



HAL
open science

Protein Dynamics from Nuclear Spin Relaxation : High-Resolution Relaxometry, Disordered Proteins and Applications to the C-Terminal Region of the Protein

Artemis

Cyril Charlier

► **To cite this version:**

Cyril Charlier. Protein Dynamics from Nuclear Spin Relaxation : High-Resolution Relaxometry, Disordered Proteins and Applications to the C-Terminal Region of the Protein Artemis. Chemical Physics [physics.chem-ph]. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066577 . tel-01333232

HAL Id: tel-01333232

<https://theses.hal.science/tel-01333232>

Submitted on 17 Jun 2016

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.

Université Pierre et Marie Curie

Ecole doctorale 388: Chimie Physique et Chimie Analytique de Paris Centre

Thèse présentée par

Mr. Cyril Charlier

En vue d'obtenir le grade de

Docteur en Sciences de l'Université Pierre et Marie Curie

Soutenance prévue le

03 juillet 2015

Protein Dynamics from Nuclear Spin Relaxation: High-Resolution Relaxometry, Disordered Proteins and Applications to the C-Terminal Region of the Protein Artemis



Jury	Prof. A.G. Palmer	Rapporteur
	Prof. Frans Mulder	Rapporteur
	Dr. Guy Lippens	Examineur
	Prof. Geoffrey Bodenhausen	Examineur
	Dr. Carine Van-Heijenoort	Examinatrice
	Dr. Sophie Zinn-Justin	Examinatrice
	Dr. Ludovic Carlier	Codirecteur
	Dr. Fabien Ferrage	Directeur

Université Pierre et Marie Curie

Ecole doctorale 388 : Chimie Physique et Chimie Analytique de Paris Centre

Thèse présentée par

Mr. Cyril Charlier

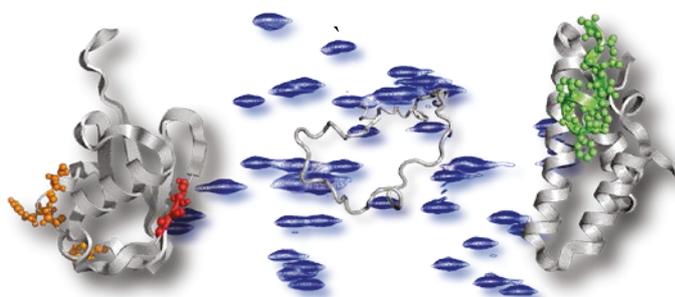
En vue d'obtenir le grade de

Docteur en Sciences de l'Université Pierre et Marie Curie

Soutenance prévue le

03 juillet 2015

Dynamique des Protéines par Relaxation des Spins Nucléaires : relaxométrie haute-résolution, protéines désordonnées et application à la région C-terminal de la protéine Artemis



Jury	Prof. A.G. Palmer	Rapporteur
	Prof. Frans Mulder	Rapporteur
	Dr. Guy Lippens	Examineur
	Prof. Geoffrey Bodenhausen	Examineur
	Dr. Carine Van-Heijenoort	Examinatrice
	Dr. Sophie Zinn-Justin	Examinatrice
	Dr. Ludovic Carlier	Codirecteur
	Dr. Fabien Ferrage	Directeur

À la mémoire de mes grand-parents,
Pierre, Jeanne-Marie,

À la mémoire de mon arrière grand-mère,
Alfreda

Acknowledgments

I am grateful to **Pr. Arthur G. Palmer III** and **Pr. Frans Mulder**, to being rapporteur of my thesis.

I would like to thank **Dr. Sophie Zinn-Justin**, **Dr. Guy Lippens**, **Dr. Carine Van-Heijenoort** and **Prof. Geoffrey Bodenhausen** for being in my committee.

Mes remerciements en premier lieu à mon directeur de thèse **Fabien Ferrage** pour son soutien et son enthousiasme durant ces trois années passées au laboratoire. Fabien, je voudrais te remercier de m'avoir fait confiance tout au long de ces 3 années. Ces recherches menées à tes côtés m'auront fait grandir tant scientifiquement qu'humainement. Merci pour ton soutien sans faille en toute circonstance et pour tes qualités humaines. Enfin, je dois admettre que le seul domaine où je n'aurais pas réussi à apprendre de toi, c'est la cuisine ! Tu restes le maître !

Je voudrais également remercier mon co-directeur de thèse, **Ludovic Carlier**, pour son soutien et son investissement durant les deux années où nous avons travaillé ensemble. Ludovic, je peux le dire sans toi ma thèse aurait été différente. Certes d'un point de vue production de protéine tu m'auras tout appris mais je retiendrais également ta bonne humeur quotidienne. Enfin, merci pour tous les bons moments partagés ensemble notamment lors de nos séances de course à pied, même si souvent j'ai dû t'attendre ou regarder dans le « rétroviseur » !!

Enfin, je voudrais remercier **Philippe Pelupessy** pour son aide précieuse durant l'ensemble de ces trois années. Philippe, tu es une personne extraordinaire tant scientifiquement qu'humainement. Merci pour ta patience et tes explications devant le spectromètre mais également pour ton aide générale durant cette thèse. Un grand merci !

Je tiens à remercier **Geoffrey Bodenhausen** de m'avoir accueilli dans le laboratoire. Geoffrey, merci pour tes conseils et pour les discussions que nous avons pu avoir pendant mon séjour au laboratoire.

Un merci tout particulier à mes collègues de bureau **Shahid** et **Samuel**. Shahid, merci pour tout ce que tu m'auras apporté pendant cette thèse 3 années passés à tes côtés. Tu

m'auras appris énormément de choses. Samuel, un grand merci à toi pour ces années que nous avons partagé à l'ENS, à Karlsruhe ou en Californie!! Nos discussions, nos moments de joies et de galères partagés autour de cafés auront rendu cette thèse très sympathique à tes côtés.

Un grand merci à **Nicola Salvi** pour son aide et sa patience sur mathematica. Un grand merci de m'avoir accueilli à Boston, ma première visite aux Etats-Unis mais surtout ce match de NBA: Boston Celtics vs Chicago Bulls !!!

Je tiens à remercier les différents collaborateurs chez Bruker BioSpin avec qui j'ai pu travailler durant ces années. **Thorsten Marquadsen, Frank Engelke et Jean-Max Tyburn**, merci pour votre travail de développement mais aussi pour votre accueil durant tous mes voyages à Karlsruhe.

Je remercie également l'ensemble des membres de mon équipe qui ont rendu cette thèse des plus agréables, **Daniel Abergel, Piotr Tekely, Luminita Duma, Lucie Khemtouri, Emeric Miclet, Olivier Lequin, Jean-Jacques Lacapère, Isabelle Correia, Dimitris Sakellariou, Akansha Sehgal, Lucie Caillon, Anaïs Terrien, Anaïs Hoffmann, Ching-Yu Chou, Lucile Senicourt, Aude Sadet, Pavel Kadeřávek, Rudra Purusottam, Sina Marhabaie, Aditya Jhajharia, Julien Bouclon et Victoire Déjean.**

Un grand merci à tous les doctorants et personnels du LBM que j'ai croisé durant ces trois ans. Un merci tout particulier à **Sandrine Sagan, Astrid Walrant et Rodrigue Marquant** pour avoir participé scientifiquement à ces travaux. Enfin, un grand merci à **Eliane Moulinié et Karine Gherdi** pour leur aide dans les aspects administratifs.

Je souhaiterais remercier les personnes que j'ai pu rencontrer durant mes différents stages qui m'ont donné la motivation, la passion et l'envie de faire de la recherche: **Guy Lippens, Denis Merlet, Nathalie Sibille, Marie-Chistine Schermann, Isabelle Billaut, Nicolas Giraud, Jean-François Bodart**,... Je tiens également à avoir une pensée particulière pour **Jean-Michel** avec qui j'ai fait mes premiers spectres RMN.

De plus, je remercie l'ensemble des amis qui ont été présents tout au long de cette thèse pour les bons moments que nous avons pu partager !

Un merci spécial à l'ensemble de ma belle famille pour votre soutien et votre bonne humeur ! Joëlle et Marcel, merci pour avoir rendu ces trois années plus agréables ! Elo, Jee,

Cedric, H el ene, Alex, Anne ainsi que les « monstres » Paul, Emma, Florian et Mathis un seul mot....merci !

Un immense merci   mes grand-parents Adrienne et Jean! Merci de m'avoir accompagn  et soutenu durant ces ann es. Vous avoir   mes c t s est une chance et j'esp re qu'elle le restera encore longtemps.

Voil  le temps de remercier mes parents, Lo titia et Didier! Par o  commencer, je ne sais pas, tellement vous m'avez apport  durant ces trois ans mais aussi de fa on plus g n rale depuis 26 ans ! Votre soutien sans faille au quotidien m'aura permis de devenir ce que je suis aujourd'hui ! Soud , nous l'avons  t  durant toutes ces ann es m me si la d faite pongiste pouvait m'emporter ! Alors je tiens   vous remercier du fond du c ur, de m'avoir  lev  (tr s haut il faut bien le dire !),  duqu  et soutenu dans ce tout que j'ai entrepris tant d'un point de vue personnel que professionnel. MERCI !!!!

Enfin, comme on dit «*last but not least*», je souhaiterais remercier mon « tit billy ». Ton manuscrit de th se mentionnait que les r les allaient bient t s'inverser et c'est chose faite. Ils se sont invers s et en plus j'ai eu la chance que tu sois   mes c t s tout au long de ces trois ann es. Un soutien sans limite ! Que dire de ces trois ann es qui auront chang  ma vie mais aussi notre vie ! En effet, le 21 juin 2014 nous avons uni nos chemins. La recherche nous a permis de nous rencontrer et va nous permettre de vivre de nouvelles choses magnifiques dans un futur proche.

Cyril

Abstract

The intimate relation between the dynamics and function of biomolecules is nowadays obvious. Indeed, proteins are very dynamic over a broad range of timescales. NMR is a unique technique to extract information on both structure and dynamics of biomolecules. Recent advances of NMR spectroscopy have substantially extended the description of protein dynamics and their relation to biological functions. The work presented here, focuses on the study of pico- to nanosecond timescale motions in proteins using nuclear spin relaxation and in particular studying the reorientation of backbone amide bond vectors in proteins. Conventional high-field relaxation measurements provide information on dynamics with a limited sampling of the density of motions.

We have used a new shuttle device designed for high-resolution relaxometry to measure relaxation rates over nearly two orders of magnitude of magnetic field. The application of high-resolution relaxometry provides unprecedented insights on internal dynamics in ubiquitin. A protocol of correction, called ICARUS, (Iterative Correction and Analysis of Relaxation Under Shuttling) has been developed in order to analyze relaxation rates. In particular it is essential to take into account relaxation pathways when the sample is in the stray field of the magnet. Defying expectations, we found that the nanosecond motions of the $\beta 1$ - $\beta 2$ turn can be revealed by relaxation measurements, in agreement with studies of residual dipolar couplings and accelerated molecular dynamics.

Over the last decades a new class of protein challenging the “structure-function paradigm” has emerged: Intrinsically Disordered Proteins (IDPs). These proteins exhibit a broad conformational ensemble due to the lack of a stable tri-dimensional conformation. However, the interpretation of relaxation data measured on IDPs requires the use of adapted models of protein dynamics. We used a fragment of the transcription factor Engrailed 2 to develop a new approach for the interpretation of relaxation data measured on IDPs called IMPACT (Interpretation of Motions by a Projection onto an Array of Correlation Times). IMPACT consists in the reconstruction of the distribution of correlation times without relying on a particular model of motions. Our results reveal the distribution of pico- to nanosecond timescale motions in the disordered region of Engrailed 2.

In parallel, we have studied a disordered domain of the DNA-repair factor Artemis. This protein, crucial for the immune system, interacts with the DNA Binding Domain of Ligase IV (DBD-LigIV). We have optimized the sample preparation and assigned backbone resonances in Artemis. We have then conducted a study of the interaction between Artemis and the DBD-LigIV. We have shown the role of 5 additional amino acids past the C-terminus of the previously identified interaction site and characterized the kinetics and thermodynamics of the interaction by NMR relaxation dispersion and isothermal titration calorimetry. Independently, we performed a series of relaxation measurements using high-resolution relaxometry (from 0.2 T to 6 T) and high-field relaxation measurements (from 11.4 T to 22.3 T). This combination of high-resolution relaxometry and high-field relaxation measurements will expand our understanding of ps-ns dynamics in intrinsically disordered proteins.

Overall, this work contributes to the development of new concepts for the interpretation of extensive nuclear spin relaxation data in proteins.

Keywords: [Biomolecular NMR, Protein Dynamics, Intrinsically Disordered Proteins (IDPs), Relaxation, High-Resolution Relaxometry, ubiquitin, Engrailed, Artemis]

Content

ACKNOWLEDGMENTS	VII
ABSTRACT	X
CONTENT	12
RÉSUMÉ	14
1. INTRODUCTION	16
2. DÉVELOPPEMENT ET APPLICATION DE LA RELAXOMÉTRIE À L'UBIQUITINE	18
3. MOUVEMENTS RAPIDES, PS-NS, DANS LES IDPS	24
4. INTERACTION ET DYNAMIQUE DE LA PROTÉINE ARTEMIS	27
GENERAL INTRODUCTION	36
1. PROTEINS: STRUCTURE, DYNAMICS, AND FUNCTION	38
2. INTRINSICALLY DISORDERED PROTEINS	40
3. THESIS OUTLINE	46
CHAPTER 1: NUCLEAR SPIN RELAXATION: CONCEPTS & METHODS	47
1. BLOCH EQUATIONS	50
2. THE SOLOMON EQUATIONS	54
3. BLOCH-WANGSNES-REDFIELD (BWR) THEORY	56
4. CHEMICAL EXCHANGE	65
5. EXPERIMENTAL METHODS	70
CHAPTER 2: INTERNAL DYNAMICS IN UBIQUITIN EXPLORED BY HIGH-RESOLUTION RELAXOMETRY	76
1. INTRODUCTION	78
2. UBIQUITIN	80
3. DESCRIPTION OF THE PNEUMATIC SHUTTLE SYSTEM	81
4. ¹⁵ N RELAXATION MEASUREMENTS	86
5. PRESENTATION OF THE ICARUS PROTOCOL	92
6. APPLICATION TO UBIQUITIN	100
7. CONCLUSION	107
CHAPTER 3: INTERPRETATION OF MOTIONS IN DISORDERED PROTEINS	108
1. INTRODUCTION	110
2. ENGRAILED	111
3. HIGH-FIELD RELAXATION MEASUREMENTS	114
4. REDUCED SPECTRAL DENSITY MAPPING	116
PRINCIPLES OF IMPACT	119
5.	119
6. APPLICATION OF IMPACT TO ENGRAILED	122
7. COMPARISON WITH 2CT AND 3CT ANALYSES	126
8. CONCLUSION	134
CHAPTER 4: INTERACTION AND DYNAMICS OF THE C- TERMINAL REGION OF ARTEMIS	137
1. BIOLOGICAL CONTEXT	139

2. SAMPLE PREPARATION OF ART^{LIGIV} AND DBD^{LIGIV}	151
3. BACKBONE ASSIGNMENT OF ARTEMIS	160
4. MAPPING THE INTERACTION BETWEEN ART^{LIGIV} AND DBD^{LIGIV}	167
5. QUANTIFYING PICO- TO NANOSECOND MOTIONS FROM ¹⁵N RELAXATION	180
6. CONCLUSION	188
<u>BIBLIOGRAPHY</u>	189

Résumé

1. Introduction

Le travail de thèse présenté ci-dessous a été réalisé sous la direction des Dr. Fabien Ferrage et Ludovic Carlier au sein de l'équipe « Structure et dynamique des biomolécules » du Laboratoire des biomolécules (LBM, UMR CNRS-UPMC 7203). Les travaux exposés dans ce manuscrit concernent l'application et le développement de nouvelles méthodes afin de caractériser la dynamique des protéines intrinsèquement désordonnées par relaxation des spins nucléaires.

Depuis sa genèse, la biologie structurale des protéines est fondée sur un dogme, aussi appelé paradigme structure-fonction, selon lequel la fonction d'une protéine n'est déterminée que par sa structure. S'appuyant sur cette hypothèse, différentes techniques biophysiques ont été développées afin de mettre en évidence les structures tri-dimensionnelles des protéines, comme la diffraction aux rayons-X de cristaux de protéines ou encore la Résonance Magnétique Nucléaire (RMN). L'exploitation des structures de protéines a permis d'éclairer les mécanismes sous-jacents à la fonction de centaines de protéines. Néanmoins, il est apparu, au cours des dernières décennies, que les mouvements des protéines étaient également essentiels à leur fonction, que ce soit pour la cinétique et la thermodynamique des interactions ou les mécanismes de la catalyse enzymatique, par exemple. A l'heure d'aujourd'hui, grâce aux nombreuses études réalisées, il est devenu évident aux yeux de la communauté scientifique que la dynamique des protéines joue un rôle indissociable de celui de la structure pour leur fonction. Les processus dynamiques sont divers et variés et sont répartis sur une large gamme d'échelles de temps allant des femtosecondes pour les plus rapides jusqu'à la minute voire l'heure pour les plus longs (Figure 1).

Le dogme de la relation structure-fonction a également été remis en cause à la fin du 20^{ème} siècle, avec la mise en évidence de protéines sans structure tri-dimensionnelle stable mais ayant, malgré tout, une fonction bien définie, les protéines et régions intrinsèquement désordonnées. Cette découverte a permis de mettre en lumière un aspect crucial de la dynamique des protéines. L'apport des techniques bioinformatiques et des caractérisations expérimentales a permis de démontrer qu'environ 35 à 51% du protéome eucaryote contient soit des régions soit des protéines complètement intrinsèquement désordonnées. Ces protéines ont été identifiées comme essentielles dans un grand nombre de processus biologiques tels que la signalisation cellulaire, la régulation du cycle cellulaire, la reconnaissance moléculaire,

Résumé

la transcription et la réplication de l'ADN. Certaines de ces protéines sont également impliquées dans le développement de pathologies humaines comme dans les maladies neurodégénératives ou les cancers. La découverte de ces régions ou protéines désordonnées, importantes pour les mécanismes biologiques, a mené le monde scientifique à développer de nouvelles méthodes pour caractériser ces protéines.

Parmi le panel de techniques expérimentales permettant d'étudier la dynamique des protéines, peu d'entre-elles permettent une étude quantitative de la dynamique à l'échelle atomique sur une gamme large de plusieurs ordres de grandeur d'échelles de temps. Grâce à l'ensemble des développements méthodologiques effectués depuis plusieurs décennies, la RMN est devenue la technique de pointe pour étudier la dynamique des protéines. En effet, celle-ci permet d'obtenir des informations précises sur les mouvements dans les protéines au niveau atomique sur quasiment l'ensemble des échelles de temps (Figure 1).

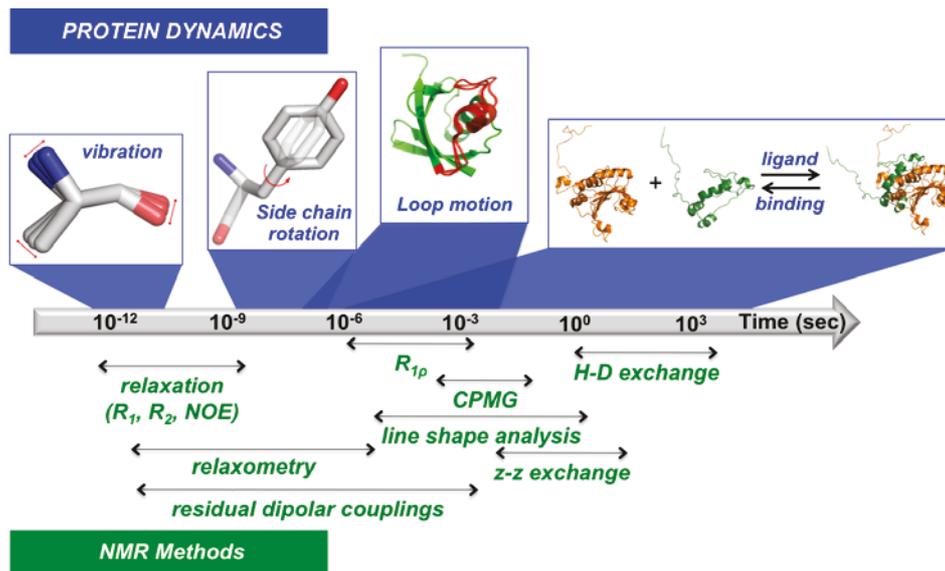


Figure 1 : Echelles de temps des mouvements dans les protéines associées aux méthodes RMN permettant de les analyser.

La Résonance Magnétique Nucléaire est également une technique de choix pour étudier les protéines intrinsèquement désordonnées, en particulier parce qu'elle permet de conserver une information à l'échelle atomique en dépit de l'absence d'ordre. De nouvelles méthodes d'acquisition du signal ont permis de caractériser ces protéines comme par exemple les expériences à haute dimensionnalité ou encore la détection du carbone-13 pour

Cyril Charlier

contrecarrer les nombreux recouvrements sur les spectres. Néanmoins, il est important de noter que dans le cadre de la dynamique, les modèles développés pour les protéines repliées ne sont pas adaptés aux protéines désordonnées.

Dans ce cadre, mon travail de thèse s'est organisé autour de trois grands axes qui seront développés dans la suite de ce résumé:

- Utilisation d'un nouvel appareil et développement d'outils d'analyse permettant de réaliser de la relaxométrie haute résolution et son application à l'ubiquitine.
- Développement d'un nouveau modèle d'interprétation des données de relaxation des spins nucléaires pour les protéines intrinsèquement désordonnées en utilisant comme modèle biologique la protéine Engrailed 2.
- Caractérisation de la dynamique de la protéine intrinsèquement désordonnée Artemis et de son interaction avec la Ligase IV.

2. Développement et application de la relaxométrie à l'ubiquitine

Une protéine est un polymère d'acides aminés. On distingue le squelette peptidique des chaînes latérales, qui sont les groupes substituants des carbones α . Les études de relaxation sur les protéines s'intéressent à différentes interactions dont les fluctuations sont liées à celles de vecteurs internucléaires aussi bien sur le squelette que sur les chaînes latérales. La majorité des études concernant la dynamique des protéines concerne la réorientation du vecteur NH du squelette peptidique des protéines à travers la mesure des vitesses de relaxation de l'azote-15. Les mouvements du vecteur NH peuvent être décrits par la fonction de corrélation. La RMN ne donne pas accès directement à cette fonction mais plutôt à sa transformée de Fourier : la fonction de densité spectrale. Les vitesses de relaxation mesurées dépendent des valeurs de la fonction de densité spectrale aux fréquences propres du système de spin, en particulier la fréquence de Larmor des noyaux le composant.

