinant of Fidelity and Translesion Specificity. *Structure* **2013**, *21* (1), 20-31.

164. Boudsocq, F.; Kokoska, R. J.; Plosky, B. S.; Vaisman, A.; Ling, H.; Kunkel, T. A.; Yang, W.; Woodgate, R., Investigating the Role of the Little Finger Domain of Y-family DNA Polymerases in Low Fidelity Synthesis and Translesion Replication. *Journal of Biological Chemistry* **2004**, *279* (31), 32932-32940.

165. Santoso, Y.; Joyce, C. M.; Potapova, O.; Le Reste, L.; Hohlbein, J.; Torella, J. P.; Grindley, N. D. F.; Kapanidis, A. N., Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proceedings of the National Academy of Sciences* **2010**, *107* (2), 715-720.

166. Washington, M. T.; Prakash, L.; Prakash, S., Yeast DNA Polymerase  $\eta$  Utilizes an Induced-Fit Mechanism of Nucleotide Incorporation. *Cell* **2001**, *107* (7), 917-927.

167. Trincao, J.; Johnson, R. E.; Escalante, C. R.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Structure of the catalytic core of *S. cerevisiae* DNA polymerase  $\eta$ : implications for translesion DNA synthesis. *Molecular cell* **2001**, *8* (2), 417-426.

168. Hansen, A. L.; Kay, L. E., Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. *Journal of Biomolecular NMR* **2011**, *50* (4), 347-355.

169. Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q., Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **2009**, *18* (5), 936-948.

170. Boivin, S.; Kozak, S.; Meijers, R., Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification* **2013**, *91* (2), 192-206.

171. Kozak, S.; Lercher, L.; Karanth, M. N.; Meijers, R.; Carlomagno, T.; Boivin, S., Optimization of protein samples for NMR using thermal shift assays. *Journal of Biomolecular NMR* **2016**, *64* (4), 281-289.