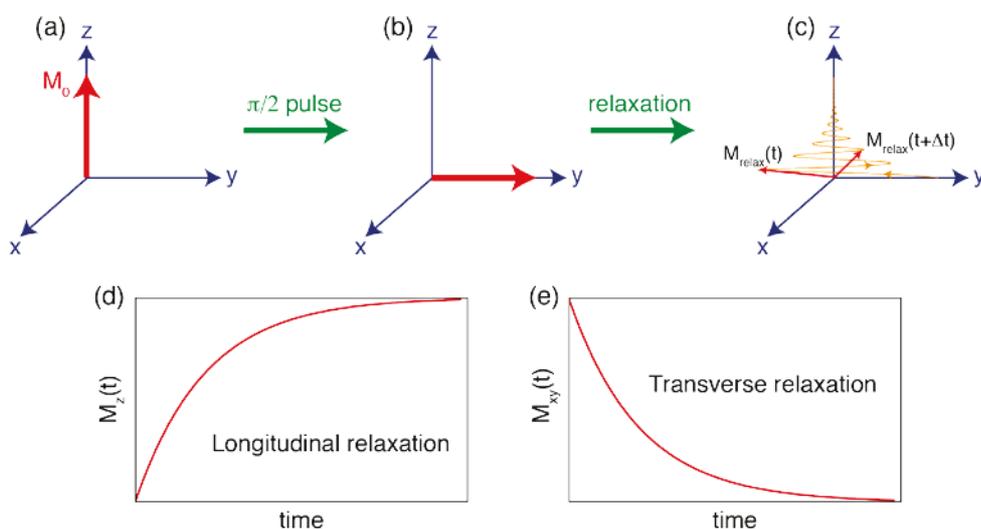


Figure 2 : Principe de la relaxation des spins nucléaires. A l'équilibre l'aimantation se trouve selon l'axe z et peut être transférée dans le plan x-y suite à l'application d'un champ radiofréquence (b). Après l'impulsion rf, l'aimantation retourne à sa position d'équilibre (c). Ce mécanisme de retour à l'équilibre est appelé relaxation. Vectoriellement la relaxation est caractérisée par le retour à l'équilibre de la composante selon z, relaxation longitudinale (d) et dans le plan x-y, relaxation transverse (e).

Le principe d'une expérience RMN est de manipuler les spins nucléaires à l'aide d'impulsions sur différents noyaux comme par exemple le proton, l'azote-15 ou le carbone-13 (Figure 2). Après une perturbation, les spins reviennent à leur état d'équilibre via un phénomène appelé relaxation (Figure 2-a-b-c). Dans le cas de la relaxation des noyaux d'azote-15 du squelette peptidique, trois vitesses de relaxation sont mesurées: vitesse de relaxation longitudinale, R_1 , (Figure 2-d), vitesse de relaxation transverse, R_2 , (Figure 2-e) et vitesse de relaxation croisée dipolaire. Cette dernière vitesse décrit l'échange de polarisation longitudinale entre le noyau d'azote-15 et les protons. On la mesure de manière indirecte en quantifiant l'effet Overhauser nucléaire (NOE). Les récents développements méthodologiques couplés à l'utilisation de champs magnétiques plus intenses ont permis d'obtenir des données de relaxation précises et exactes dans de nombreuses biomolécules. Comme décrits dans la Figure 1, la relaxation permet d'obtenir des informations sur les mouvements rapides, picoseconde à nanoseconde, dans les protéines. Classiquement les vitesses de relaxation (R_1 , R_2 et NOE) sont mesurées en utilisant des hauts champs magnétiques, typiquement de 9.4 à 23.5 T (400 MHz à 1 GHz pour la fréquence de Larmor du proton). L'apport de ces champs magnétiques intenses permet en règle générale d'obtenir un rapport signal sur bruit et une résolution des spectres suffisant pour l'étude de biomolécules. Concernant l'étude des

mouvements rapides, les expériences à haut champ produisent une information sur des gammes de fréquences restreintes. Dans le cas d'une étude concernant le vecteur NH, les données sont obtenues à la fréquence de Larmor de l'azote-15, ω_N (zone verte Figure 3) et autour de la fréquence de Larmor du proton, ω_H (zone grisée Figure 3). La sensibilité aux paramètres des différents modèles dans ces deux zones peut être limitée (Figure 3). Afin de remédier à cette limite, il apparaît nécessaire de décrire ces différentes fonctions de densité spectrale à des fréquences plus faibles. Ceci doit permettre de cartographier de façon plus complète la fonction de densité spectrale et ainsi observer les différences entre les modèles. La technique permettant l'acquisition de données à des champs faibles est appelée relaxométrie.

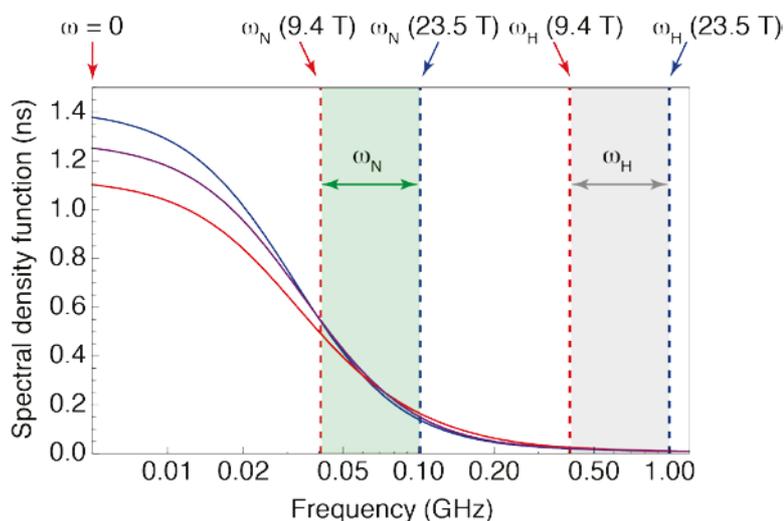


Figure 3 : Fonction de densité spectrale (« Model Free » ou « Extended Model Free ») pour différents jeux de paramètres avec les zones accessibles pour l'azote (vert) et le proton (gris) avec la mesure classique vitesses de relaxation allant de 9.4T (pointillés rouges) à 23.5T (pointillés bleu). Les courbes représentent des modèles de fonction de densité spectrale : en l'absence de mouvements internes (courbe bleue) et en présence de mouvements internes de l'ordre de 2 nanosecondes (courbe rouge et de 10 nanosecondes (Courbe violet).

La relaxométrie haute résolution est une technique développée dans les années 2000 par A. Redfield. Cette technique permet de mesurer des vitesses de relaxation longitudinale à bas champs magnétiques. Pour cela, l'idée développée et appliquée par A. Redfield est d'utiliser le champ de fuite d'un aimant haut champ pour y effectuer la relaxation. L'échantillon est polarisé à haut champ où le terme de spin dont la vitesse de relaxation doit être mesurée est généré. L'échantillon est ensuite transféré physiquement à une position choisie hors de l'aimant pour effectuer la relaxation et finalement il est repositionné à haut

Résumé

champ pour acquérir le signal (Figure 4). Cette méthode combine les intérêts de la sensibilité et de la résolution à haut champ et ceux de la relaxation à bas champ. Dans le cadre de nos travaux, nous avons utilisé et contribué à développer un prototype construit par Bruker Biospin. Ce système pneumatique permet de transférer rapidement l'échantillon (vitesse de l'ordre de 10 m/s) depuis le haut champ (position basse suivant l'axe z) vers le bas champ (position haute). La position d'arrêt à bas champ est réglable en fonction du champ magnétique désiré. Contrairement au dernier système de A. Redfield où le transfert de l'échantillon se fait de façon mécanique, le transfert est assuré d'un côté par de l'azote sous pression et, de l'autre, par une sous-pression. Ce prototype a été installé sur un système 600 MHz, Bruker, muni d'une sonde spécialement développée pour cette application. Dans le but de développer le système, les séquences d'impulsions et l'analyse de données, nous avons utilisé un échantillon d'ubiquitine.

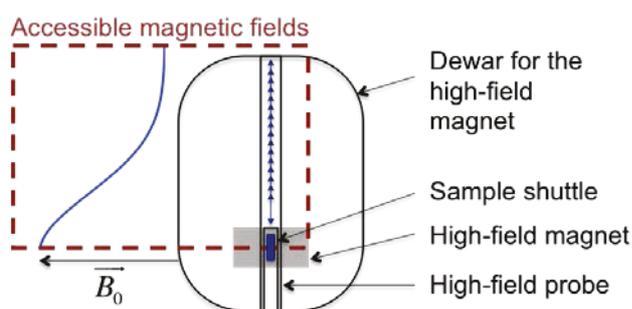


Figure 4 : Principe général de la relaxométrie à haute résolution

L'ubiquitine est une protéine de 8.5 kDa comportant 76 acides aminés. Ses rôles biologiques sont multiples et essentiels : l'ubiquitine est directement impliquée dans de nombreux processus biologiques tels que le cycle cellulaire, la signalisation cellulaire, et aussi dans la dégradation de protéines ayant adopté une mauvaise conformation. La découverte de son rôle dans la dégradation des protéines est l'objet de l'attribution du prix Nobel de chimie 2004 à Aaron Ciechanover, Avram Hershko et Irwin Rose. L'ubiquitine n'agit jamais seule dans ces cascades de signalisation mais avec de nombreux partenaires. Ces diverses interactions sont rendues possibles par sa relative flexibilité. La structure tri-dimensionnelle de l'ubiquitine est parfaitement connue (Figure 5-a) par contre sa dynamique reste un domaine de recherche ouvert, surtout dans la communauté RMN. De plus, cette protéine montre une grande stabilité et une forte solubilité ce qui permet de travailler avec des échantillons concentrés et stables pendant plusieurs mois voire plusieurs années.

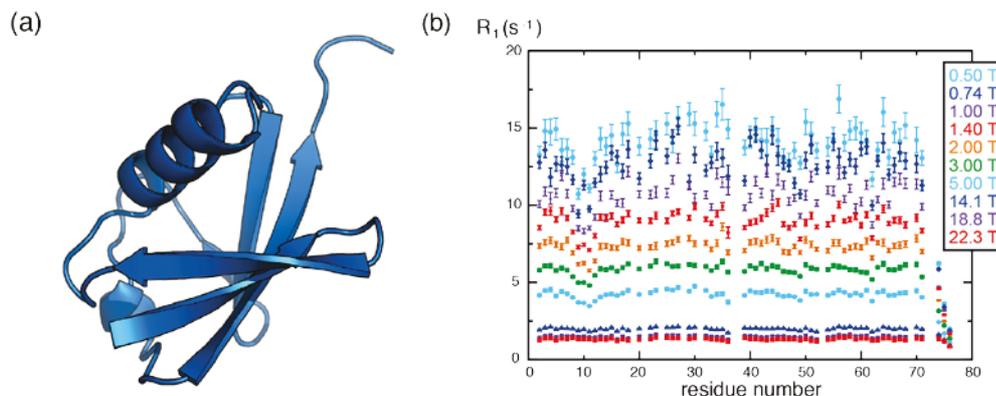


Figure 5 : (a) Structure tri-dimensionnelle de l'ubiquitine (code pdb : 1D3Z) (b) Vitesses de relaxation longitudinale de l'azote-15 mesurées sur l'ubiquitine en fonction du champ magnétique. Du haut vers le bas: $0,5 \text{ T} < B_0^{LF} < 5,00 \text{ T}$ (relaxométrie), $14,1 \text{ T} < B_0^{HF} < 22,3 \text{ T}$ (relaxation à haut champ).

En utilisant notre nouveau système de relaxométrie, nous avons mesuré, sur un échantillon d'ubiquitine à 3 mM, les vitesses de relaxation longitudinale de l'azote-15 pour des champs magnétiques allant de 5 T à 0,5 T (Figure 5-b). Pour compléter notre étude, nous avons mesuré les vitesses de relaxation à 22,3 T, 18,8 T et 14,1 T en utilisant les techniques conventionnelles.

La seconde étape après la mesure des vitesses de relaxation est l'interprétation en terme de paramètres de dynamiques de la protéine. Le passage des données brutes RMN à l'interprétation en terme de dynamique se fait grâce à des modèles de la fonction de densité spectrale. Dans le cas des protéines repliées, les plus populaires d'entre eux, sont le « model free » et son extension « extended mode free ». Le « model free » décrit la dynamique de la molécule avec un temps de corrélation pour la rotation globale, un temps de corrélation effectif pour les mouvements locaux et un paramètre d'ordre qui correspond à la limite aux temps longs de la fonction de corrélation des mouvements du vecteur NH dans le référentiel moléculaire. Quant au « extended model free », il utilise une échelle de temps globale et deux échelles de temps pour les mouvements locaux : une rapide et une lente, ainsi que deux paramètres d'ordre.

Contrairement aux expériences de relaxation qui peuvent être analysées simplement par des méthodes bien établies, celles de relaxométrie nécessitent une analyse de données plus complexe. En effet à partir du moment où l'échantillon quitte sa position à haut champ, il n'est plus possible de manipuler le système de spins à l'aide de champs radiofréquence. Ceci conduit à l'expression de nombreux chemins de relaxation croisée, pendant les transferts et le

délat de relaxation à bas champ. Afin de prendre ces effets en compte, nous avons développé un protocole de correction permettant de corriger les vitesses de relaxation expérimentales : ICARUS (« Iterative Correction and Analysis of Relaxation data Under Shuttling »). Grâce à cela nous avons obtenu des paramètres d'ordre, S^2 , montrant des variations le long de la séquence plus marquées que celles obtenues à partir de données de relaxation à haut champ uniquement (Figure 6-a). Ceci est d'autant plus marqué et visible dans les parties flexibles, comme le coude $\beta 1$ - $\beta 2$ par exemple (Figure 6).

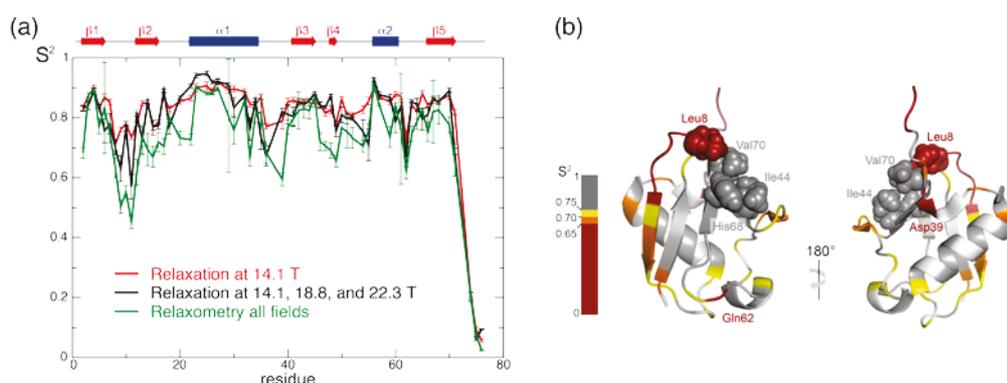


Figure 6 : (a) Paramètres d'ordres S^2 de l'ubiquitine obtenus à partir des vitesses de relaxation de l'azote-15 à 14,1 T uniquement (rouge), à 14,1 T, 18,8 T et 22,3 T (noir) et avec l'ensemble des données de relaxation de 0,5 à 22,3 T (vert). (b) Structure de l'ubiquitine colorée en fonction des valeurs du S^2 . Les résidus les plus mobiles sont en rouge alors que les résidus les plus rigides sont en gris.

Cette étude montre que la relaxométrie à haute résolution, combinée à une analyse de type « ICARUS » nous confère des informations complémentaires sur la dynamique nanoseconde au sein des protéines. Une description détaillée des échelles de temps des mouvements dans les protéines désordonnées n'est toujours pas accessible. Il nous est donc apparu évident d'appliquer la relaxométrie à haute résolution à ce type de protéines. De plus, les temps de relaxation des protéines désordonnées étant plus long, cette technique devrait être encore plus favorable d'un point de vu expérimental (les pertes de polarisation pendant les transferts étant plus courts).

Avant de se lancer dans ce projet, il nous a fallu nous intéresser aux modèles de la fonction de densité spectrale pour le cas des protéines désordonnées. Pour ce type de protéines, il est impossible de parler de temps de rotation globale et d'utiliser uniquement une ou deux échelles de temps internes. Les autres études traitant de la fonction de densité spectrale se tournent vers l'utilisation de distribution d'échelles de temps. Néanmoins,

l'utilisation de ces distributions ou le développement de nouvelles distributions se heurtent au choix de la forme de la distribution. Nous nous sommes donc tournés vers le développement d'une nouvelle approche permettant de décrire la fonction de densité spectrale dans le cadre des protéines désordonnées. Pour réaliser ces développements, nous avons utilisé comme protéine modèle la protéine partiellement désordonnée Engrailed 2.

3. Mouvements rapides, ps-ns, dans les IDPs

Dans le cadre de ce projet, réalisé avec Shahid Nawaz Khan, nous avons travaillé sur le facteur de transcription Engrailed 2. Cette protéine comporte un domaine structuré liant l'ADN (homéodomaine) et une région non structurée d'environ 200 acides aminés (Figure 7). Cette dernière est décrite dans la littérature comme interagissant avec des régulateurs de la transcription, lui conférant un rôle important notamment dans le développement du système nerveux central.

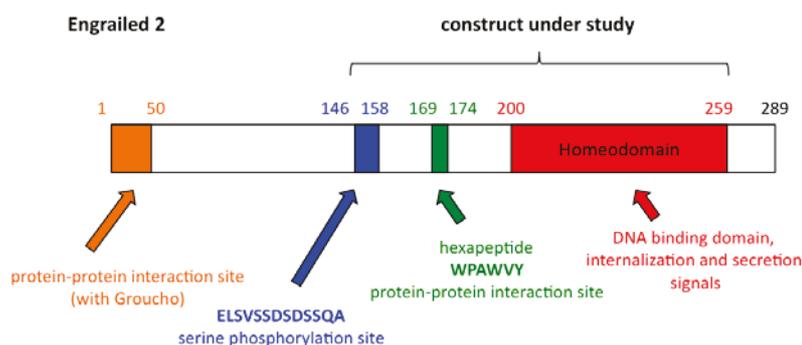


Figure 7 : Régions fonctionnelles de la protéine Engrailed 2. La construction étudiée comprend les résidus 146 à 259. Elle inclut deux régions désordonnées (158-169 et 174-200) ainsi que l'homéodomaine (200-259).

Le travail présenté ci-dessous repose sur l'étude d'une construction comportant une région désordonnée de 54 résidus (146-199) ainsi que l'homéodomaine (200-259) (Figure 7). Une partie du travail de thèse de Shahid a été de comprendre et de décrire l'espace conformationnel d'Engrailed 2. Nous avons également étudié la cinétique d'exploration de cet espace. Pour cela, les vitesses de relaxation de l'azote-15 ont été mesurées à cinq champs magnétiques allant de 9,4 T à 23,5 T (fréquence de Larmor du proton de 400 MHz à 1 GHz) :

- les vitesses de relaxation longitudinale (R_1),
- les vitesses de relaxation transverse soit avec un simple echo (R_2^{echo}) ou un train d'impulsions $\pi \square \square R_2^{\text{cpmg}} \square$,

- les effets Overhauser nucléaires (NOEs),
- les vitesses de relaxation croisée transverse ($\eta_{\square\square}$) et longitudinale (η_{\square}) dues aux fluctuations de l'anisotropie de déplacement chimique de l'azote-15 (^{15}N -CSA) et du couplage dipolaire entre l'azote-15 et le proton amides.

Pour chaque champ magnétique, il est possible d'exploiter ces vitesses de relaxation avec la méthode de cartographie de la fonction de densité spectrale réduite (« reduced spectral density mapping »). Cette méthode permet d'approximer les équations théoriques des vitesses de relaxation R_1 , R_2 et NOEs.

Dans notre travail, nous avons choisi de reconstituer la distribution des temps de corrélation à partir des vitesses de relaxation. Cette reconstruction est analogue aux techniques de régularisation employées pour résoudre les problèmes d'inversion « mal posés ». Nous avons optimisé une grille de temps de corrélation sur laquelle nous projetons la distribution des temps de corrélation. L'ajustement de la fonction de densité spectrale vis-à-vis des données expérimentales se fait simplement sur l'intensité relative des contributions de chacun de ces différents temps de corrélations. Grâce à cette approche, il est possible d'augmenter le nombre de temps de corrélations sans pour autant augmenter le nombre de paramètres à optimiser (Figure 8-a-b). Notre modèle n'est fondé sur aucun modèle physique. La seule hypothèse de l'approche développée est que la fonction de corrélation peut être approximée par une somme d'exponentielles. Celle-ci mène à une fonction de densité spectrale de type :

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

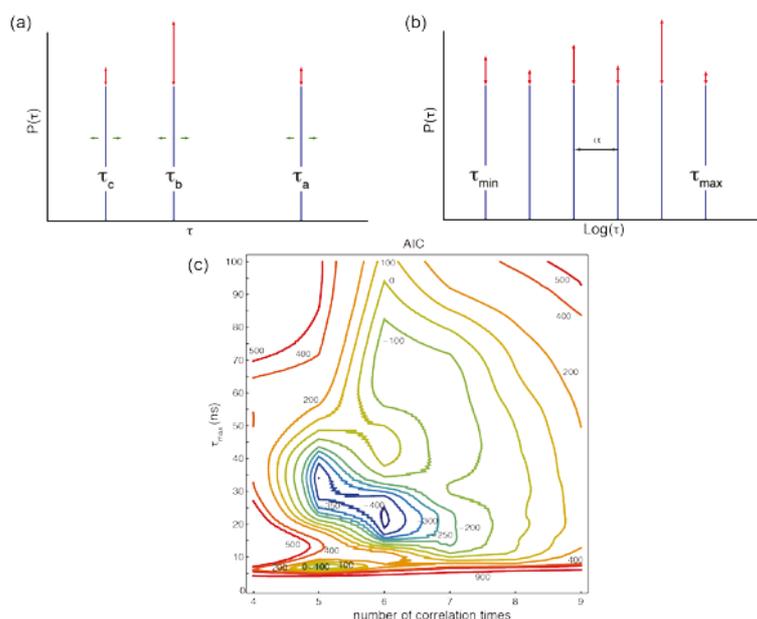


Figure 8 : Principe et optimisation de l’approche IMPACT. (a) Dans l’analyse 3CT (« extended model free ») les temps de corrélations et leurs contributions relatives doivent être ajustés. (b) Dans IMPACT les valeurs des temps de corrélations sont fixées et espacées régulièrement sur une échelle logarithmique, seul leurs contributions relatives doivent être ajustées. (c) Optimisation de l’approche IMPACT. « AICc » est montrée. Le nombre de temps de corrélations à été exploré de $n=4$ à 9 et la plage $[\tau_{\min}, \tau_{\max}]$ de temps de corrélation caractérisé par IMPACT fut exploré de $[1\text{ps}, 1\text{ns}]$ à $[100\text{ps}, 100\text{ns}]$.

La première étape d’IMPACT (« Interpretation of Motions by a Projection onto an Array of Correlation Times ») consiste en l’optimisation du de la grille de temps de corrélation : leur nombre (n) ainsi que l’intervalle $[\tau_{\min}, \tau_{\max}]$. Une des questions critiques dans l’approche IMPACT concerne la sélection du modèle. Nous avons optés pour une optimisation utilisant le critère d’information de Akaike (« AIC »). Cette étape permettant d’évaluer le critère statistique de chaque jeux de paramètres $\{n, [\tau_{\min}, \tau_{\max}]\}$. Cette optimisation, appliquée au données de la protéine Engrailed 2 a permis d’identifier un minimum pour $[\tau_{\min}, \tau_{\max}] = [21\text{ps}, 21\text{ns}]$ avec $n=6$ (Figure 8-c).

Afin d’illustrer nos résultats, nous avons utilisé une figure sous forme de « code barre ». Pour chaque résidu analysé, un histogramme montre la contribution de chaque temps de corrélation (Figure 9). Cette figure montre clairement les différents régimes de dynamique dans Engrailed 2 ainsi que la continuité des variations de la distribution des temps de corrélation. On peut noter, par exemple, la diminution des mouvements très rapides (sous la

nanoseconde) dans l'hexapeptide cordonnée avec l'apparition des mouvements plus lents (supérieures à la nanoseconde). De plus, le changement dans les temps de corrélation est facilement identifiable entre les parties non-structurées avec une majorité de mouvements aux alentours de la nanoseconde et l'homéodomaine avec des mouvements plus longs. De même au sein de l'homéodomaine, structuré en 3 hélices (Figure 9-b), il est possible d'observer les variations au niveau des boucles entre ces hélices.

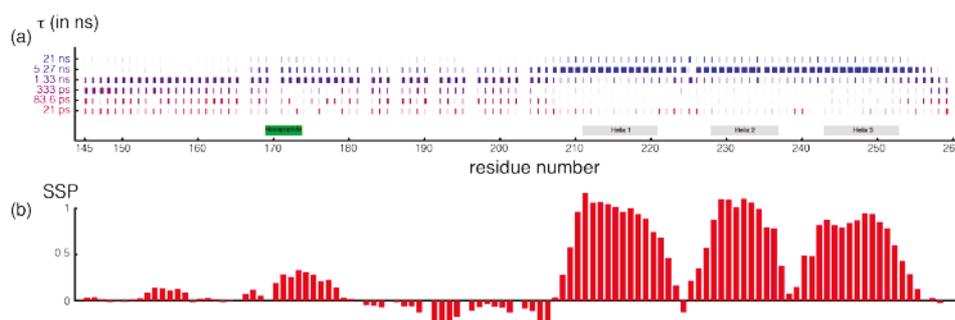


Figure 9 : (a) Représentation graphique de l'analyse IMPACT de la fonction de densité spectrale de la protéine Engrailed 2 (b) SSP score représenté pour la compréhension de la dynamique d'engrailed 2.

Nous avons ici développé une méthode permettant d'analyser la dynamique des protéines désordonnées. Comme nous l'avons montré dans le cas de l'ubiquitine, la relaxométrie pourra certainement permettre de compléter et d'affiner ces informations sur la dynamique d'Engrailed 2. Pour cela il sera nécessaire de coupler les approches ICARUS et IMPACT. Dans le cas de la protéine Engrailed 2 et à l'heure de rédiger ce manuscrit, nous avons uniquement des données à hauts champs. Il est prévu dans les prochains mois de faire l'acquisition de données à des champs plus faibles afin de compléter cette étude.

Parallèlement à ce projet, nous avons commencé l'étude d'une autre protéine désordonnée, la région C-terminale de la protéine Artemis. Ce projet est présenté dans la troisième partie de ce résumé.

4. Interaction et dynamique de la protéine Artemis

La troisième partie de ce travail de thèse concerne l'étude de la protéine Artemis. Le système immunitaire adaptatif est une barrière cruciale que l'organisme a développé au cours de son évolution afin de lutter contre des agressions extérieures. Ce système complexe, est fondé sur la reconnaissance anticorps-antigène. Afin de générer une large diversité d'anticorps, capables de s'adapter aux pathogènes étrangers, les cellules immunitaires, ont

développés un système d'adaptabilité. Ce mécanisme de recombinaison de l'ADN, entre autres, le mécanisme V(D)J, permet de générer une grande diversité de récepteurs de cellules T et d'immunoglobulines. Typiquement les anticorps sont composés de chaînes lourdes et légères, composées de parties constante et d'une partie variable. Les gènes codant les parties variables, sont composée de trois types de segments différents V (« Variable »), D (« Diversité ») et J (« Jonction »). Le principe du mécanisme V(D)J est de sélectionner trois segments V, D et J afin de reconnaître l'antigène. Ce mécanisme divisé en deux étapes est crucial pour le bon fonctionnement du système immunitaire. La première étape consiste en une sélection d'un segment de chaque partie, V, D et J sur le gène et pour cela l'ADN est coupé entre les segments. Cette première étape mène à la formation de cassures double-brin de l'ADN. La seconde étape consiste en une réparation de ces cassures grâce au mécanisme de jonction non homologues des extrémités (NHEJ).

Il a été montré que la protéine Artemis est impliquée dans ces deux mécanismes avec notamment un rôle de nucléase. Naturellement, Artemis a une activité exonucléase. Son interaction et sa phosphorylation par DNA-PKcs lui font acquérir une activité endonucléase. L'absence de cette protéine ou la présence de formes tronquées a des effets néfastes menant à des syndromes d'immunodéficience pouvant être létaux chez les patients. La protéine entière est composée de 692 acides aminés. Elle est divisée en deux parties : une partie structurée sous la forme des domaines β -lactamase et β -CASP et une longue partie désordonnée C-terminal d'environ 300 acides aminés. De plus il a été montré ces dernières années qu'Artemis interagit avec un autre partenaire, la Ligase IV, via un domaine d'interaction de 10 résidus dans la région C-terminale d'Artemis.

Durant ce travail, nous avons abordé deux aspects. Dans un premier temps nous avons étudié l'interaction entre Artemis et la Ligase IV et, dans un second temps, nous avons étudié la dynamique rapide d'Artemis. Ce travail s'effectue en collaboration avec Aneel K. Agarwal et Patricia Cortes du Icahn School of Medicine at Mount Sinai à New-York qui nous ont fournis les plasmides permettant de purifier le fragment 480-575 d'Artemis (Art^{LigIV}) et le domaine de liaison à l'ADN de la ligase IV (^{DBD}LigIV).

Nous avons tout d'abord exprimé et purifié Art^{LigIV} ainsi que ^{DBD}LigIV. Pour les besoins des expériences futures de RMN, nous avons préparé des échantillons d'Artemis uniformément enrichi soit en azote-15 (¹⁵N-Art^{LigIV}) soit en azote-15 et carbone-13 (¹⁵N-¹³C-Art^{LigIV}). Ces marquages ont été réalisés selon le protocole de Marley. La Ligase IV a quant à elle été exprimée et purifiée sans marquage isotopique.

Le premier volet du travail concerne l'interaction entre Art^{LigIV} et ^{DBD}LigIV. Pour approfondir cela, nous avons réalisé un titrage suivi par RMN. En partant d'un échantillon, d'Art^{LigIV} seule, nous avons ajouté étape par étape ^{DBD}LigIV jusqu'à en obtenir 2 équivalents molaires. A chaque étape du titrage, un spectre ¹⁵N-¹H HSQC a été enregistré sur un spectromètre 600 MHz. Deux comportements ont pu être clairement identifiés. Le premier se caractérisant par une disparition des pics (485-499) et le second par un déplacement des signaux au fur et à mesure de l'ajout de ^{DBD}LigIV (483-504-521-552 par exemple). Le premier comportement est caractéristique d'un échange lent entre les deux formes. En effet, dans le cas d'un échange lent, le signal de la forme libre disparaît tandis que celui de la forme liée apparaît. Dans notre cas nous n'avons pu observer la réapparition des pics. Quant au second comportement, il est caractéristique d'un échange rapide à deux sites car la trajectoire des pics est linéaire alors que pour des équilibres plus compliqués les signaux suivent des courbes non linéaires. Grâce à ce titrage, nous avons pu mettre en évidence une zone d'interaction plus longue (485-499), dans laquelle les perturbations structurales sont notables, que celle identifiée dans la littérature (485-495). De plus il est intéressant de noter que les perturbations de déplacement chimique à l'extérieur du domaine d'interaction sont très limitées, ce qui nous amène à croire que l'interaction n'affecte que les résidus du domaine d'interaction. Grâce à ces données, nous avons pu obtenir une constante de dissociation de $1,38 \pm 0,77 \mu\text{M}$. Cette valeur de Kd est légèrement différente de celle publiée dans la littérature qui est de $4,5 \mu\text{M}$. Cela peut s'expliquer par le fait que cette mesure de constante de dissociation a été réalisée uniquement avec le peptide de 10 acides aminés (485-495) représentant le domaine minimal d'interaction. Or, nos résultats suggèrent que les résidus (496-499) sont impliqués directement dans l'interaction.

Pour vérifier notre hypothèse, les deux peptides sPepArt (485-495) et lPepArt (485-500) ont été synthétisés afin de réaliser des expériences de calorimétrie permettant d'obtenir une mesure précise de constante de dissociation entre les différentes formes d'Art^{LigIV} et la ^{DBD}LigIV. La constante de dissociation mesurée pour sPepArt est de $11,7 \pm 0,6 \mu\text{M}$ alors que celle de lPepArt et ^{DBD}LigIV sont respectivement de $2,4 \pm 0,2 \mu\text{M}$ et $2,2 \pm 0,1 \mu\text{M}$. Ces résultats confirment ceux issus du titrage : l'interaction entre Art^{LigIV} et ^{DBD}LigIV implique les résidus 496-499. De plus l'absence de différence de constante de dissociation entre lPepArt et Art^{LigIV} confirme que le reste de la partie désordonnée étudiée (501-575) ne régule pas l'interaction.

Résumé

Nous avons également réalisé des expériences de dispersion CPMG sur Art^{LigIV} seule ainsi que sur Art^{LigIV} + 10% ^{DBD}LigIV afin de caractériser la dynamique milliseconde de l'interaction (Figure 1). La dispersion CPMG est une technique de RMN permettant de caractériser les phénomènes d'échange dit chimique. Le principe est de mesurer la vitesse de relaxation transverse, R_2 , avec une série d'échos de spin $\square\square$. En faisant varier le nombre d'échos et leur durée, on obtient un profil de dispersion complet. Nous avons réalisé ces expériences à 500 MHz avec différents ratios Art^{LigIV}:^{DBD}LigIV (Figure 11). Les équations de Carver-Richards ont été minimisées sur ces données expérimentales et ont permis d'obtenir une vitesse d'échange entre la forme libre et la forme liée d'environ 60 s^{-1} . La minimisation réalisée permet également d'obtenir la différence de déplacement chimique entre la forme libre et liée ainsi que la vitesse de relaxation transverse dans la forme liée. Nous avons également pu montrer que les résidus au cœur de la région interagissant avec ^{DBD}LigIV présentaient des vitesses de relaxation anormalement grandes dans la forme liée, ce qui laisse à penser que le mécanisme de liaison/dissociation implique plus de deux états. L'acquisition de données à différents champs magnétiques ainsi que sur différents noyaux devrait permettre d'éclairer les détails de ce mécanisme.

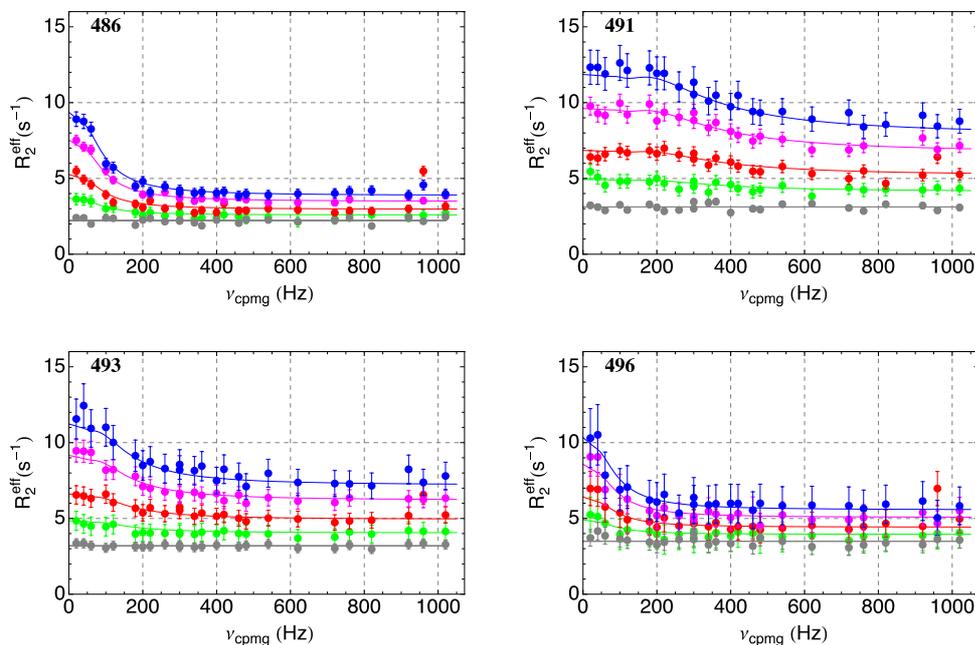


Figure 11 : Courbes de dispersion de relaxation CPMG pour une série de résidus sélectionnés de $\text{Art}^{\text{LigIV}}$ mesurées à 500 MHz avec 0% (gris), 3% (vert), 6% (rouge), 10% (magenta) and 13% (bleu) équivalent molaire de $^{\text{DBD}}\text{LigIV}$. Les lignes solides correspondent à un fit global de la vitesse d'échange et local de la différence de déplacement chimique et de la vitesse de relaxation transverse dans la forme liée.

Le second volet de notre étude d'Artemis concerne la dynamique ps-ns d'Artemis en solution. Pour cela, nous avons obtenu des données de relaxation à différents champs magnétiques permettant de caractériser les mouvements aux échelles de temps piconnanoseconde (Figure 1). Ces expériences de relaxation ont été acquises sur différents échantillons : $^{15}\text{N}\text{-Art}^{\text{LigIV}}$ (Figure 12-a-b), $^{15}\text{N}\text{-Art}^{\text{LigIV}} + 0,1$ équivalent molaire $^{\text{DBD}}\text{LigIV}$ (Figure 12-c-b) et $^{15}\text{N}\text{-Art}^{\text{LigIV}} + 1$ équivalent molaire $^{\text{DBD}}\text{LigIV}$ (Figure 12-e-f). Dans ce résumé, uniquement les vitesses de relaxation longitudinale et les effets Overhauser nucléaires sont présentées. Les valeurs des NOE sont typiques des protéines désordonnées avec des valeurs négatives à des champs magnétiques intermédiaires et faibles (~ 0.5) aux champs élevés (Figure 12-b).

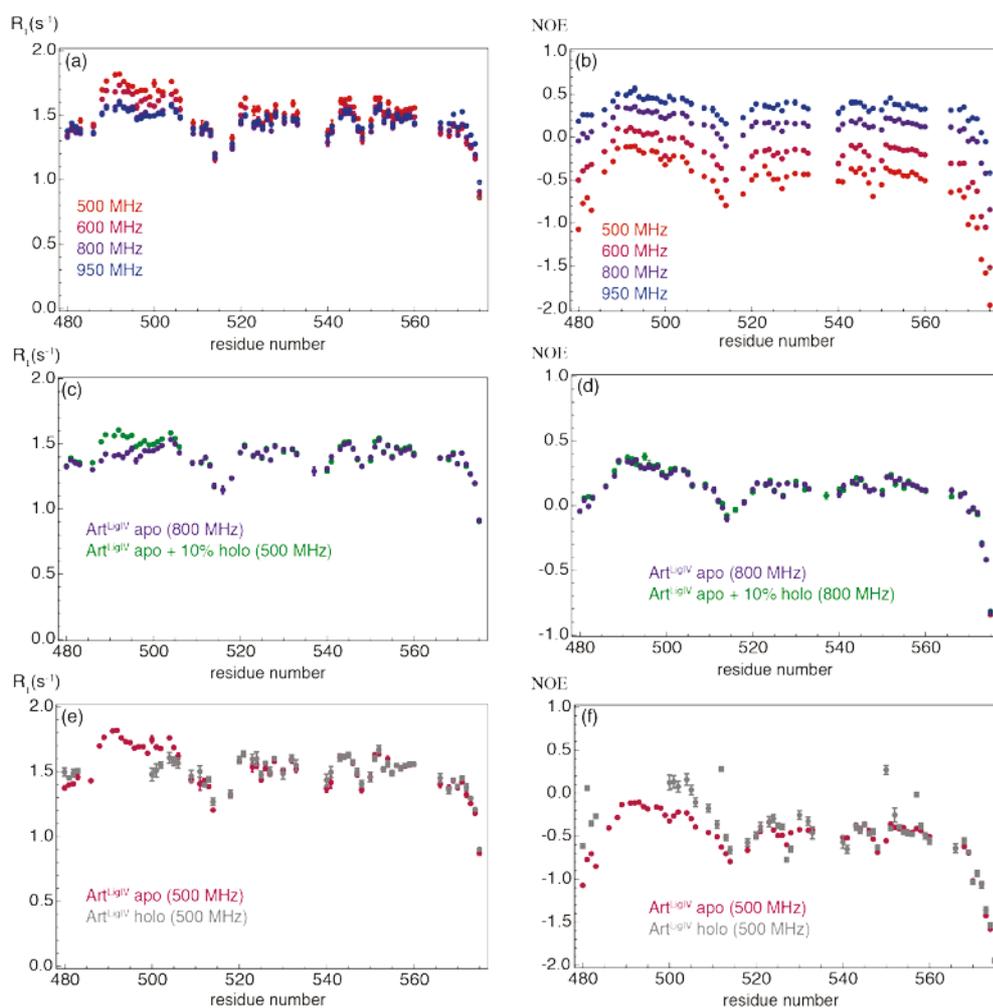


Figure 12 : Vitesses de relaxation de l'azote-15 (R_1 et NOE) pour (a-b) Art^{LigIV}; (c-d) Art^{LigIV} Apo (violet) and Art^{LigIV} Apo + 0.1 eq DBD-LigIV (vert) à 800 MHz; (e-f) Art^{LigIV} (cerise) and Art^{LigIV} Apo + 1eq DBD-LigIV (gris) à 500 MHz.

De plus, comme nous l'avons montré dans le cas de l'ubiquitine, il apparaît important de mesurer les vitesses de relaxation longitudinale à des champs faibles afin de décrire de façon plus complète la fonction de densité spectrale et, notamment, la présence de temps de corrélation nanoseconde. Pour cela, nous avons mesuré avec Ching-Yu Chou une série de vitesses de relaxation longitudinales pour ¹⁵N-Art^{LigIV} (Figure 13) en utilisant la relaxométrie haute résolution de 0,22 à 6 T.

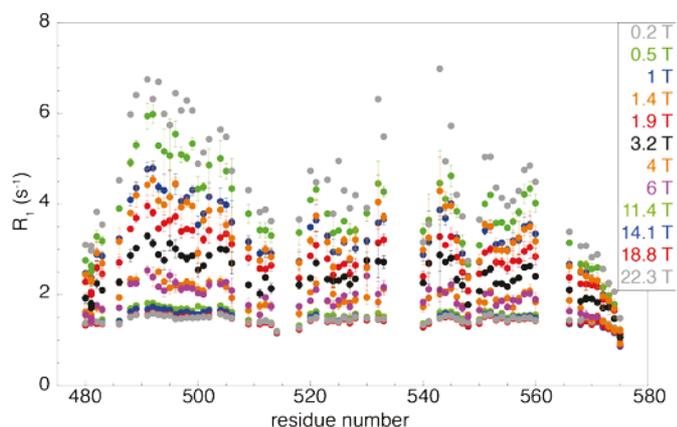


Figure 13 : Vitesses de relaxation longitudinale de l'azote-15 mesurées sur Art^{LigIV} en fonction du champ magnétique. Du haut vers le bas: 0,22 T < B₀^{LF} < 6,00 T (relaxométrie), 11,1 T < B₀^{HF} < 22,3 T (relaxation à haut champ).

Dans le cadre du travail réalisé sur l'ubiquitine, le protocole de correction, ICARUS, était fondé sur l'utilisation de fonctions de densité spectrale de type « model-free » pour extraire les paramètres de dynamiques. Pour les protéines désordonnées, l'approche « model-free » est inadaptée, *a priori*. Par contre l'approche IMPACT, développée pour la protéine Engrailed peut être utilisée pour interpréter les données de relaxation en terme de dynamique. Ainsi il va être nécessaire de développer une approche combinant ICARUS pour la correction des données et IMPACT pour obtenir la fonction de densité spectrale. A l'heure de rédiger ce manuscrit, cette analyse est encore à une étape très préliminaire.

L'ensemble des travaux présentés a permis de développer de nouvelles méthodes pour l'analyse des données de relaxation de spins nucléaires dans les protéines. Les travaux réalisés sur le développement de la relaxométrie à haute résolution et son application à l'ubiquitine ont permis pour la première fois de mesurer des vitesses de relaxation sur deux ordres de grandeurs de champs magnétiques ainsi que de mettre en évidence des mouvements nanoseconde jusqu'alors non identifiés par la relaxation « classique » à haut champ. Avant d'appliquer la technique de relaxométrie haute résolution aux protéines désordonnées, nous avons développé une méthode permettant d'extraire les paramètres de dynamiques des protéines désordonnées à partir de données de relaxation des spins nucléaires, s'affranchissant de l'emploi d'un modèle physique des mouvements dans les protéines. Cette méthode, appelée IMPACT, a été d'abord appliquée au facteur de transcription Engrailed 2 et a permis d'obtenir une vue plus complète des distributions locales de temps de corrélation. La

Résumé

troisième partie de ce travail concerne l'étude d'une protéine désordonnée : Artemis. Nous avons pu caractériser la cinétique et la thermodynamique de son interaction avec la ligase IV. De plus, pour la première fois des vitesses de relaxation à bas champs ont été mesurées en utilisant la relaxométrie haute résolution. L'analyse de ces données et l'extraction des paramètres de dynamiques sont en cours au moment de la rédaction de ce manuscrit.

Cyril Charlier

General Introduction

1. Proteins: Structure, dynamics, and function

Proteins are polymers of tens to hundreds of monomers called amino acids, linked together to form long chains also called polypeptides or polypeptidic chains. There are 20 different natural amino acids, which differ by the substituent of the α carbon, called side-chain. Amino acids are linked by an amide bond, called peptide bond. The ensemble of atoms of the series of peptide planes in a protein chain constitutes the protein backbone. The protein organization is split in four levels:

- The *primary structure* of a protein describes the sequence of the amino acids, somehow as a list.
- The *secondary structure* describes typical regular conformations of the protein backbone. The conformation of the protein backbone can be stabilized by hydrogen bonds to form stable regular structures, mostly alpha helices and beta sheets.
- The *tertiary structure* corresponds to the complete 3D structure of a polypeptide chain. Various constraints such as steric interactions or non-covalent interactions define the tri-dimensional arrangements of atoms of a protein.
- In proteins containing several polypeptide chains, the *quaternary structure* is the overall organization in space of all the chains.

For several decades, a central dogma of molecular biology states that the 3D structure of protein is required for its function. This dogma is called the structure/function paradigm. It was somehow born with the “lock-and-key” model introduced in 1894 by Fisher¹: the substrate of a protein (key) binds to a protein where the binding site has an exactly matching structure (lock). It has been assumed that proteins fold spontaneously into a unique three-dimensional structure, which is the most energetically stable conformation².

The role of proteins is critical for all living systems. They are involved in a broad range of biological functions such as cell cycle regulation, cellular signaling, metabolism, immune responses or cell adhesion for instance. Some proteins are used to catalyze reactions (enzymes), to maintain the shape of cells (actin, myosin). Any change in the sequence of a protein, also called mutation, may dramatically affect the function of the protein. It has been demonstrated that mutations in proteins are crucial for the maintenance of organisms and can

General Introduction

lead in some cases to diseases such as cancers. In order to understand the link between the sequence of a protein and its function, it is essential to be able to determine their structure.

The birth of structural biology may be traced back to the 1950's when Max Perutz and John Kendrew from Cambridge University solved the first 3D structures of proteins by X-Ray crystallography. The low-resolution structure of myoglobin was solved in 1956 while the high-resolution structure appeared in 1959^{3,4}. In addition, in 1959, the low-resolution structure of hemoglobin was described⁵. Since then, the continuous improvements of X-ray crystallography techniques have permitted the determination of thousands of proteins, including much larger 3D structures⁶.

More recently other techniques such as cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR) have been used for structure determination. In addition to electron microscopy, which reveals information at relatively high resolution ($\sim 4 \text{ \AA}$ at best)⁷⁻⁹, NMR provides three-dimensional molecular structures of proteins at atomic resolution. The progress of NMR in the field of the biology started in large part with the work of Kurt Wüthrich and his colleagues during the 1980's¹⁰. Developments of NMR over that last thirty five years have allowed the structure determination of thousands of proteins by solution-state NMR. Although protein size has proven to be a serious limitation for NMR, the structures of some large proteins have been solved by solid-state NMR^{11,12} as well as with liquid-state NMR^{13,14}. The advancement of high-field spectrometers combined with multi-dimensional NMR experiments^{15,16} applied to uniformly or selectively isotopically-labeled samples^{14,17-20} enable the characterization of the structure of a host of macromolecules at atomic resolution.

More protein structures have been solved by NMR than by X-ray crystallography for proteins smaller than 25 kDa but size is much more a limitation for NMR than for X-ray crystallography, so that the vast majority of the structures of proteins larger than 30 kDa have been solved by X-ray crystallography²¹.

A key advantage of NMR, compared to other biophysical techniques, is that it can provide information about protein dynamics at atomic resolution. Indeed a large amount of evidence demonstrated that dynamics is essential for protein function. Such evidence has made evolve the structure-function paradigm toward a new structure-dynamics-function paradigm. Protein motions are crucial for the understanding of protein function, and occur

over a broad range of timescales and length scales from the femtosecond up to minute and hours (Figure 1) and from the picometer to the nanometer. Most biological processes rely on molecular mechanisms that take place in this range of timescales. The NMR community has developed a large variety of experiments to probe this entire range of timescales.

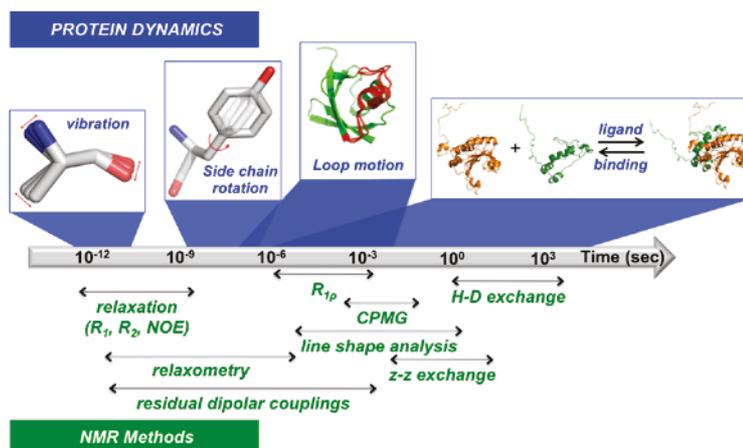


Figure 1: Timescales for protein dynamics and NMR techniques.

2. *Intrinsically disordered proteins*

2.1 *A challenge to the Structure-Function paradigm*

The structure-function paradigm described above was further challenged at the end of the 20st century by the discovery of proteins, now called intrinsically disordered proteins (IDPs), which lack a stable 3D structure in their functional form²². A bioinformatic screening based on protein sequences showed that over 15000 proteins in the *Swiss Protein Database* contain long intrinsically disordered regions (IDRs), with at least 40 consecutive residues²³. Other types of programs estimated that the percentage of proteins with long disordered regions (> 40 amino acids) increases with the complexity of species. Global estimations depend on the computational predictors used: 2.0%, 4.2% and 33% in archea, bacteria and eukaryotes have long disordered regions with 30 or more disordered residues using DISOPRED2 or 23~56%, 15~40% and 25~78% in archea, bacteria and eukaryotes exhibit disordered regions of more than 40 residues with DisEMBL^{24,25}. However, the global trend is consistent. The abundance of IDPs and IDRs in organisms indicates that they are essential for numerous functions.

Here, we will use the terms intrinsically disordered proteins or intrinsically disordered regions but in the early literature on these proteins referred to them as pliable, floppy,

General Introduction

rheomorphic, flexible, mobile, partially folded, natively denatured, natively unfolded, intrinsically unstructured, intrinsically disordered, intrinsically unfolded, vulnerable, chameleon, malleable, 4D, dancing proteins, protein clouds, 3^2 proteins, natively disordered proteins^{26,27}.

Structured proteins can be classified by the structural motifs that constitute them. Classification of IDPs is less straightforward. Any structured protein can be unfolded by denaturation but not all disordered proteins can fold. So it is important to distinguish between the foldable, non-foldable IDPs and conditional IDPs^{28,29}.

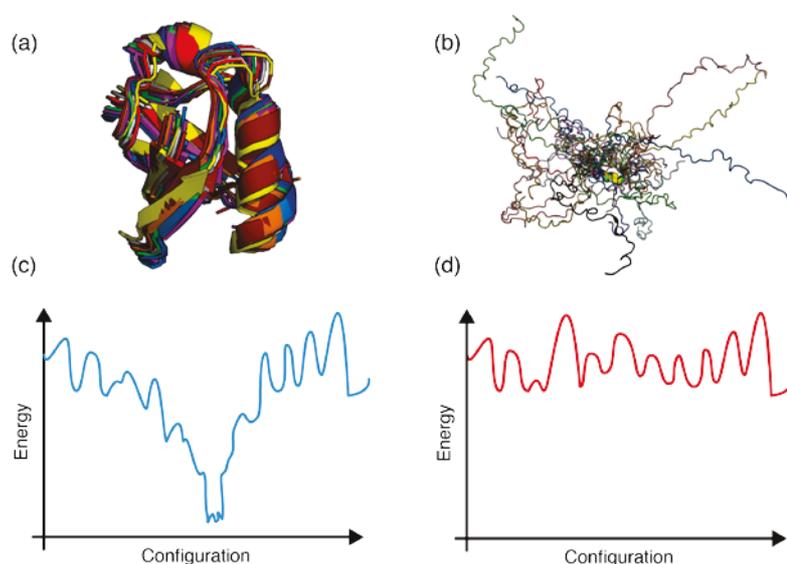


Figure 2: Schematic representations of the conformational ensemble and diagram of folding energy for (a, c) a folded protein (ubiquitin, pdb code: 1d3Z³⁰), and a (b-d) disordered protein (Spinach Thylakoid Soluble Phosphoprotein, pdb code: 2FFT³¹)

2.2 Characterization of intrinsically disordered proteins

The disorder of a protein is encoded in its amino-acid sequence. The sequences of IDPs/IDRs differ substantially from the sequences of structured proteins (Figure 2-a-b). Disordered proteins are often characterized by a low complexity of sequence. In addition, IDPs are significantly depleted of so-called “order promoting” residues either hydrophobic aliphatic (Ile, Leu and Val) or aromatic residues (Trp, Tyr and Phe), which leads to the absence of the hydrophobic core around which folded domains are usually organized. In contrast, IDPs and IDRs are enriched in “disordered promoting” residues, such as polar and charged residues (Arg, Gly, Gln, Ser, Pro, Glu and Lys) as well as structure-breaking residues

Cyril Charlier

(Gly and Pro)^{26,32}. It has been shown that there is a difference in terms of amino acid composition between short (<30 a.a) and long (>30 a.a) disordered regions. Indeed, short disordered regions are more depleted in Ile, Val and Leu and enriched in Gly and Asp. In contrast long disordered regions are more enriched in Lys, Glu and Pro, less enriched in Asp and depleted in Gly and Asn³³.

The absence of hydrophobic core and stable secondary structure is responsible for the enhanced dynamics of IDPs and IDRs. The backbone atoms of well-folded proteins, stabilized by hydrogen-bond networks and organized around hydrophobic cores, fluctuate within a narrow well of the energy landscape (Figure 2-c). In contrast, IDPs are characterized by shallow energy landscapes and significant variations of their Ramachandran angles (Figure 2-d)³². This absence of a real global minimum leads to a system without a dominant well-defined conformation and is an explanation of their higher plasticity and adaptability^{26,34}.

The particular amino acid composition of IDPs opened the way to the design of primary-sequence based predictors of protein disorder. The prediction of the three-dimensional structure of a folded protein from its amino acid composition is still challenging. Folding of a protein *in silico* has seen dramatic progress but is still limited to small and medium size domains and requires important computation resources^{35,36}. It is also possible to predict the structure of a protein when a domain with a known structure shows a high sequence homology. In contrast, predicting the disorder of any region or protein is straightforward³⁷ (even though predicting the whole conformational space of an IDP is still impossible at the time of writing). The first predictor was developed by Romero and co-workers³⁸ and nowadays more than 50 algorithms are available online such as DISOPRED³⁹, DISOPRED2³⁹, PONDR^{23,40-42}. Most of the algorithms designed to predict disordered are very similar and differ mostly from their graphical outputs^{24,26,33,43-45}.

2.3 Functions of intrinsically disordered proteins

The biological significance of IDPs in cells has been largely reviewed in the literature^{22,28,46-50}. It has been showed that IDPs are ubiquitous in the cell and in particular in the nucleus²⁸.

The structural plasticity of IDPs often allows for interactions that are highly specific (with a large interaction surface) yet with weak affinities because of the entropic cost of binding. These interactions may involve numerous different targets sequentially or

General Introduction

concomitantly. Reversible binding and the ability to interact simultaneously with multiple partners are often required in signaling pathway. In parallel, IDPs have been identified in processes as recognition and regulation in cells⁵¹.

Protein functions are also governed by numerous post-transcriptional modifications (PTMs) such as phosphorylation and glycosylation. It has been shown that IDRs in proteins are well-suited targets for PTMs. Indeed, the flexibility of IDRs facilitates the access of enzymes to PTM sites. For instance, the c-Src tyrosine kinase is phosphorylated only when its activation loop is disordered^{52,53}. This behavior is a typical example of the implication in protein-protein interactions and regulation of protein functions. Furthermore, IDPs show an implication as chaperones for both RNA and proteins^{28,54,55}. As previously mentioned, IDRs are present in most types of proteins. Several oncogenes overexpressed or tumor suppressors inactivated in cancer progression have been identified as IDRs or IDPs: p53 or c-Myc, transcription factors c-Fos/c-Jun or cell-cycle inhibitors p27 and p57kip1. Many drivers of the imbalance leading to a cancer are partially or fully disordered without their partners and become structured upon their binding, for instance the protein p53 and its well-known partner Mdm2. In addition, p53 is an illustrative example of multiple-partner interaction hub showing the use of disordered regions to modulate the interactions with other proteins. Indeed, this protein interacts with a large number of partners and plays an important role in many cellular processes. Overall, 70% of the interaction of p53 are mediated by its IDRs and around 90-95% of its PTMs appear in these IDRs²⁶. One of the predictions realized on the link between IDPs/IDRs and cancer pointed out that a large proportion of proteins involved in molecular mechanisms of cancer are IDPs⁵². Understanding the mechanism of action will help in the future to develop potential therapy for cancer focusing on disordered regions⁵². Even if the strategy to block protein-protein interactions to treat disease has not been successful yet, IDPs would be a relevant target for drugs which modulate protein-protein interactions.

2.4 Experimental characterization of intrinsically disordered proteins

Intrinsically disordered proteins have been characterized using several techniques that will be briefly introduced here. The most popular technique to determine the structure of a well-folded protein, X-ray crystallography, can be used to probe disorder in crystallized proteins. Typically, the absence of electron density for part of a protein is a sign of disorder or structural heterogeneity for that region. Thus, evidence of disorder in protein X-ray crystallography is at best a lack of information. Even if a “structure” of a certain crystal could

be obtained, it would only represent a single conformation with little relevance with respect to the large conformational ensemble of the protein.

Biophysical studies of IDPs present unusual experimental challenges due in part to their broad conformational heterogeneity. The combination of structural and biophysical techniques can provide information at the atomic level on the structure, dynamics and the molecular mechanisms underlying the function of a protein. Biophysical techniques such as circular dichroism (CD), Raman or infrared spectroscopies can detect the absence of structure of a protein. Hydrodynamic dimension of the conformational ensemble of disordered proteins can be studied with size-exclusion chromatography or dynamic light scattering⁴³ for instance to provide overall dimensions of the studied IDP. Additional knowledge on compactness of a protein can be extracted from fluorescence experiments including fluorescence resonance energy transfer (FRET)⁴³. However, the combination of small-angle X-ray scattering (SAXS)^{56,57}, single-molecule fluorescence^{34,58} and NMR with computational approaches is particularly powerful for the characterization of IDPs.

Biomolecular NMR is probably the most adapted technique to study IDPs. Unfolded proteins were shown to provide high-resolution spectra about two decades ago⁵⁹⁻⁶⁶, despite a severe drop in resolution as compared to well-folded proteins. The dependence of chemical shifts upon the local primary sequence combined with narrow line width due to slow relaxation provides sufficient resolution to apply typical multidimensional NMR approaches to IDPs and IDRs. The main proof of disorder in a typical ¹⁵N-¹H HSQC experiment, is the low dispersion of the resonances in the proton dimension (Figure 3). In contrast, the somehow preserved dispersions of ¹³C and ¹⁵N chemical shifts are essential for resonance assignments. Due to the higher gyromagnetic ratio of the proton, most of the traditional experiments are based on proton signal detection. To benefit from the dispersion of ¹³C chemical shifts, several groups have developed carbon-detected experiments that appear very well suited for the study of IDPs⁶⁷⁻⁶⁹. A major goal of recent methodological developments adapted to IDPs is to increase the spectral resolution thanks to higher dimensionality (4D, 5D, 6D and 7D) experiments. Such experiments with long transfer delays are possible in part thanks to the long transverse relaxation times encountered in IDPs/IDRs. Combined with fast acquisition schemes and/or non-uniform sampling, such high-dimensionality experiments have allowed for the assignment of backbone and side chains of challenging IDPs with severe overlap in typical 2D experiments^{70,71}.

As discussed above, IDPs cannot be described using a single set of coordinates (conformation) but should be described in terms of continuum in coordinate space. A range of

NMR spectroscopy techniques have been proposed to characterize the conformational space of IDPs^{45,72–77}. Techniques to probe structure and dynamics in IDPs are briefly summarized here:

- Structure propensity of the backbone and local structure information: The first observables on any NMR spectra are the chemical shifts. Chemical shifts depend strongly of the local physicochemical environment of the observed nucleus reflecting the secondary structure in both ordered and disordered proteins⁷⁸. In principle, the deviation of the chemical shift from the random coil value allows for the identification of the secondary structure. Information regarding secondary structure propensity can also be extracted from scalar couplings between nuclei of the protein backbone and can be interpreted in terms of dihedral angles.
- Constraints for conformational ensembles: The inherent flexibility of IDPs creates transient long-range contacts, which can be due to transient and specific binding or due to random encounters. The measurement of nuclear Overhauser effects (NOE) between protons, developed for well-folded proteins is suitable to observe short and medium range contacts. Paramagnetic relaxation enhancements (PRE) are probably the most frequently used technique to probe long-range contacts in IDPs⁷⁶. Finally, residual dipolar couplings (RDC) measured between pairs of nuclei in partially aligned protein solutions provide information on the ensemble of backbone conformers in IDPs. This powerful technique can be combined with other techniques such as chemical shift or PREs to estimate the conformational ensemble^{79,80}.
- Information on dynamics: NMR relaxation measurements are frequently used to provide information on backbone dynamics. ¹⁵N relaxation measurements are employed to characterize ps-ns timescale (with longitudinal relaxation R_1 and heteronuclear NOE) motions, as well as μ s-ms timescale (R_2^{cpmg}) motions, which give rise to chemical exchange (in the NMR sense). In addition, ¹⁵N longitudinal relaxation rates in the rotating frame ($R_{1\rho}$) or the cross-correlated relaxation rate between the ¹⁵N-¹H dipole-dipole and ¹⁵N chemical shift anisotropy interactions (η) can be used to probe chemical exchange and separate the contributions of ps-ns and μ s-ms processes to relaxation. Usually a combination of ¹⁵N- R_1 , ¹⁵N-NOE and ¹⁵N- R_2 are used to obtain a “picture” of dynamics in an IDPs⁸¹. More recently, advances in relaxation dispersion experiments have allowed for the extraction of kinetics information on ligand binding events, revealing transient low populated states⁸².

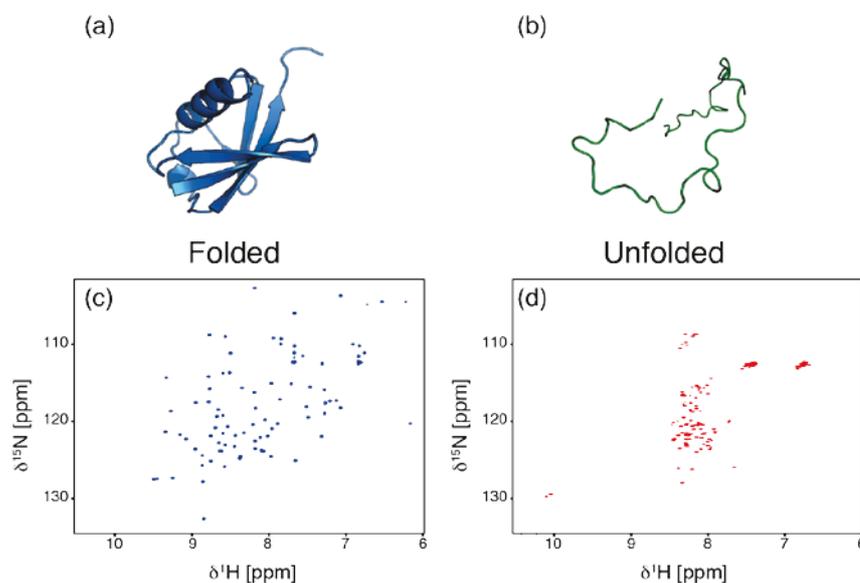


Figure 3: (a) The ribbon representation of ubiquitin (pdb code 1D3Z³⁰) and (b) a schematic representation of a backbone conformation of an IDP are shown on top. 2D ¹H-¹⁵N HSQC spectra acquired on two different proteins: (c) a folded one (ubiquitin) and (d) an intrinsically disordered one (Artemis).

3. Thesis outline

As shown in Figure 1, proteins are dynamic over a broad range of timescales. The present work will be focused on the study of motions in proteins, mostly in the pico- and nanosecond range but also in the μ s-ms range.

In the **first chapter** we introduce nuclear spin relaxation, its measurement and analysis in NMR. We will start from the basic Bloch Equations and present the concepts of relaxation in biomolecules to probe ps-ns motions, we will discuss the most popular models of protein dynamics. In addition, we will discuss some features about chemical exchange as measured in relaxation dispersion experiments. All the methods and pulse sequences used for data acquisition during the present thesis will be presented in this section.

The **second chapter** presents an original approach to perform relaxometry and to interpret high-resolution relaxometry data. We show the development, in close collaboration with Bruker, of an apparatus to perform high-resolution relaxometry on a 600 MHz spectrometer. This method was applied to the model protein ubiquitin. We have developed a

framework for the analysis of such relaxometry datasets. We demonstrate that this technique offers unprecedented access to nanosecond motions in proteins. This chapter is based on⁸³:

Charlier, C. †; Khan, S. N. †; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F. Nanosecond timescale motions in proteins revealed by high-resolution NMR relaxometry; *J. Am. Chem. Soc.* **2013**, *135*, 18665.

In a **chapter 3**, a new model for the interpretation of relaxation measurements in disordered proteins will be discussed. Most of the models of protein dynamics have been developed in the particular case of globular proteins. Here, we introduce an approach, which reconstructs the distribution of correlation times for ps-ns motions, with no requirement of a particular physical model of protein dynamics. This approach, also called IMPACT, has been applied to a transcription factor containing both an IDR and a well-folded domain, Engrailed 2. This chapter is based on:

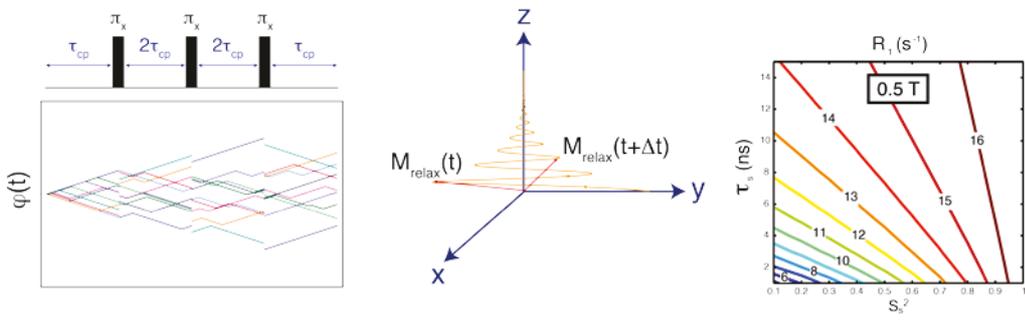
Khan, S. N. †; **Charlier, C.** †; Augustyniak, R.; Salvi, N.; Déjean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F. Distribution of pico-nanosecond motions in disordered proteins from nuclear spin relaxation: a simple array of correlation times; *Under revision*

Finally in **chapter 4**, we report the study of an intrinsically disordered region of the protein Artemis. This project is separated in two parts. We have first studied the interaction between Artemis and its partner Ligase IV. For this purpose, we performed both NMR and isothermal calorimetry (ITC) titrations and we measured transverse relaxation dispersion experiments in order to elucidate the thermodynamics and kinetics of binding. We have also characterized in detail ps-ns motions in the free state of Artemis. Here, we present the first measurements and interpretation of high-resolution relaxometry data for a disordered protein.

Charlier, C.; Pelupessy, P.; Walrant, A.; Marquant, R.; Lavielle, S.; Sagan, S.; Bodenhausen, G.; Cortes, P.; Aggarwal, A. K.; Carlier, L.; Ferrage, F.; Interactions and dynamics of the disordered C-terminal domain of Artemis. *In Preparation*

Charlier, C.; Chou, C.Y.; Pelupessy, P.; Bodenhausen, G.; Cortes, P.; Aggarwal, A. K.; Carlier, L.; Sakellariou, D.; Ferrage, F.; Nanosecond motions of the disordered C-terminal domain of Artemis by a combination of high-resolution relaxometry and high-field relaxation. *In preparation*

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods



In this first chapter we will discuss the concept of relaxation in nuclear magnetic resonance. First, will introduce the Bloch and Solomon Equations as well as Bloch-Wangsness-Redfield theory. The second part will present models of spectral density functions used to describe fast motions in proteins in conventional approaches. The third part will give a brief overview of the mechanisms of relaxation in biomolecules. The fourth part focuses on the study of chemical exchange by NMR in particular using transverse relaxation measurements by relaxation dispersion techniques. In the last section, we will present methods for high-field relaxation measurements used in the present work. This full chapter will follow the approach of Cavanagh *et al.*⁸⁴ and Korzhnev *et al.*⁸⁵

1. Bloch Equations

Relaxation is the irreversible evolution of a spin system towards a steady state. In the absence of perturbation, at thermal equilibrium, magnetizations lie along the z-axis. The basic principle of any NMR experiment is to manipulate this magnetization by applying radio-frequency perturbations also called *rf*-pulses. Relaxation drives the return of the magnetization to the equilibrium position.

1.1 Spin dynamics in the absence of relaxation

The return to equilibrium of the bulk magnetization can be described using the Bloch Equations⁸⁶. The equilibrium state is defined as a state where the populations of energy levels are predicted by the Boltzmann distribution and for which there is *no* transverse component of the magnetization. The vector model is appropriate to describe a system of non-interacting spin 1/2 nuclei in a static field. We will first focus on the vector model in the absence of relaxation. In this formalism, the evolution of the bulk magnetization vector $\mathbf{M}(t)$, which is the sum of all the nuclei momenta in the sample is described by:

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma\mathbf{B}(t) \quad (1.1)$$

where γ magnetic ratio of the nuclei of interest and $\mathbf{B}(t)$ is the magnetic field. In the absence of perturbation, the magnetic field is constant and applied along the z axis of the laboratory frame:

$$\mathbf{B} = \mathbf{B}_0 = \begin{bmatrix} 0 \\ 0 \\ B_0 \end{bmatrix} \quad (1.2)$$

Equation (1.1) can be written as:

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

$$\frac{d}{dt} \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \gamma \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ B_0 \end{bmatrix} = \begin{bmatrix} \gamma M_x B_0 \\ -\gamma M_y B_0 \\ 0 \end{bmatrix} \quad (1.3)$$

The magnetization precesses in a cone with an angular frequency (*Larmor frequency*):

$$\omega_0 = -\gamma B_0 \quad (1.4)$$

The way to perturb the magnetization out of equilibrium is to apply irradiation near the frequency of the system (ω_0) using a radiofrequency field. The magnetic field is the sum of the static field \mathbf{B}_0 and the radiofrequency $\mathbf{B}_{rf}(t)$ also called *rf*-field:

$$\mathbf{B}(t) = \mathbf{B}_0 + \mathbf{B}_{rf}(t) \quad (1.5)$$

The magnetic component of an *rf*-field that is linearly polarized along the x-axis of the laboratory frame is written as:

$$\begin{aligned} \mathbf{B}_{rf}(t) &= 2B_1 \cos(\omega_{rf}t + \phi) \mathbf{i} \\ \mathbf{B}_{rf}(t) &= B_1 \{ \cos(\omega_{rf}t + \phi) \mathbf{i} + \sin(\omega_{rf}t + \phi) \mathbf{j} \} \\ &+ B_1 \{ \cos(\omega_{rf}t + \phi) \mathbf{i} - \sin(\omega_{rf}t + \phi) \mathbf{j} \} \end{aligned} \quad (1.6)$$

where $2B_1$ is the amplitude of the applied *rf*-field, ω_{rf} is the angular frequency of the *rf*-field also called *carrier frequency*, ϕ the field and \mathbf{i} and \mathbf{j} are unit vectors defining the x and y axes respectively. In Equation (1.6) the linearly polarized *rf*-field is decomposed in two circularly polarized components rotating in opposite directions around the z-axis. Only the component rotating in the same sense as the magnetic moment can interact with the nuclear spins. The second component produces a small and neglected effect known as Bloch-Siegert shift, which is proportional to $(B_1/2B_0)^2$. The expression of the *rf*-field (1.6) can be simplified to:

$$\mathbf{B}_{rf}(t) = \mathbf{B}_1(t) = B_1 \{ \cos(\omega_{rf}t + \phi) \mathbf{i} + \sin(\omega_{rf}t + \phi) \mathbf{j} \} \quad (1.7)$$

These Equations can be written in a rotating frame of reference that rotates around z at ω_{rf} . In this rotating frame, $\mathbf{B}_1(t)$ is stationary. In the laboratory frame, the evolution of the magnetization $\mathbf{M}(t)$ is defined as:

$$\frac{d\mathbf{M}}{dt} = \mathbf{M} \times \gamma \mathbf{B}_0 + \mathbf{M} \times \gamma \mathbf{B}_1(t) \quad (1.8)$$

While in the rotating frame, the \mathbf{B}_1 is time independent:

$$\left[\frac{d\mathbf{M}}{dt} \right]_{\text{rot}} = \mathbf{M}_{\text{rot}} \times \gamma \mathbf{B}_0 + \mathbf{M}_{\text{rot}} \times \gamma \mathbf{B}_{1\text{rot}} - \boldsymbol{\omega}_{\text{rf}} \times \mathbf{M}_{\text{rot}} \quad (1.9)$$

This Equation can be written in the rotating frame with a rotation vector $\boldsymbol{\omega}_{\text{eff}}$:

Cyril Charlier

$$\left[\frac{d\mathbf{M}}{dt}\right]_{\text{rot}} = \boldsymbol{\omega}_{\text{eff}} \times \mathbf{M}_{\text{rot}} \quad (1.10)$$

where $\boldsymbol{\Omega}_0 = \boldsymbol{\omega}_0 - \boldsymbol{\omega}_{rf}$ is the offset, ω_1 is the *rf*-field and $\boldsymbol{\omega}_{\text{eff}} = \boldsymbol{\Omega}_0 + \boldsymbol{\omega}_1$ is the *effective field*. Usually we define θ as the *tilt angle*, which is the angle between the z-axis and the effective field \mathbf{B}^{eff} :

$$\tan(\theta) = \frac{\Omega_0}{\omega_1} \quad (1.11)$$

A spin system is manipulated by applying *rf*-field pulses with duration τ_p , leading to a flip angle β :

$$\beta = \omega^{\text{eff}}\tau_p = -\gamma B^{\text{eff}}\tau_p = -\frac{\gamma B_1 \tau_p}{\sin\theta} = \frac{\omega_1 \tau_p}{\sin\theta} \quad (1.12)$$

After the application of an *rf*-pulse, which moves the bulk magnetization out of the z axis, the magnetization precesses around the static field with the Larmor frequency ω_0 . Although this model can be used to describe important features of NMR, the previous Equations lead to the unphysical prediction that, after this *rf*-pulse, the magnetization would evolve in the transverse plane forever.

1.2 Introduction of relaxation in the Bloch Equations

As any physical system following a perturbation the system should return to equilibrium. To take this into account, in 1946, Felix Bloch⁸⁶ introduced modified Equations describing the behavior of a nuclear spin in a magnetic field following the application of a perturbation. Bloch assumed two different first-order relaxation processes: one along the z-axis and one in the x-y plane (Figure 1.1). The former is called longitudinal relaxation or spin-lattice relaxation defined by a rate $R_1=1/T_1$ (Figure 1.1-d). The second, called transverse relaxation or spin-spin relaxation, is defined by the rate $R_2=1/T_2$ (Figure 1.1-e). These two processes are described by the Equations:

$$\begin{aligned} \frac{dM_x(t)}{dt} &= -R_2 M_x(t) \\ \frac{dM_y(t)}{dt} &= -R_2 M_y(t) \\ \frac{dM_z(t)}{dt} &= R_1 [M_0 - M_z(t)] \end{aligned} \quad (1.13)$$

where M_0 is the magnetization at the equilibrium state. These three Equations are solved using an exponential solution:

$$\begin{aligned} M_x(t) &= M_x(0)\exp(-R_2 t) \\ M_y(t) &= M_y(0)\exp(-R_2 t) \\ M_z(t) &= M_0 - [M_0 - M_z(0)]\exp(-R_1 t) \end{aligned} \quad (1.14)$$

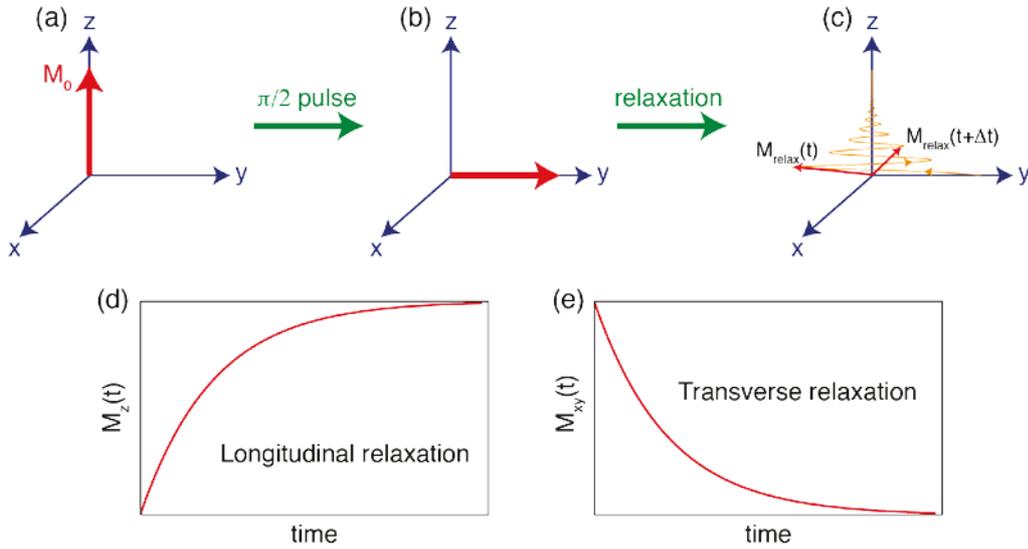


Figure 1.1: Description of the magnetization. (a) At thermal equilibrium the magnetization is along the z-axis. After a perturbation (a $\pi/2$ -pulse here) the magnetization is moved to the x-y plane (c) and finally returned to equilibrium following the Bloch-Equations (c). The longitudinal (d) and transverse (e) components are shown as a function of time following Equations (1.14).

Using Equations (1.1) and (1.13) the Bloch Equations can be written:

$$\begin{aligned} \frac{dM_x(t)}{dt} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_x - R_2 M_x(t) \\ \frac{dM_y(t)}{dt} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_y - R_2 M_y(t) \\ \frac{dM_z(t)}{dt} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_z - R_1 [M_z(t) - M_0] \end{aligned} \quad (1.15)$$

and using the definition of the vector product, the evolution of magnetization in a magnetic field and in absence of *rf-field* is given by:

$$\begin{aligned} \frac{dM_x(t)}{dt} &= \gamma[M_y(t)B_z(t) - M_z(t)B_y(t)] - R_2 M_x(t) \\ \frac{dM_y(t)}{dt} &= \gamma[M_z(t)B_x(t) - M_x(t)B_z(t)] - R_2 M_y(t) \\ \frac{dM_z(t)}{dt} &= \gamma[M_x(t)B_y(t) - M_y(t)B_x(t)] - R_1 [M_z(t) - M_0] \end{aligned} \quad (1.16)$$

To take into account the effect of an *rf-field*, Equations (1.16) can be rewritten in the rotating frame:

$$\frac{d}{dt} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} -R_2 & -\Omega & \omega_1 \sin \phi \\ \Omega & -R_2 & -\omega_1 \cos \phi \\ -\omega_1 \sin \phi & \omega_1 \cos \phi & -R_1 \end{bmatrix} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} + R_1 M_0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1.17)$$

In absence of an applied *rf*-pulse, $\omega_1=0$, for instance during free precession, Equation (1.17) can be simplified as follows:

$$\frac{d}{dt} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} -R_2 & -\Omega & 0 \\ \Omega & -R_2 & 0 \\ 0 & 0 & -R_1 \end{bmatrix} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} + R_1 M_0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1.18)$$

During the *rf*-pulse and if *rf*-pulse length τ_p is short compared to relaxation times $\tau_p \ll T_1, T_2$, Equations (1.17) can be simplified to:

$$\frac{d}{dt} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} 0 & -\Omega & \omega_1 \sin \phi \\ \Omega & 0 & -\omega_1 \cos \phi \\ -\omega_1 \sin \phi & \omega_1 \cos \phi & 0 \end{bmatrix} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} \quad (1.19)$$

The Bloch Equations fail to describe systems more complex than non-interacting spin 1/2 nuclei. In order to take into account interactions Solomon developed an extension of the Bloch Equations⁸⁷.

2. The Solomon Equations

While the Bloch Equations are limited to non-interacting spin 1/2 nuclei, the Solomon Equations⁸⁷ are a convenient way to describe relaxation in more complex spin systems.

2.1 Determination of the Equations of relaxation for a two spin-1/2 system

Here, we consider a two-spin system (I and S). The four energy levels are shown in Figure 1.2 and are labeled with the spin states of the two spins. The rate constants for transitions between the energy levels are denoted by W_0, W_1, W_S and W_2 and are distinguished according to which spins change spin state during the transition:

- W_1 denotes a relaxation process involving an I spin flip
- W_S denotes a relaxation process involving an S spin flip
- W_0 is a relaxation process in which both spins are flipped in opposite senses (flip-flop transitions)
- W_2 is a relaxation process in which both spins are flipped in the same sense (flip-flip transition)

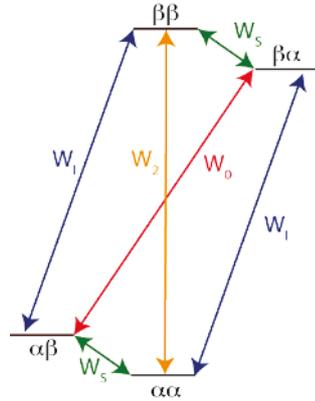


Figure 1.2: Transitions and associated rate constants for a two-spin system

The differential Equations regarding the evolution of the populations n_{ij} for all states are:

$$\begin{cases} \frac{dn_{\alpha\alpha}}{dt} = -(W_s + W_1 + W_2)n_{\alpha\alpha} + W_s n_{\alpha\beta} + W_1 n_{\beta\alpha} + W_2 n_{\beta\beta} \\ \frac{dn_{\alpha\beta}}{dt} = -(W_s + W_1 + W_0)n_{\alpha\beta} + W_s n_{\alpha\alpha} + W_1 n_{\beta\beta} + W_0 n_{\beta\alpha} \\ \frac{dn_{\beta\alpha}}{dt} = -(W_s + W_1 + W_0)n_{\beta\alpha} + W_s n_{\beta\beta} + W_1 n_{\alpha\alpha} + W_0 n_{\alpha\beta} \\ \frac{dn_{\beta\beta}}{dt} = -(W_s + W_1 + W_2)n_{\beta\beta} + W_s n_{\beta\alpha} + W_1 n_{\alpha\beta} + W_2 n_{\alpha\alpha} \end{cases} \quad (1.20)$$

In reality the populations in Equations (1.20) correspond to the deviation of the population of the states from equilibrium population, *e.g.* $\Delta n_{\alpha\alpha} = n_{\alpha\alpha} - n_{\alpha\alpha}^0$ leading to:

$$\begin{cases} \frac{d\Delta n_{\alpha\alpha}}{dt} = -(W_s + W_1 + W_2)\Delta n_{\alpha\alpha} + W_s \Delta n_{\alpha\beta} + W_1 \Delta n_{\beta\alpha} + W_2 \Delta n_{\beta\beta} \\ \frac{d\Delta n_{\alpha\beta}}{dt} = -(W_s + W_1 + W_0)\Delta n_{\alpha\beta} + W_s \Delta n_{\alpha\alpha} + W_1 \Delta n_{\beta\beta} + W_0 \Delta n_{\beta\alpha} \\ \frac{d\Delta n_{\beta\alpha}}{dt} = -(W_s + W_1 + W_0)\Delta n_{\beta\alpha} + W_s \Delta n_{\beta\beta} + W_1 \Delta n_{\alpha\alpha} + W_0 \Delta n_{\alpha\beta} \\ \frac{d\Delta n_{\beta\beta}}{dt} = -(W_s + W_1 + W_2)\Delta n_{\beta\beta} + W_s \Delta n_{\beta\alpha} + W_1 \Delta n_{\alpha\beta} + W_2 \Delta n_{\alpha\alpha} \end{cases} \quad (1.21)$$

The series of Equations can be expressed in a more compact fashion if we introduce:

- The I spin magnetization equal to the population difference across the two I spin transitions, $|\alpha\alpha\rangle \rightarrow |\beta\alpha\rangle$ and $|\alpha\beta\rangle \rightarrow |\beta\beta\rangle$:

$$I_z = n_{\alpha\alpha} - n_{\beta\alpha} + n_{\alpha\beta} - n_{\beta\beta} \quad (1.22)$$

- The S spin magnetization similar to I spin magnetization:

$$S_z = n_{\alpha\alpha} - n_{\alpha\beta} + n_{\beta\alpha} - n_{\beta\beta} \quad (1.23)$$

Cyril Charlier

As the same way we defined the population as the deviation from the equilibrium state we define $\Delta I_z(t) = I_z(t) - I_z^0$ where I_z^0 is the I spin magnetization at the equilibrium. The derivative form of Equations (1.22), (1.23) and (1.24) give:

$$\begin{cases} \frac{d\Delta I_z(t)}{dt} = \frac{d\Delta n_{\alpha\alpha}}{dt} - \frac{d\Delta n_{\beta\alpha}}{dt} + \frac{d\Delta n_{\alpha\beta}}{dt} - \frac{d\Delta n_{\beta\beta}}{dt} \\ \frac{d\Delta S_z(t)}{dt} = \frac{d\Delta n_{\alpha\alpha}}{dt} - \frac{d\Delta n_{\alpha\beta}}{dt} + \frac{d\Delta n_{\beta\alpha}}{dt} - \frac{d\Delta n_{\beta\beta}}{dt} \end{cases} \quad (1.24)$$

Combining Equations (1.20) and (1.21) give:

$$\begin{cases} \frac{d\Delta I_z(t)}{dt} = -(W_2 + W_0 + 2W_I)\Delta I_z - (W_2 - W_0)\Delta S_z \\ \frac{d\Delta S_z(t)}{dt} = -(W_2 + W_0 + 2W_S)\Delta S_z - (W_2 - W_0)\Delta I_z \end{cases} \quad (1.25)$$

Here, we can identify rates:

$$\begin{cases} \rho_I = (W_2 + W_0 + 2W_I) \\ \rho_S = (W_2 + W_0 + 2W_S) \\ \sigma_{IS} = W_2 - W_0 \end{cases} \quad (1.26)$$

and finally obtain the Solomon Equations for a two-spin system:

$$\begin{cases} \frac{d\Delta I_z(t)}{dt} = -\rho_I \Delta I_z(t) - \sigma_{IS} \Delta S_z(t) \\ \frac{d\Delta S_z(t)}{dt} = -\rho_S \Delta S_z(t) - \sigma_{IS} \Delta I_z(t) \end{cases} \quad (1.27)$$

2.2 Interpretation of Solomon Equations

Interestingly Equations (1.27) show that the variation of $I_z(t)$ depends also of the S spin magnetization. In other words, the return to equilibrium of the I spin also depends of the return to equilibrium of the S spin. This exchange of polarization is called *cross-relaxation*. σ_{IS} is called the *cross-relaxation rate constant*. ρ_I and ρ_S in Equation (1.27) corresponds to the *auto-relaxation rate constants* and can be written R_{1I} and R_{1S} in the Bloch formalism developed previously.

Solomon Equations can be extended to N interacting spins⁸⁴ and are also a very convenient way to describe the measurements of nuclear Overhauser effect (NOE)^{84,88}. However for practical applications, it is necessary to introduce the semi-classical approach developed Bloch, Wangsness and Redfield^{89,90}.

3. Bloch-Wangsness-Redfield (BWR) theory

The Bloch-Wangsness and Redfield approach^{89,91}, also referred as Redfield theory, consists in the treatment of the spin systems with quantum mechanics while the surrounding is described in a classical way. This semiclassical approach is valid in the high-limit

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

temperature and requires corrections at finite temperature. The BWT approach has been described in several books and reviews^{84,85,92,93}.

3.1 The Master Equation

The Hamiltonian, in the absence of *rf-field*, is written as a sum of a spin system Hamiltonian (\mathcal{H}_0) and a stochastic Hamiltonian $\mathcal{H}_1(t)$ that couples the spin system to the lattice:

$$\mathcal{H}(t) = \mathcal{H}_0 + \mathcal{H}_1(t) \quad (1.28)$$

Here \mathcal{H}_0 is time-independent and $\mathcal{H}_1(t)$ is a time-dependent perturbation to \mathcal{H}_0 . The fluctuations of the perturbation $\mathcal{H}_1(t)$, are due to motions of the molecule, and cause relaxation of the spin system towards equilibrium. The Liouville-von Neumann Equation, describing the evolution of the time-dependent density operator $\sigma(t)$, is given by:

$$\frac{d\sigma(t)}{dt} = -i[\mathcal{H}_0 + \mathcal{H}_1(t), \sigma(t)] \quad (1.29)$$

The effect of \mathcal{H}_0 on the density operator, $\sigma(t)$, can be effectively suppressed by a transformation into the *interaction frame*. An arbitrary operator B in the laboratory frame becomes in the interaction frame B^T :

$$B^T(t) = \exp(i\mathcal{H}_0 t) B(t) \exp(-i\mathcal{H}_0 t) \quad (1.30)$$

In this representation, the evolution of the density matrix is given by:

$$\frac{d\sigma^T(t)}{dt} = -i[\mathcal{H}_1^T(t), \sigma^T(t)] \quad (1.31)$$

This Equation is solved using successive assumptions. First of all, $\sigma^T(t)$ is written as a series $\sum_i \sigma_i^T(t)$ where

$$\begin{aligned} \sigma_0^T &= \sigma^T(0) \\ \sigma_{i+1}^T(t) &= -i \int_0^t [\mathcal{H}_1^T(t'), \sigma_i^T(t')] dt' \end{aligned} \quad (1.32)$$

We may integrate the Equation (1.31) up to the second order and introduce the results into The Liouville-von Neumann Equation leading to:

$$\frac{d\sigma^T(t)}{dt} = -i[\mathcal{H}_1^T(t), \sigma^T(0)] - \int_0^t [\mathcal{H}_1^T(t) [\mathcal{H}_1^T(t-\tau), \sigma^T(0)]] d\tau \quad (1.33)$$

Where $\tau = t - t'$. A list of additional assumptions is required to solve this Equation:

- The ensemble average of all the random Hamiltonians $\mathcal{H}_1(t)$ is zero, $\langle \mathcal{H}_1(t) \rangle = 0$. The first term of Equation (1.33) vanishes when we take the ensemble average of Equation 1.33.

Cyril Charlier

- $\mathcal{H}_1(t)$ and $\sigma^T(t)$ are uncorrelated and can be averaged separately.
- $\tau_c \ll t \ll 1/R$, where τ_c is the correlation time of the molecule relevant for fluctuations of $\mathcal{H}_1^T(t)$, yielding to the replacement of the upper limit of the integral by $+\infty$.
- The evolution of density operator $\sigma^T(t)$ caused by a random Hamiltonian $\mathcal{H}^T(t)$ is slow (*i.e.* $t \ll 1/R$ with R the relevant relaxation rate constant) so that $\sigma^T(0)$ on the right side of the Equation (1.33) can be replaced by $\sigma^T(t)$.
- It is permissible to replace $\sigma^T(t)$ by $\sigma^T(t) - \sigma_0$, in which σ_0 is the equilibrium density operator written as $\sigma_0 = (\exp[-\mathcal{H}_0/\hbar_B T]/\text{Tr}[\exp[-\mathcal{H}_0/\hbar_B T]])$ where \hbar_B is the Boltzmann constant, T the temperature and Tr means the trace operator.

Using these assumptions Equation (1.33) becomes:

$$\frac{d\sigma^T(t)}{dt} = - \int_0^\infty dt \overline{\mathcal{H}_1^T(t), [\mathcal{H}_1^T(t-\tau)]} - \sigma^T(t) - \sigma_0 \quad (1.34)$$

In order to transform (1.33) back into the laboratory frame, the transformation properties have to be established. First the stochastic Hamiltonian has to be decomposed as a sum of random functions of spatial variables $F_k^q(t)$ proportional to the spherical harmonic and a tensor spin operator \mathbf{A}_k^q where k is the rank of the tensor, typically one or two for NMR spectroscopy. In particular for $k = 2$ for dipole-dipole relaxation, relaxation due to chemical shift anisotropy and for quadrupolar relaxation:

$$\mathcal{H}_1(t) = \sum_{q=-k}^k (-1)^q F_k^{-q}(t) \mathbf{A}_k^q \quad (1.35)$$

The transformation of spin operators \mathbf{A}_k^q to the interaction frame results in:

$$\mathbf{A}_k^{qT} = \exp(i\mathcal{H}_0 t) \mathbf{A}_k^q \exp(-i\mathcal{H}_0 t) = \sum_p \mathbf{A}_{kp}^q \exp\{i\omega_p^q t\} \quad (1.36)$$

where \mathbf{A}_k^q satisfy the relationship $[\mathcal{H}_0, \mathbf{A}_{kp}^q] = \omega_p^q \mathbf{A}_{kp}^q$, and \mathbf{A}_{kp}^q and ω_p^q are the eigenfunctions and the eigenfrequencies of the Hamiltonian \mathcal{H}_0 . Substitution from $\mathcal{H}_1^T(t)$ from Equation (1.35) into (1.33) yields

$$\begin{aligned} \frac{d\sigma^T(t)}{dt} = & - \sum_{q=-k}^k \sum_{p,p'} \exp\{i(\omega_p^q - \omega_{p'}^q)t\} \left[\mathbf{A}_{kp'}^{-q}, [\mathbf{A}_{kp}^q, \sigma^T(t) - \sigma_0] \right] \\ & \times \int_0^\infty \overline{F_k^q(t) F_k^{-q}(t-\tau) \exp(-i\omega_p^q \tau)} dt \end{aligned} \quad (1.37)$$

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

where it is assumed that the random process $F_k^q(t)$ and $F_k^{q'}(t)$ are statistically independent unless $q = q'$. A second simplification is obtained by the secular approximation. The terms that oscillate much faster than typical relaxation times can be averaged out and do not affect the evolution of the density matrix. Furthermore, in the absence of degenerate eigenfrequencies, terms in Equations (1.37) are not vanishing only if $p = p'$. Thus,

$$\frac{d\sigma^T(t)}{dt} = - \sum_{q=-k}^k \sum_p \left[A_{kp}^{-q}, [A_{kp}^q, \sigma^T(t) - \sigma_0] \right] \times \int_0^\infty \overline{F_k^q(t) F_k^{-q}(t-\tau)} \exp(i\omega_p^q \tau) d\tau \quad (1.38)$$

The terms $\overline{F_k^q(t) F_k^{-q}(t-\tau)}$ are the *correlation functions*. The real part of the integral in Equation (1.38) is called *power spectral density function*, $j^q(\omega)$:

$$\begin{aligned} j^q(\omega) &= 2\text{Re} \left\{ \int_0^\infty \overline{F_k^q(t) F_k^{-q}(t-\tau)} \exp(-i\omega\tau) d\tau \right\} \\ j^q(\omega) &= \text{Re} \left\{ \int_{-\infty}^\infty \overline{F_k^q(t) F_k^{-q}(t-\tau)} \exp(-i\omega\tau) d\tau \right\} \\ j^q(\omega) &= \text{Re} \left\{ \int_{-\infty}^\infty \overline{F_k^q(t) F_k^{-q}(t+\tau)} \exp(-i\omega\tau) d\tau \right\} \end{aligned} \quad (1.39)$$

The power spectral density function is an even function of τ and ω .

3.2 Spectral density functions

For the particular case of isotropic liquids and in the high-temperature limit, it has been shown⁹⁴ that only one spectral density function has to be evaluated:

$$j^q(\omega) = (-1^q) j^0(\omega) \equiv (-1)^q j(\omega) \quad (1.40)$$

Regarding the relaxation mechanisms of interest, tensor operators of rank 2 are used. The random functions $F_2^0(t)$ can be factored as:

$$F_2^0(t) = c_0(t) Y_2^0[\Omega(t)] \quad (1.41)$$

Here, $c_0(t)$ is function of physical constants and spatial variables (Table 1.1), $Y_2^0[\Omega(t)]$ is a modified second-order spherical harmonic function (Table 1.2) and $\Omega(t) = \{\theta(t), \phi(t)\}$ are the polar angles in the laboratory frame. The polar angles are used to define the orientation of a unit vector that points in the principal direction for the interaction. Equation (1.41) can be substituted into (1.39):

$$\begin{aligned} j(\omega) &= \text{Re} \left\{ \int_{-\infty}^\infty c_0(t) c_0(t+\tau) Y_2^0[\Omega(t)] Y_2^0[\Omega(t+\tau)] \exp(-i\omega\tau) d\tau \right\} \\ j(\omega) &= \text{Re} \left\{ \int_{-\infty}^\infty C(\tau) \exp(-i\omega\tau) d\tau \right\} \end{aligned} \quad (1.42)$$

With the stochastic correlation function:

$$C(\tau) = \overline{c_0(t)c_0(t+\tau)Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]} \quad (1.43)$$

Table 1.1: Spatial functions for relaxation mechanisms. The chemical shift tensor is assumed to be axially symmetric with principal values $\sigma_{zz} = \sigma_{\parallel}$, $\sigma_{xx} = \sigma_{yy} = \sigma_{\perp}$ and $\Delta\sigma = \sigma_{\parallel} - \sigma_{\perp}$.

Interaction	c(t)
Dipolar	$-\sqrt{6}(\mu_0/4\pi)\hbar\gamma_I\gamma_S r_{IS}^{-3}(t)^{-3}$
CSA	$\Delta\sigma\gamma_I B_0/\sqrt{3}$

Table 1.2: Second-Order spherical harmonics

q	unmodified Y_2^q	modified Y_2^q
0	$\sqrt{15/16\pi}(3\cos^2\theta - 1)/2$	$(3\cos^2\theta - 1)/2$
1	$\sqrt{15/8\pi}\sin\theta\cos\theta e^{-i\phi}$	$\sqrt{3/2}\sin\theta\cos\theta e^{-i\phi}$
2	$\sqrt{15/32\pi}\sin^2\theta e^{-i2\phi}$	$\sqrt{3/8}\sin^2\theta e^{-i2\phi}$

For a rigid molecule, the spatial variable is time-independent, $c_0(t) = c_0$, and the spectral density function is $j(\omega) = d_{00}J(\omega)$ where $d_{00} = c_0^2$. The orientational spectral density function $J(\omega)$ is defined as:

$$J(\omega) = \text{Re} \left\{ \int_{-\infty}^{\infty} C_{00}^2(\tau) \exp(-i\omega\tau) d\tau \right\} \quad (1.44)$$

in which the orientational correlation function $C_{00}^2(\tau)$ is

$$C_{00}^2(\tau) = \overline{Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]} \quad (1.45)$$

For the particular case of *isotropic* rotational diffusion of a rigid rotor, the orientational correlation function is

$$C_{00}^2(\tau) = C_1(\tau) = \frac{1}{5} e^{-\frac{\tau}{\tau_c}} \quad (1.46)$$

in which, τ_c is the correlation time of the molecule. The correlation time depends of the size of the molecule, the viscosity of the solvent and the temperature. The correlation function is shown in Figure 1.3 with different correlation time values. As explained above, the spectral density function is

$$J(\omega) = \frac{2}{5} \frac{\tau_c}{(1 + \omega^2\tau_c^2)} \quad (1.47)$$

The spectral density function $J(\omega)$ in the Equation (1.47) is a Lorentzian function. Figure 1.3 shows the spectral density function $J(\omega)$ versus ω shows that a Lorentzian is relatively

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

constant for $\omega^2\tau_c^2 \ll 1$. If the molecular motions are fast enough $(\omega^q\tau_c)^2 \ll 1$, then $J(\omega_p^q) \approx J(0)$. The limit is called *extreme narrowing regime*. On the opposite, if $(\omega^q\tau_c)^2 \gg 1$, then $J(\omega_p^q) \propto \omega_p^{q-2}$, known as the *slow tumbling regime*.

Fluctuations of orientation-dependent interactions depend on motions of nuclei in a molecular reference frame as well as overall rotational Brownian motion. In the particular case of isotropic rotational diffusion, the total correlation should be factored as:

$$C(\tau) = C_0(\tau)C_1(t) \quad (1.48)$$

where $C_0(\tau)$ is the correlation function of overall motions, i.e. $C_{00}^2(\tau)$ above and $C_1(\tau)$ is the correlation function for internal motions. This factorization is valid as long as the overall and local motions are statistically independent.

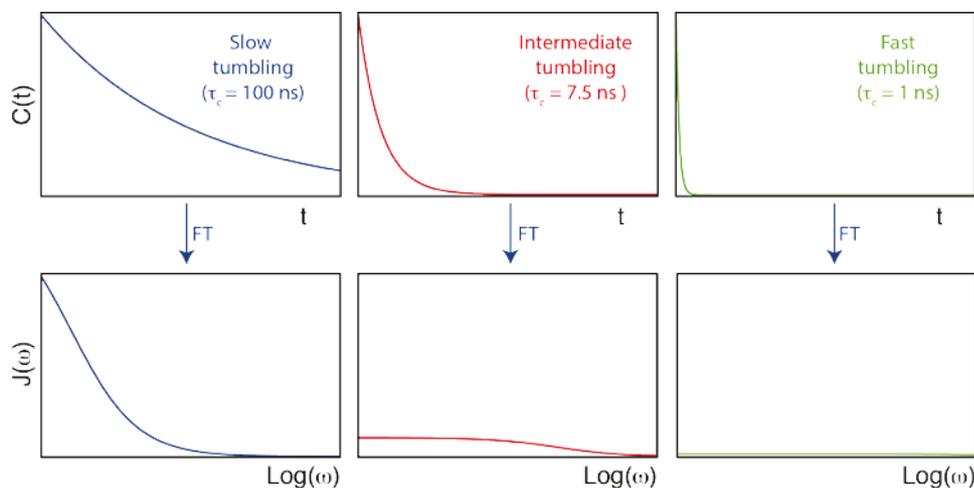


Figure 1.3: Correlation functions, $C(t)$ and the corresponding spectral density functions, $J(\omega)$. Note that a log scale was used for $J(\omega)$ plots. The simulation were carried out with $\tau_c = 100$ ns (slow tumbling), $\tau_c = 7.5$ ns (intermediate tumbling) and $\tau_c = 1$ ns (fast tumbling) using Equations (1.46) and (1.47)

3.3 Model-Free approach

The most popular approach to develop the correlation function of internal motions in Equation (1.48) is based on the statistical independence of the overall and internal motions even in the case of anisotropic molecular reorientation. This approach does not contain any assumptions on a physical model of motions but postulates the form of the correlation function. The correlation function of the model-free⁹⁵⁻⁹⁷ approach, based on a single-exponential correlation function is

$$C_1^{MF}(\tau) = S^2 + (1 - S^2)e^{-\frac{\tau}{\tau_1}} \quad (1.49)$$

Cyril Charlier

in which, S^2 is the generalized order parameter and τ_i is an internal correlation time. When S^2 is close to one, motions are very restricted. If S^2 is close to zero internal motions are unrestricted. In the case of isotropic overall rotational diffusion, the spectral density function is given by the Fourier transform:

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau_e}{1 + \omega^2 \tau_e^2} \right) \quad (1.50)$$

where τ_e is defined as

$$\frac{1}{\tau_e} = \frac{1}{\tau_c} + \frac{1}{\tau_i} \quad (1.51)$$

The first version of the model-free approach describes internal motions with a single correlation time. This model provides a good approximation of the spectral density function if (i) internal motions are much faster than the overall motions (ii) the internal motions are in the extreme narrowing limit. Typically for ^{15}N measurements from 9.4 to 18.8 T this holds if $\tau_e < 50\text{--}100$ ps and $\tau_c > 1$ ns. Few years latter, an extension was suggested⁹⁸ where the correlation functions contains one additional internal correlation time:

$$C_i^{EMF}(\tau) = S_f^2 S_s^2 + (1 - S_f^2) e^{-\frac{\tau}{\tau_f}} + S_f^2 (1 - S_s^2) e^{-\frac{\tau}{\tau_s}} \quad (1.52)$$

Here, the internal motions are subdivided in the fast internal motions (corresponding to the slow correlation time τ_f) and slow internal motions (corresponding to fast correlation time τ_s). S_f^2 and S_s^2 have the same meaning than S^2 in model-free for fast and slow motions, respectively. For convenience, we usually define $S^2 = S_f^2 S_s^2$. In the case of isotropic overall rotation, the spectral density function is:

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S_f^2) \tau_{f'}}{1 + \omega^2 \tau_{f'}^2} + \frac{(S_f^2 - S^2) \tau_{s'}}{1 + \omega^2 \tau_{s'}^2} \right) \quad (1.53)$$

where,

$$\frac{1}{\tau_{f'}} = \frac{1}{\tau_c} + \frac{1}{\tau_f} \quad \text{and} \quad \frac{1}{\tau_{s'}} = \frac{1}{\tau_c} + \frac{1}{\tau_s} \quad (1.54)$$

Simulated dependence of nitrogen-15 longitudinal relaxation rates on internal motions with nanosecond timescales are various magnetic fields are shown in Figure 1.4. The model-free approach has been the most successful framework used to interpret relaxation rates in terms of ps-ns motions for over twenty years. Nevertheless, many other models have been developed with particular forms of the correlation function such as *Gaussian Axial Fluctuation* (GAF)⁹⁹ or the *Slow Relaxing Local Structure* (SRLS)¹⁰⁰ for instance. A major difference between the model-free approach and the SRLS model¹⁰⁰, consists in the introduction of a coupling

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

between internal and overall motions. All the models discussed here, assumed a functional form of the correlation function. By contrast, relaxation rates can also be interpreted with an approach called *spectral density mapping*, which consists in the determination of the spectral density function at a few frequencies defined by the spin system and the magnetic field^{101,102}. This method will be detailed in the chapter 3.

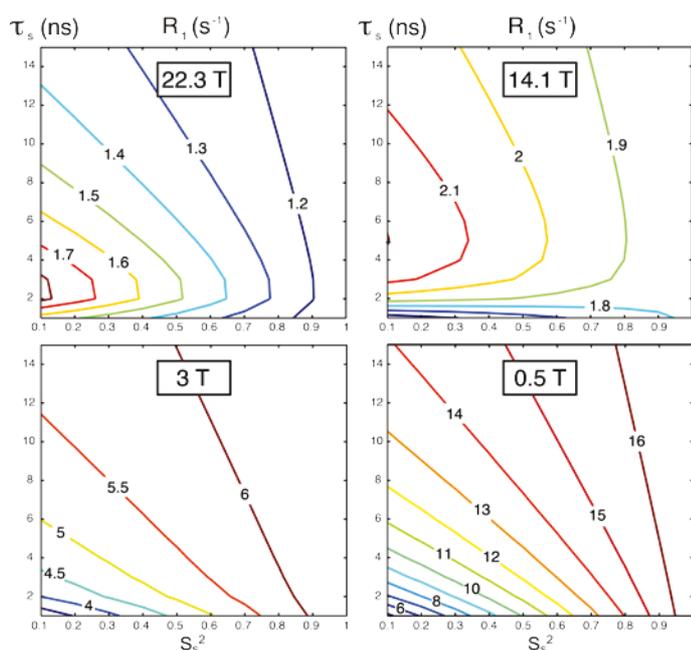


Figure 1.4: Simulated dependence of nitrogen-15 longitudinal relaxation rates for different B_0 on internal motions with nanosecond timescales, i.e., below and above the correlation time for overall tumbling $\tau_c = 5$ ns. An extended model-free spectral density function was used with the following parameters: $\tau_c = 5$ ns; correlation time for fast internal motions $\tau_{\text{fast}} = 10$ ps; order parameters for fast internal motions $S_{\text{fast}}^2 = 0.8$; order parameters for slow internal motions $0.1 < S_{\text{slow}}^2 < 1.0$ (x axis); $1 < \tau_{\text{int}} < 15$ ns (y axis)

3.4 Relaxation mechanisms

Relaxation Mechanisms are the interactions which fluctuations give rise to nuclear spin relaxation. A large number of physical interactions are capable of mediating spin relaxation but only a few are really important for spin-1/2 nuclei in biomolecules. For spin-1/2 nuclei in diamagnetic biological macromolecules, the dominant relaxation mechanisms are the magnetic dipolar and anisotropic chemical shift mechanisms. Here, we will limit ourselves to the description of intramolecular magnetic dipolar and anisotropic chemical shift.

3.4.1 Dipolar relaxation

Intramolecular dipolar relaxation is due to fluctuations of the magnetic dipolar field with internal motions and the tumbling of the molecule. The terms A_{2p}^q for the dipolar interaction are given Table 1.3 in and the relaxation rate constants are shown in Table 1.4.

Table 1.3: Tensors Operators for the dipolar Interaction

q	p	A_{2p}^q	$A_{2p}^{-q} = (-1)^q A_{2p}^{qi}$	ω_p^q
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	$-(1/2)I_z S^+$	$-(1/2)I_z S^-$	ω_S
1	1	$-(1/2)I^+ S_z$	$(1/2)I^- S_z$	ω_I
2	0	$(1/2)I^+ S^+$	$(1/2)I^- S^-$	$\omega_I + \omega_S$

Table 1.4: 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	$(d_{00}/4)\{J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)\}$
	$I_z \leftrightarrow S_z^b$	$(d_{00}/4)\{-J(\omega_I - \omega_S) + 6J(\omega_I + \omega_S)\}$
	$2I_z S_z$	$3(d_{00}/4)\{J(\omega_I) + J(\omega_S)\}$
	ZQ_x, ZQ_y	$(d_{00}/8)\{2J(\omega_I - \omega_S) + 3J(\omega_I) + 3J(\omega_S)\}$
0	I^+, I^-	$(d_{00}/8)\{4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_S) + 6J(\omega_I + \omega_S)\}$
	± 1	$(d_{00}/8)\{4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)\}$
	$2I^+ S_z, 2I^- S_z$	$(d_{00}/8)\{4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)\}$
± 1	$2I_z S^+, 2I_z S^-$	$(d_{00}/8)\{4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)\}$
	± 2	$(d_{00}/8)\{3J(\omega_I) + 3J(\omega_S) + 12J(\omega_I + \omega_S)\}$

3.4.2 Chemical shift anisotropy

The chemical shift interaction originates from the interaction of the electrons in the vicinity of the nucleus with the static magnetic field¹⁰³. Due to the geometry of molecular orbitals around the nuclei, the chemical shift is generally not isotropic. Thus, the chemical shift interaction changes with the orientation of the molecule with respect to the magnetic field. In isotropic solutions, fast overall tumbling of the molecule averages the chemical shift and only one isotropic chemical shift value is observed in NMR spectra. The chemical shift interaction is decomposed in two contributions: the isotropic chemical shift and the, traceless, chemical shift anisotropy (CSA).

In biomolecular NMR, the CSA of ^1H , ^{15}N , ^{13}C and ^{31}P have significant contributions to relaxation except for ^1H and ^{13}C in aromatic systems. CSA relaxation rate constants have a quadratic dependence on the applied magnetic field strength. Interestingly, high-magnetic are

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

not always favorable to signal-to-noise ratio due to broadening of the resonances linewidths coming from the increase of CSA contributions to relaxation. The terms for A_{2p}^q for the CSA are given in Table 1.5 and the relaxation rate constants are shown in Table 1.6. These results are calculated for axially symmetric chemical shift CSA tensors.

Table 1.5: Tensors Operators for the CSA Interaction

q	p	A_{2p}^q	$A_{2p}^{-q} = (-1)^q A_{2p}^{q\dagger}$	ω_p^q
0	0	$(2/\sqrt{6})I_z$	$(2/\sqrt{6})I_z$	0
1	0	$-(1/2)I^+$	$(1/2)I^-$	ω_I
2	0	–	–	$2\omega_I$

Table 1.6: CSA Relaxation Rate Constants

Coherence level	Operator	Relaxation rate constant
Populations	I_z	$d_{00}J(\omega_I)$
± 1	I^+, I^-	$(d_{00}/6)\{4J(0) + 3J(\omega_I)\}$

4. Chemical exchange

In NMR, the term *chemical exchange* describes the case of a nucleus in exchange between at least two sites with different chemical shifts due to chemical reactions or conformation fluctuations. Chemical exchange has major effects on the positions and the line width of the peaks detected in an NMR spectrum. In this section, we will describe these effects and introduce a technique to derive structural, kinetic and thermodynamic parameters of chemical exchange from the study of line broadening.

4.1 Theory

In the following discussion we consider the effect of chemical exchange in absence of scalar couplings interactions by an extension of the Bloch Equations. We treat here the case of two-site chemical exchange between two species A_1 and A_2 where k_1 and k_{-1} are the *first order forward* and *reverse rate constants*:



The concentrations of A_1 and A_2 are described by the Equations

$$\begin{cases} \frac{d[A_1]}{dt} = -k_1[A_1] + k_{-1}[A_2] \\ \frac{d[A_2]}{dt} = k_1[A_1] - k_{-1}[A_2] \end{cases} \quad (1.56)$$

also written in matrix form as:

Cyril Charlier

$$\frac{d}{dt} \begin{bmatrix} [A_1](t) \\ [A_2](t) \end{bmatrix} = \begin{bmatrix} -k_1 & k_{-1} \\ k_1 & -k_{-1} \end{bmatrix} \begin{bmatrix} [A_1](t) \\ [A_2](t) \end{bmatrix} \quad (1.57)$$

The particular Equations for a simple two-site exchange site can be generalized to a set of N coupled reactions between *i*th and *j*th species as



Then, the Equations (1.56) can be generalized to

$$\frac{d\mathbf{A}(t)}{dt} = \mathbf{K}\mathbf{A}(t) \quad (1.59)$$

where the elements of the rate matrix K are given by:

$$\begin{cases} K_{ij} &= k_{ji} \\ K_{ii} &= -\sum_{\substack{j=1 \\ j \neq i}}^N k_{ij} \end{cases} \quad (1.60)$$

For such system McConnell derived from Bloch Equations, the so-called McConnell Equations:

$$\begin{aligned} \frac{dM_{jx}(t)}{dt} &= \gamma(1 - \sigma_j)[\mathbf{M}_j(t) \times \mathbf{B}(t)]_x - R_{2j}M_{jx}(t) + \sum_{k=1}^N K_{jk} M_{kx}(t) \\ \frac{dM_{jy}(t)}{dt} &= \gamma(1 - \sigma_j)[\mathbf{M}_j(t) \times \mathbf{B}(t)]_y - R_{2j}M_{jy}(t) + \sum_{k=1}^N K_{jk} M_{ky}(t) \\ \frac{dM_{jz}(t)}{dt} &= \gamma(1 - \sigma_j)[\mathbf{M}_j(t) \times \mathbf{B}(t)]_z - R_{1j}[M_{jx}(t) - M_{j0}(t)] + \sum_{k=1}^N K_{jk} M_{kz}(t) \end{aligned} \quad (1.61)$$

with

$$M_{j0}(t) = \frac{M_0[A_j](t)}{\sum_{j=1}^N [A_j](t)} \quad (1.62)$$

In the McConnell Equations, the index *j* refers to the same spins in different chemical environments. If the system is in chemical equilibrium, then $[A_j](t) = [A_j]$

The above Equations can be generalized to higher-order chemical reactions by defining the pseudo-order order rate constants:

$$k_{ij} = \frac{\xi_{ij}(t)}{[A_i](t)} \quad (1.63)$$

Where $\xi_{ij}(t)$ is the rate constant for the conversion of the *i*-species into the *j*-th one.

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

4.2 Effect on NMR spectra

The McConnell Equations can be solved as the Bloch Equations⁸⁴ and lead to the identification of three regimes of exchange based on the comparison between the exchange rate $k = k_{-1} + k_1$ and the difference of precession frequency between the two sites $\Delta\Omega = \Omega_1 - \Omega_2$ as shown in Figure 1.5:

- In a *slow exchange* regime, $\Delta\Omega \gg k$, two resonances are observed at Ω_1 and Ω_2 .
- In an *intermediate exchange regime* or *coalescence*, peaks are broaden and when $\Delta\Omega \sim k$ lines coalesce.
- In a *fast exchange* regime, $\Delta\Omega \ll k$, a single resonance line is observed at the population average of the two chemical shifts.

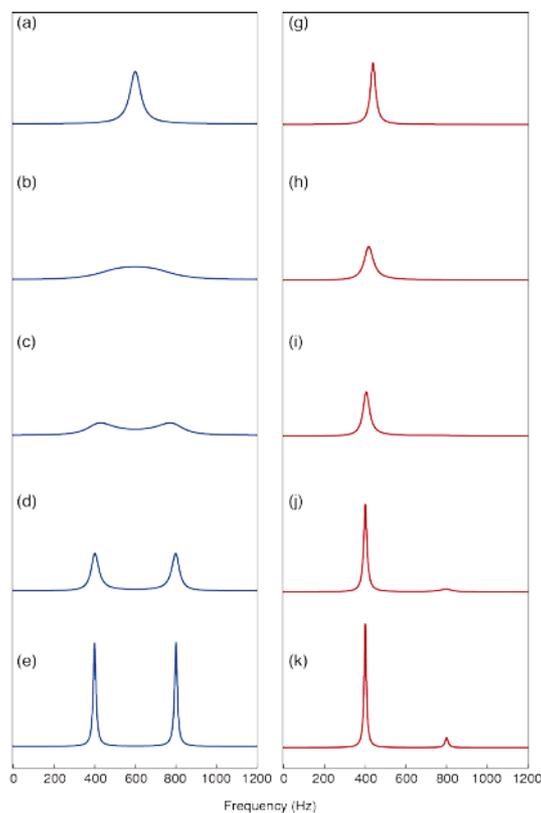


Figure 1.5: Chemical exchange for a two-site system [Modified from Cavanagh *et al.*⁸⁴]. Shown are the Fourier transformations of FIDs. In (a-f), $p_A = p_B = 0.5$ in (g-l), $p_A = 0.1$ and $p_B = 0.9$. Calculations were performed for values of the exchange rate k , equal to (a, g) 10000 s^{-1} , (b, h) 2000 s^{-1} , (c-i) 900 s^{-1} , (d, j) 200 s^{-1} , (e, k) 20 s^{-1} .

4.3 Quantifying chemical exchange with CPMG measurements

Over the past decades, the improvement in methodology allowed to quantify motions

in biomolecules on micro- to millisecond timescale dynamics with the measurement of chemical exchange process in large macromolecules, using for instance CPMG relaxation dispersion, $R_{1\rho}$ relaxation dispersion or ZZ exchange¹⁰⁴. Here, we simply present the concept of CPMG relaxation dispersion, which will be employed in the work presented in Chapter 4.

Introduced in 1954 by Carr and Purcell¹⁰⁵ and improved by Meiboom and Gill in 1958¹⁰⁶, the CPMG (Carr-Purcell-Meiboom-Gill) pulse sequence is a train of spin echoes. CPMG is the most commonly used technique to measure transverse relaxation rates. By varying the delay between refocusing pulses, so-called relaxation dispersion is one of the main techniques to quantify chemical exchange in macromolecules. Typically applied on ^1H , ^{15}N or ^{13}C nuclei, CPMG relaxation dispersion minimizes the effect of the exchange contribution (R_{ex}) to the transverse relaxation rates. In this technique, the relaxation of the transverse magnetisation is observed during a train of π pulses separated by delays τ_{cp} in a constant period $T_{\text{cpmg}} = (\tau_{\text{cp}} - \pi)_{\text{cpmg}}$ where n is an integer (Figure 1.6)^{107,108}.

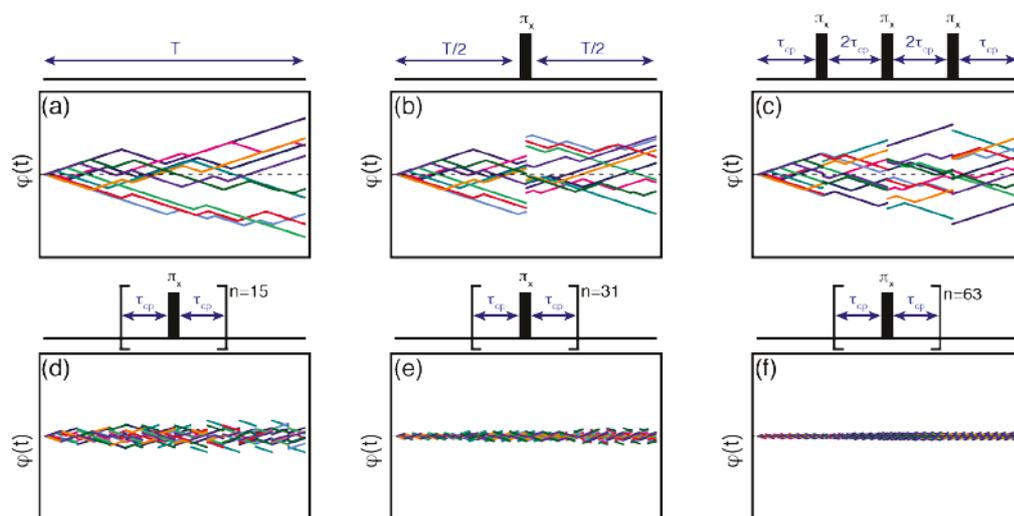


Figure 1.6: Evolution of the phase during a CPMG experiments with a variable number of π -pulses (a) 0; (b) 1; (c) 3; (d) 15; (e) 31 and (f) 63.

The principle of CPMG dispersion is illustrated in Figure 1.6. In the absence of rf-pulses (Figure 1.6-a), chemical exchange leads to a defocussing of transverse polarizations and an apparent increase of the relaxation rate R_2 due to the contribution of R_{ex} . When the pulsing frequency is smaller than the kinetic constants of exchange, this defocussing is still

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

present (Figure 1.6-b-c). When the pulsing frequency becomes larger than the exchange kinetics, chemical shift evolutions are refocused in between jumps between the two states, thus $R_{ex}(\tau_{cp})$ decreases (Figure 1.6-d-e). Practically, the effective relaxation rate (R_2^{eff}) is measured at various τ_{cp} leading to a dispersion curves. Information about kinetic processes are obtained by fitting dispersion curves using theoretical expressions of two- or three-site model. Experimental investigations of CPMG relation dispersion is widely used to study ligand binding, conformational changes, enzyme catalysis or protein folding¹⁰⁹⁻¹¹² for two or three site exchange in general^{82,113,114}.

4.4 Interpretation of CPMG experiments

The transverse relaxation rates R_2^{eff} are calculated from the extracted intensities of each 2D ^1H - ^{15}N spectra via:

$$R_2^{eff} = -\frac{1}{T_{cpmg}} \ln \left[\frac{I_{cpmg}}{I_0} \right] \quad (1.64)$$

where I_{cpmg} are the intensities at given frequency ν_{cpmg} and I_0 are the intensities without π -pulse (reference experiments).

The experimental dependence of the transverse relaxation rates on the frequency can be analysed with the use of a model of the exchange process. Carver and Richards derived the general Equation for a two states model A and B in 1972¹¹⁵. Particular Equations depending of the exchange regime treated have been developed¹¹⁶. Here we present the more general Equations working for all exchange regimes:

$$R_2 \left(\frac{1}{\tau_{cp}} \right) = \frac{1}{2} \left(R_{2A}^0 + R_{2B}^0 + k_{ex} - \frac{1}{\tau_{cp}} \cosh^{-1} [D_+ \cosh(\eta_+) - D_- \cos(\eta_-)] \right) \quad (1.65)$$

in which,

$$D_{\pm} = \frac{1}{2} \left[\pm 1 + \frac{\Psi + 2\Delta\omega^2}{(\Psi^2 + \xi^2)^{\frac{1}{2}}} \right] \quad (1.66)$$

$$\eta_{\pm} = \frac{\tau_{cp}}{\sqrt{2}} \left[\pm \Psi + (\Psi^2 + \xi^2)^{\frac{1}{2}} \right]^{\frac{1}{2}} \quad (1.67)$$

$$\Psi = (R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex})^2 - \Delta\omega^2 + 4p_A p_B k_{ex}^2 \quad (1.68)$$

$$\xi = 2\Delta\omega (R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex}) \quad (1.69)$$

Where R_{2A}^0 and R_{2B}^0 are the transverse relaxation rates in the free and the bound states, respectively; p_A and p_B are the populations in the free and the bound state; $\Delta\omega$ is the difference in chemical shift between the two states; and k_{ex} is the exchange rate conversion. In the case of slow-to-intermediate chemical exchange, all these parameters can be

determined from experimental CPMG dispersion data, providing information on the structure of exchanging species ($\Delta\omega$), the kinetics (k_{ex}) and the thermodynamics (p_A and p_B) of the exchange process.

5. *Experimental methods*

As mentioned above, nuclear spin relaxation is a powerful method to study protein dynamics. In the following section, we briefly introduce the pulse sequences used in this work for measuring ^{15}N relaxation rates. The experimental methods for measuring classical relaxation rates, (R_1 , R_2 and NOE) are well established for both ^{15}N and ^{13}C nuclei^{16,84,117,118}. The original pulse sequences were later modified to improve their sensitivity and accuracy^{117,119–122}.

5.1 Measurement of auto-relaxation rates

The common scheme for auto-relaxation measurements (R_1 , R_2) at high field is subdivided in 5 blocks: *preparation*, *relaxation*, *frequency labelling*, *mixing* and *detection*. The *preparation* period is composed by a refocused INEPT^{123,124} to transfer polarization from proton to nitrogen-15. During the *relaxation* period T the magnetization returns effectively towards a fully saturated state¹²⁵. The relaxation of the longitudinal nitrogen-15 polarization (N_z) is measured during the relaxation period of the R_1 experiment while the decay of the transverse nitrogen-15 coherence (N_x and N_y) is measured in the R_2 experiments. During the relaxation delay of a transverse relaxation measurement, the effect of the scalar coupling with the proton has to be suppressed in order to preserve an in-phase nitrogen-15 polarization. This is achieved either by a train of π -pulses (R_2^{cpmg}), which also suppresses the effect of supra-ms chemical exchange or a single echo (R_2^{echo}) with proton composite-pulse decoupling, which preserves the contribution of chemical exchange to transverse relaxation R_{ex} . R_1 and R_2 experiments are acquired as series of 2D spectra with various relaxation delays T . The relaxation rates (R_1 , R_2) are extracted by fitting the decrease of the intensities to a mono-exponential decay. During R_1 and R_2 experiments it is necessary to suppress the effects of cross-relaxation due to the cross correlation of CSA and DD interactions. This is achieved using ^1H pulses during the relaxation period. The t_1 *frequency-labelling* period, used to generate the second dimension of the spectrum, differs depending on the type of experiment. The *mixing* period is used for the coherence transfer from the nitrogen to the proton using reverse INEPT. Finally the proton magnetization is detected during the t_2 *acquisition* period.

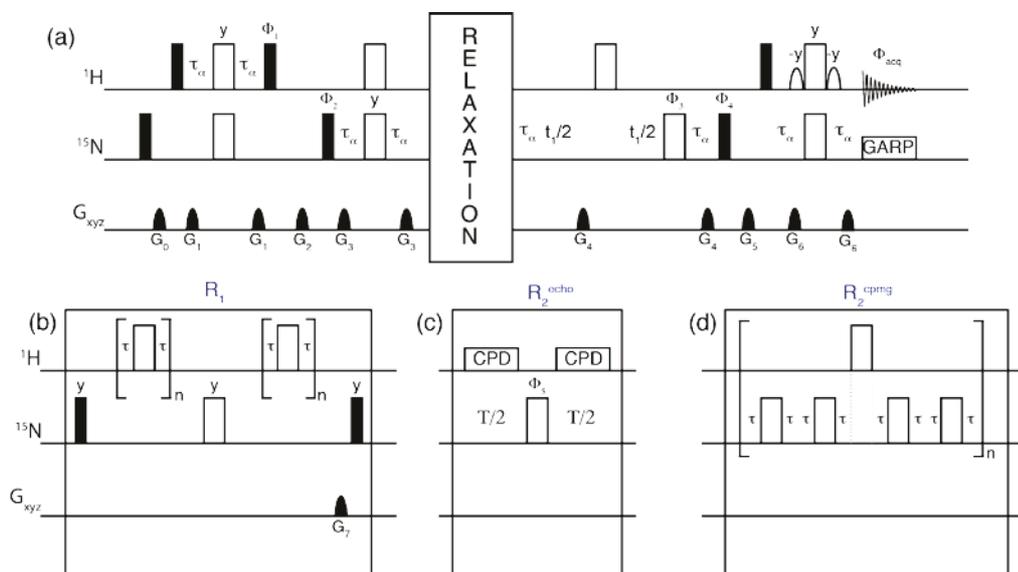


Figure 1.7: Pulse sequences used for measuring ^{15}N auto-relaxation rates. (a) General scheme for the generation and detection of in-phase nitrogen-15 polarization. (b) Relaxation sequence for measuring the longitudinal relaxation rate R_1 ; (c) relaxation sequence for measuring the transverse relaxation rate under a single echo R_2^{echo} ; (d) relaxation sequence for the measurement of the transverse relaxation rate under a CPMG echo train R_2^{cpmg} . All narrow (filled) and (wide) open rectangles represent 90° and 180° pulses. Pulses are along the x -axis of the rotating frame unless otherwise mentioned. (c) Proton composite-pulse decoupling during the relaxation delay was performed with a WALTZ-16 scheme ($T/2$ must be equal to the duration of an integer number of cycles of the decoupling sequence). Composite-pulse decoupling during acquisition was performed on the ^{15}N channel with a GARP scheme. The delay τ_{cr} is 2.65 ms. The phases cycles were: $\phi_1 \square \{y,y,-y,-y\}$; $\phi_2 \square \{x,-x\}$; $\phi_3 \square \{y,y,y,y,-y,-y,-y,-y\}$ ($\phi_3 = y$, not cycled for R_2^{echo}); $\phi_4 \square \{x,x,x,x,-x,-x,-x,-x\}$ ($\phi_4 = x$, not cycled for R_2^{echo}); $\phi_5 \square \{x,x,x,x,y,y,y,y\}$; $\phi_{acq} \square \{x,-x,-x,x,-x,x,-x,-x\}$; The amplitude profile of the pulsed field gradient had a sine-bell shape. The durations and peak amplitudes over the x , y and z orientations (when triple axis gradients are available) were respectively: G_0 : 1000 μs , [0, 0, 7.5] G/cm; G_1 : 550 μs , [0, 0, 3.5] G/cm; G_2 : 2500 μs , [13.5, 0, 24.5] G/cm; G_3 : 550 μs , [0, 0, 11.5] G/cm; G_4 : 550 μs , [0, 0, 9.5] G/cm; G_5 : 1000 μs , [0, 5.5, 18.5] G/cm; G_6 : 1000 μs , [33.5, 33.5, 33.5] G/cm; G_7 : 900 μs , [0, 0, 15.5] G/cm. Frequency sign discrimination in the indirect dimension was performed with the States-TPPI scheme¹²⁶.

5.2 Measurement of steady-state $^{15}\text{N}\{-^1\text{H}\}$ nuclear Overhauser effect

For the NOE experiments the experimental scheme is slightly different. The steady state $^{15}\text{N}\{-^1\text{H}\}$ NOE is measured by two interleaved spectra one with and one without proton saturation^{127,128}. Because this experiment starts from the nitrogen-15 polarization, its sensitivity is lower. The NOE values are obtained by the ratio of saturated over unsaturated intensities¹¹⁸.

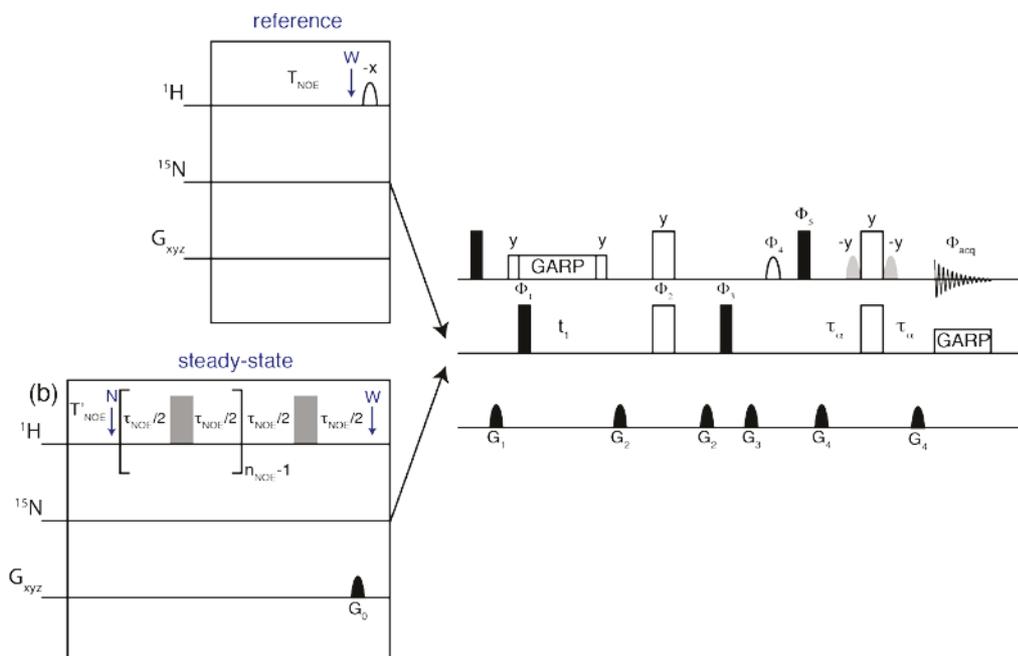


Figure 1.8: Pulse sequence for recording steady-state $^{15}\text{N}\{-^1\text{H}\}$ nuclear Overhauser effects. Reference and steady-state experiments were recorded in an interleaved manner. In the reference experiment, the “reference” box was used. At the end of the recovery delay $T_{\text{NOE}} \geq 10$ s, the proton carrier is placed on resonance with the water signal and a very selective waster flip-back pulse is applied (3 ms *sinc* shaped). To record the steady-state experiment, the boxed sequence “reference” is substituted by the scheme shown in the “steady-state” box for the effective saturation of amide proton resonances. After an initial delay $T'_{\text{NOE}} = 2$ s for stable detection of the lock signal, the proton carrier is placed in the center of the amide region (at 8.2 ppm) as shown by the arrow labeled by N. The motif (delay $\tau_{\text{NOE}}/2 - \pi$ pulse - delay $\tau_{\text{NOE}}/2$) is repeated n_{NOE} times for a total duration $T'_{\text{NOE}} \geq 4 \cdot T_1(\text{max})$, where $T_1(\text{max})$ is the longest nitrogen-15 relaxation time in the protein. We always have $T_{\text{NOE}} > T'_{\text{NOE}}$. The interpulse delay, τ_{NOE} is 22 ms. A gradient G_1 is applied at the end of the last $\tau_{\text{NOE}}/2$ delay to suppress all transverse components of the proton polarization. The carrier was moved on-resonance with the water signal as indicated by the W arrow. Proton composite-pulse decoupling during the t_1 delay was performed with a GARP scheme. Composite-pulse decoupling during acquisition was performed on the ^{15}N channel with a GARP. The delay τ_a is 2.56 ms. The phase cycles were: $\phi_1 \{y, -y\}$; $\phi_2 \{y, y, y, y, -y, -y, -y, -y\}$; $\phi_3 \{x, x, x, x, -x, -x, -x, -x\}$; $\phi_4 \{-x, -x, x, x\}$; $\phi_5 \{x, x, -x, -x\}$; $\phi_{\text{acq}} \{x, -x, -x, x, -x, x, x, -x\}$; The amplitude profiles of the pulse field gradient was a sine bell shape. The durations and peak amplitudes over x, y and z orientations were, respectively: G_0 : 1000 μs , [0, 0] G/cm; G_1 : 1000 μs , [0, 15.5] G/cm; G_2 : 900 μs , [0, 0, 9.5] G/cm; G_3 : 1500 μs , [5, 18.5] G/cm; G_4 : 1000 μs , [33.5, 33.5] G/cm. Frequency sign discrimination in the indirect dimension was performed with the States-TPPI scheme.

5.3 Measurement of CSA/DD cross-correlated cross-relaxation rates

Over the last two decades, the measurement of the transverse (η_{\perp}) and longitudinal (η_z) cross-correlated relaxation rates due to correlated fluctuations of $^1\text{H}\text{-}^{15}\text{N}$ DD and ^{15}N CSA^{125,129,130} have been demonstrated to be widely informative to describe the CSA¹³¹ and to

Chapter 1: Nuclear Spin Relaxation: Concepts & Methods

identify chemical exchange in proteins¹¹¹. Indeed, these relaxation rates are not contaminated by chemical exchange, so that an exchange-free R_2 can be estimated using η_{xy} , η_z , R_1 and σ_{NH}^{131} :

$$R_2^0 = (R_1 - 1.25\sigma_{NH}) \frac{\eta_{xy}}{\eta_z} - 1.08\sigma_{NH} \quad (1.70)$$

In addition, transverse cross-correlated relaxation rate (η_{xy}) is very sensitive for the characterization of order and disorder in proteins. η_{xy} cross-correlated relaxation rates (η_{xy}) are measured using symmetrical reconversion^{125,130}. These experiments are subsided in 6 blocks: *coherence transfer*, coherence selection, *relaxation*, second coherence selection, second *coherence transfer* and *acquisition*. During the first coherence selection either N_y (N_z) or $2N_yH_z$ ($2N_zH_z$) is selected for transverse (longitudinal) cross-correlated relaxation rates. The CSA/DD cross-correlated relaxation due to the fluctuations of CSA of ^{15}N and ^1H - ^{15}N DD interaction leads to the interconversion of these operators: $N_y \leftrightarrow 2N_yH_z$ ($N_z \leftrightarrow 2N_zH_z$). In this type of system, with two operators, four relaxation pathways are possible:

$$\begin{array}{lcl} P \leftrightarrow P & I & \\ P \leftrightarrow Q & II & \\ Q \leftrightarrow P & III & \\ Q \leftrightarrow Q & IV & \end{array} \Leftrightarrow \quad (1.71)$$

II (I) and III (IV) are called cross-relaxation pathways, I (II) and IV (III) are the auto-relaxation pathways. Each relaxation pathways is measured and the rates are obtained using the Equation 1.72:

$$\begin{aligned} \tanh|\eta_{xy}T| &= \sqrt{\frac{A_{II}(T)A_{III}(T)}{A_I(T)A_{IV}(T)}} \quad (A) \\ \tanh|\eta_zT| &= \sqrt{\frac{A_I(T)A_{IV}(T)}{A_{II}(T)A_{III}(T)}} \quad (B) \end{aligned} \quad (1.72)$$

Due to the swapping block in the middle of the relaxation period in the sequence for the measurement of the longitudinal rate, the right part of the Equation (1.72-B) is the reciprocal of Equation (1.72-A).

5.4 Measurement of ^{15}N - ^1H CPMG relaxation dispersion

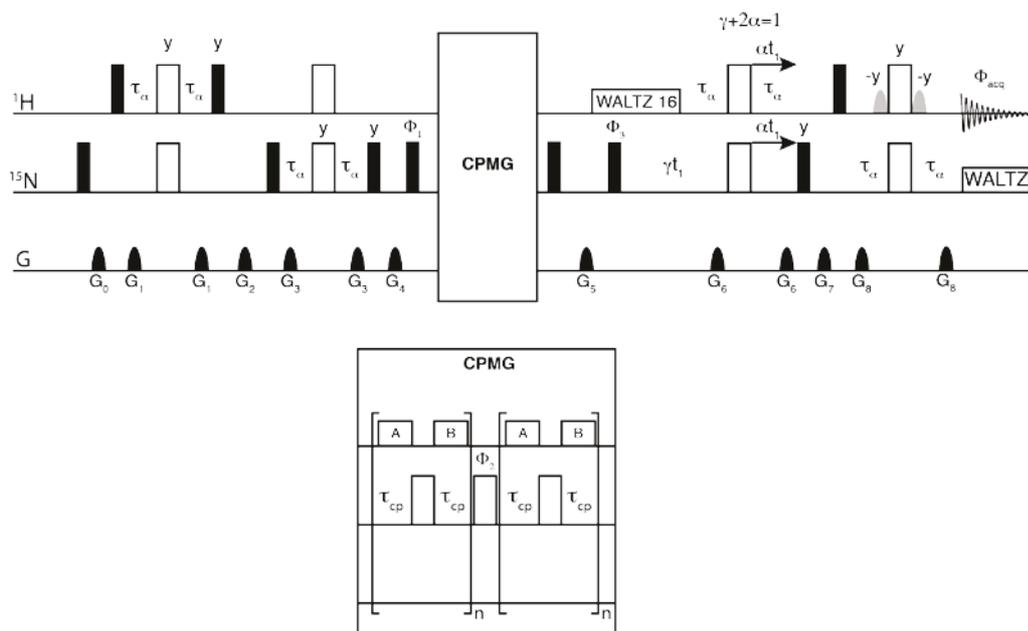
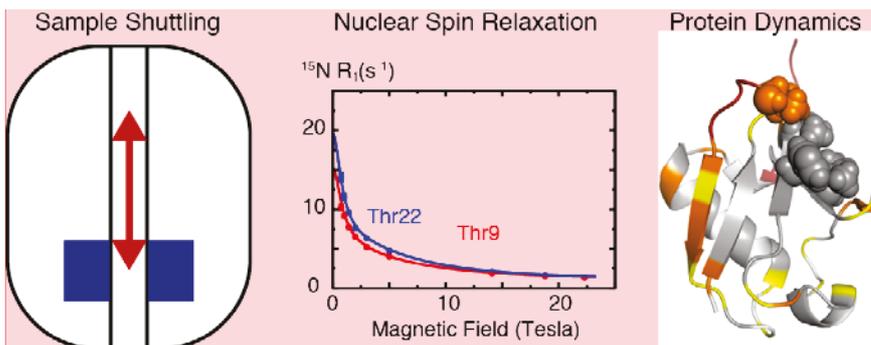


Figure 1.10: Pulse sequence for measuring CPMG dispersion profiles. All narrow (filled) and (wide) open rectangles represent 90° and 180° . Pulses are along the x-axis of the rotating frame unless otherwise mentioned. Proton composite pulse decoupling during the delay t_1 was performed with a WALTZ-16 scheme. Composite-pulse decoupling during acquisition was performed on the ^{15}N channel with a WALTZ scheme. The delay τ_a is 2.65 ms. The frequency labeling in the indirect dimension was done using semi-constant time manner. The phase cycles were: $\phi_1 = \{x, -x\}$; $\phi_2 = \{x, x, -x, -x\}$; $\phi_3 = \{x, x, -x, -x\}$; $\phi_{\text{acq}} = \{x, -x, -x, x\}$; Gradient durations and peak amplitudes over the x, y and z orientations were respectively: G_0 : 1000 [0, 0, 0] G/cm; G_1 : 3000 [0, 0, 0] G/cm; G_2 : 1000 [0, 0, 0] G/cm; G_3 : 5000 [0, 0, 0] G/cm; G_4 : 1000 [0, 0, 0] G/cm; G_5 : 1000 [0, 0, 0] G/cm; G_6 : 1000 [0, 0, 0] G/cm; G_7 : 1000 [0, 0, 0] G/cm; G_8 : 1000 [0, 0, 0] G/cm; Frequency sign discrimination was performed using States-TPPI. Decoupling block A and B were optimized for each experiment. Depending of the duration between π -pulses on the nitrogen, either WALTZ-8 or WALTZ-16 were used. If the duration of the delay between the π -pulses did not suffices to accommodate two entire cycles of the decoupling scheme the cycle was splitted in two parts, A and B. Otherwise A and B were equal to each other.

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry



Commentaire [CC1]: Position en hauteur

1. Introduction

Nuclear spin relaxation has been extensively used to study protein dynamics. Over the last decades the improvement in protein labeling, NMR methodology as well as the development of high-field magnet have allowed to obtain atomic resolution information about motions in biomolecules. The main advantages for NMR experiments performed using very high magnetic field are the sensitivity and the resolution of spectra. Nowadays, conventional experiments to study structure or dynamics with NMR are performed on high-field magnets typically ranging from 9.4 T (400 MHz) to 23.5 T (1000 MHz). Theoretically, relaxation rates are described using the spectral density function, $J(\omega)$, which is the Fourier transform of the correlation function, $C(t)$. By measuring relaxation rates between 9.4 T and 23.5 T, the spectral density function is explored on few narrow ranges of frequencies (Figure 2.1).

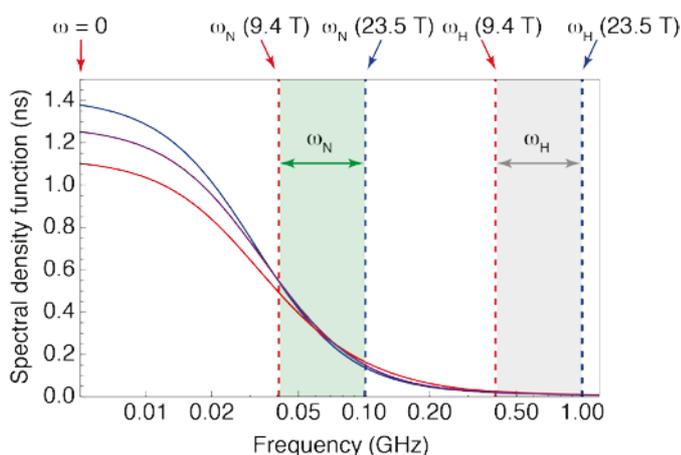


Figure 2.1: Spectral density function shown for three models: model-free with $\tau_c = 5$ ns, $S^2 = 0.7$, $\tau_e = 0.1$ ns (blue); extended model-free with $\tau_c = 5$ ns, $S^2 = 0.5$, $S_f^2 = 0.7$, $\tau_f = 10$ ns, $\tau_s = 0.1$ ns (purple) a $\tau_f = 2$ ns (red). The green and gray areas correspond to the range of nitrogen and proton Larmor frequencies accessible with conventional high-field relaxation measurements.

In order to complete information about internal motions in proteins accessible from conventional NMR, the mapping of the spectral density function at lower frequencies appears to be essential. The classical technique to probe longitudinal relaxation rates at lower frequencies is called relaxometry. This method, introduced in the nineteen sixties has been in continuous development since then, including the development of fast field cycling relaxometry¹³². Fruitful developments have been performed in order to measure field-

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

dependent NMR relaxation providing information on molecular motions, for instance water dynamics^{133–135}.

Although these methods can be used to obtain general information about oligomerization and motional modes in proteins¹³⁶, based on ^1H detection at low field are not suitable for site-specific studies of proteins. Indeed, the study of proteins with NMR requires high sensitivity and highly resolved spectra. To overcome this limitation, A.G. Redfield¹³² proposed an approach, which combines the advantages of the resolution and sensitivity obtained at high field and relaxation at low field, which we call *high-resolution relaxometry*. The principle is to prepare a spin term at high field, then shuttle the sample at a lower magnetic field for the relaxation period and go back to high field for detection (Figure 2.2-b). The shuttling time has to be as short as possible (lower than the relaxation time of the spin term prepared at high field) in order to avoid loss of magnetization during the transfer period. This approach has been conducted in two different ways. In the first case the lower magnetic field is created by a second magnet^{137,138} and in the second case the relaxation happens in the stray field of the high-field magnet^{132,139–141} (Figure 2.2-a). In these studies, relaxometry has been applied to lipids, DNA and proteins^{142,143} using ^{31}P and ^{13}C . To date only two studies about proteins came out in the literature on a limited range of frequencies^{144,145}.

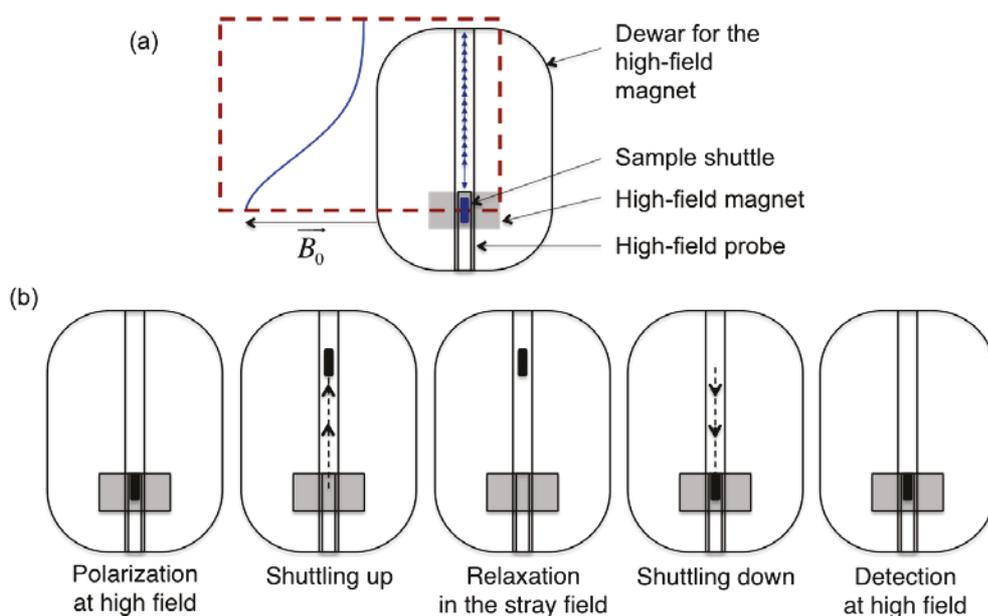


Figure 2.2: Principle of high-resolution relaxometry using the stray field of a high-field magnet. (a) Description of the apparatus with the map of the stray field; (b) Description of a typical experiment for the measurement of longitudinal relaxation rates at low field.

Cyril Charlier

Here, we report the design and use of a new shuttle system, installed on a 600 MHz spectrometer, where the polarization builds up and the signal is detected at high field while longitudinal relaxation takes place at various points of the stray field. We have measured longitudinal nitrogen-15 relaxation rates in ubiquitin over nearly two orders of magnitude of magnetic field from 0.5 T to 22.3 T. In order to analyze low field relaxation rates and correct for effects of relaxation during shuttling, we have developed an iterative protocol called Iterative Correction and Analysis of Relaxation Under Shuttling (ICARUS). In the following sections, we will describe the development of the pneumatic system, the ICARUS protocol as well as its application to ubiquitin. The article published in the *Journal of American Chemical Society* inspires this part⁸³.

2. Ubiquitin

Human ubiquitin has been widely used as a model system for NMR methodology. Indeed, the good solubility of ubiquitin and long stability of ubiquitin samples are great advantages in order to develop methods. Human Ubiquitin is generally accepted as reference molecule for methodological developments.

Ubiquitin, which is only found in eukaryotes, is highly conserved with only 3 mutations from yeast to humans, suggesting that all residues are crucial for its function. Ubiquitin exhibits various critical functions such as cell cycle regulation, DNA repair, embryogenesis or apoptosis (programmed cell-death with poly-ubitination)^{146,147}.

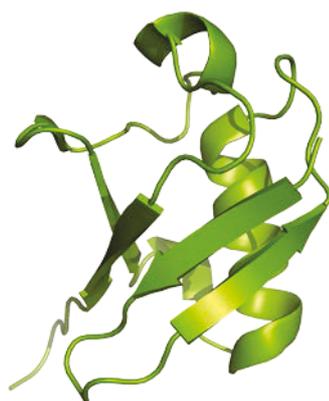


Figure 2.3: Structure of human ubiquitin (pdb code: 1D3Z)¹⁴⁸

Ubiquitin is a 76 residues protein, which crystal structure has been first determined in 1985^{149,150}. NMR ¹H, ¹⁵N and ¹³C assignments have been reported^{151–153}. The structure of

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

ubiquitin comprises four β -sheets, one α -helix and several loop regions (Figure 2.3). The structure of human ubiquitin is well understood (Figure 2.3), and its dynamics have been studied extensively by NMR^{111,154–162}. Motions in Ubiquitin are essential for its function. For instance, the β 1- β 2 turn has been shown recently to be important for the binding of specific partners¹⁶³.

3. Description of the pneumatic shuttle system

Our collaborators at *Bruker Biospin* have developed a pneumatic system for fast shuttling, based on a system that was originally developed for liquid-state dynamic nuclear polarization (DNP) studies where proton polarization observed at 14.1 T can be enhanced by saturating EPR transitions at 0.34 T¹⁶⁴. Our shuttle consists of a custom-designed probe, a transfer system and a control unit, as described in following sections.

3.1 Probe

The probe uses two saddle coils, like in standard high-resolution probes. The inner coil is doubly tuned for ^{13}C and ^{15}N , while the orthogonal outer coil is doubly tuned for ^1H (observation) and ^2H (field-frequency lock) (Figure 2.4) in a manner similar to so-called “observe” probes. This design reduces interactions between the sample and the *electric* component of the *rf-field*, albeit at the expense of a slight loss of sensitivity. Special care was taken to attenuate vibrations arising from shocks when the shuttle is stopped suddenly at the upper and lower ends of its displacement. Careful investigations of these effects were necessary to obtain line shapes of similar quality as in standard high-resolution NMR spectroscopy. A shielded z-gradient coil (Figure 2.4) allows one to use standard NMR experiments.

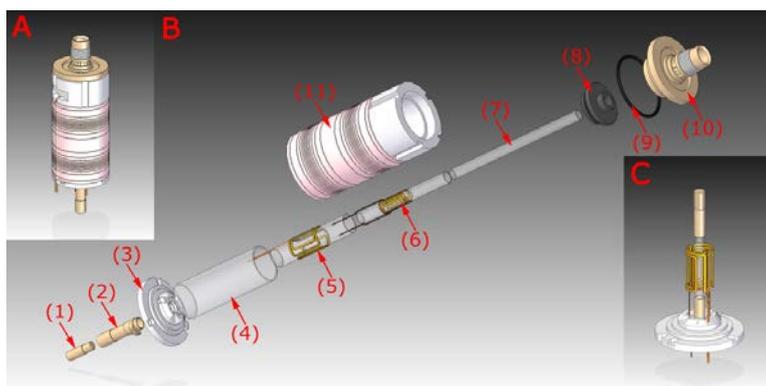


Figure 2.4: Schematic views of the upper part of the shuttle probe. A. Complete view. B. Exploded view. C. Coil assembly with amorphous quartz shuttle container. (1) Shuttle touchdown pad. (2) Lower attenuating connector. (3) Lower insert. (4) Thermal glass shield. (5) Guiding glass tube and

Cyril Charlier

outer *rf*-coil. (6) Guiding glass tube and inner *rf*-coil. (7) Shuttle protection glass tube. (8) Glass gauge. (9) Vibration damper. (10) Upper insert and second attenuating connector. (11) Z-gradient coil.

A vertical metal tube runs from the bottom of the probe to a position just below the *rf*-coils. This tube is used to insert or eject the shuttle container and to guide the shuttle stopper and the shuttle touchdown pad (Figure 2.4) The shuttle stopper is fixed with two springs at the bottom of the coaxial tube. The shuttle stopper and spring serve as shock absorbers that center the shuttle container with respect to the *rf*-coils (Figure 2.4-C). The “lower attenuating connector” marks the transition between the central tube and the detection area at the top of the probe (Figure 2.4-B-2). A sensor was integrated in the lower attenuating connector to detect the mechanical position of the shuttle container. A shuttle protection glass tube ensures that the shuttle container is properly aligned (Figure 2.4-B-7). The interconnection between the top of the probe and the shuttle transfer system consists of an upper insert and a second attenuating connector (Figure 2.4-B-10). An O-ring (Figure 2.4-B-9) inserted between the gradient system (Figure 2.4-B-11) and the upper insert further improves vibration damping.

The design of the *rf*-circuit is similar to many high-resolution probes. Low-susceptibility and/or susceptibility-compensated materials were selected, particularly for the coil and capacitor wires and all surrounding components. As a result, the spectral resolution and line shape were comparable to those of a standard 600 MHz high-resolution probe. However, as expected, the NMR sensitivity of the shuttle probe is reduced by about an order of magnitude due to the lower filling factor and reduced volume of the sample.

3.2 Shuttle transfer system

The shuttle transfer system allows one to stop the shuttle at a predetermined position at a chosen height in the stray field above the magnetic center (Figure 2.5). The shuttle guide consists of a tube connected to the top of the probe at its lower end, and to the top of the cryostat of the main magnet at its upper end. At the top, a second tube, coaxial with the shuttle guide, is equipped with a ‘stopper’ that prevents the shuttle container to move beyond a well-defined position.

This inner tube has been isolated from the outer one and a damping system has been designed to reduce vibrations and shock-waves from the shuttle motion and sudden stop. Inside the stopper a second optical sensor has been integrated to detect the precise position of the shuttle in the low field. The position of the stopper was measured before and after the experiments to ensure that the value of the magnetic field B_0^{low} was constant during the course

of the experiment.

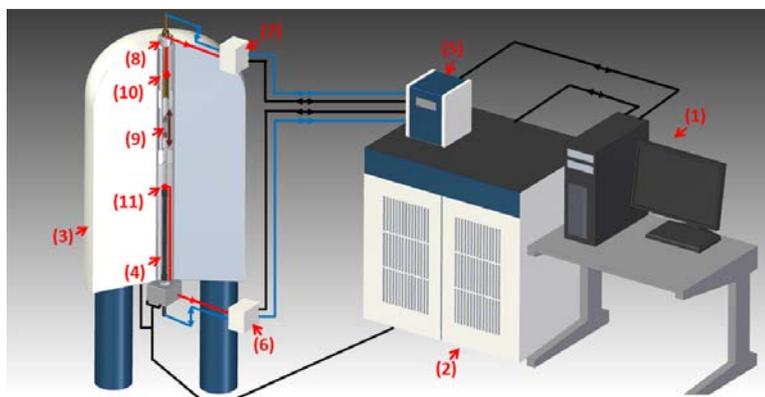


Figure 2.5: Schematic view of the shuttle system: (1) Workstation; (2) Console; (3) Magnet; (4) Shuttle probe; (5) Main shuttle control; (6) & (7) Shuttle control satellites; (8) Transfer system; (9) Shuttle container; (10) & (11) Optical sensors and touchdown pads.

3.3 Shuttle Controller Unit

The pneumatic shuttle control consists of a main unit and two satellite units (Figure 2.5). The main unit is equipped with a microcontroller board which allows communication with the NMR workstation and the spectrometer console, controls the valves and pressures, and allows one to determine the position of the sample. In addition, a pneumatically driven vacuum pump generates a suitable under pressure. The two satellites are equipped with a proportional pressure valve that controls the shuttle motion and an optical sensor to detect the end positions of the shuttle. The main unit is designed to be installed either next to the NMR workstation or to the spectrometer console, while the two satellite units should be close to the magnet, one at the bottom, close to the shuttle probe, and the other one at the top, close to the shuttle transfer system.

A simple script file is used to program the shuttle controller. This file lists pressure settings for the transfers and at the two static positions. The pressure settings are kept constant during each phase. The shuttle motion is activated by the TCU (timing control unit) of the spectrometer and information from the optical sensors about the positions (top for low field, bottom for high field) is sent back to the spectrometer. The shuttle controller creates a report in the form of a table with the timing of all shuttle motions that can be displayed by the NMR workstation.

Cyril Charlier

3.4 Shuttle container

A special quartz container was chosen for protein samples (Figure 2.6). Synthetic amorphous quartz glass with a low magnetic susceptibility can resist a large number of shocks. The shuttle system was developed to withstand fast motions and strong shocks caused by sudden stops. (A single container was used for more than 500 000 shuttling events in the course of this study). This is particularly important for the sample container (Figure 2.6). A special synthetic amorphous quartz was chosen for the glass parts (Figure 2.6-4-7) and a high performance polyimide resin for the caps (Figure 2.4-1-2). Two O-rings (Figure 2.4-3) were integrated in the caps as dampers to reduce shocks to the glass body. All of these materials have a low magnetic susceptibility to reduce distortions of NMR signals.

The total sample volume (V_2) is about 110 μL , while the active volume (V_1) of the sample is about 60 μL (Figure 2.6). Once filled, the container is sealed with glue. The shuttle container has a “bubble catcher” with a volume of about 10 μL to confine air bubbles that may appear in the sample and to accommodate thermal expansion of the sample. Bubbles appearing in the active volume of the sample can be easily centrifuged into the bubble catcher through a thin capillary. In our hands the samples are stable for several hours, and bubbles are predominantly confined in the bubble catcher.

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

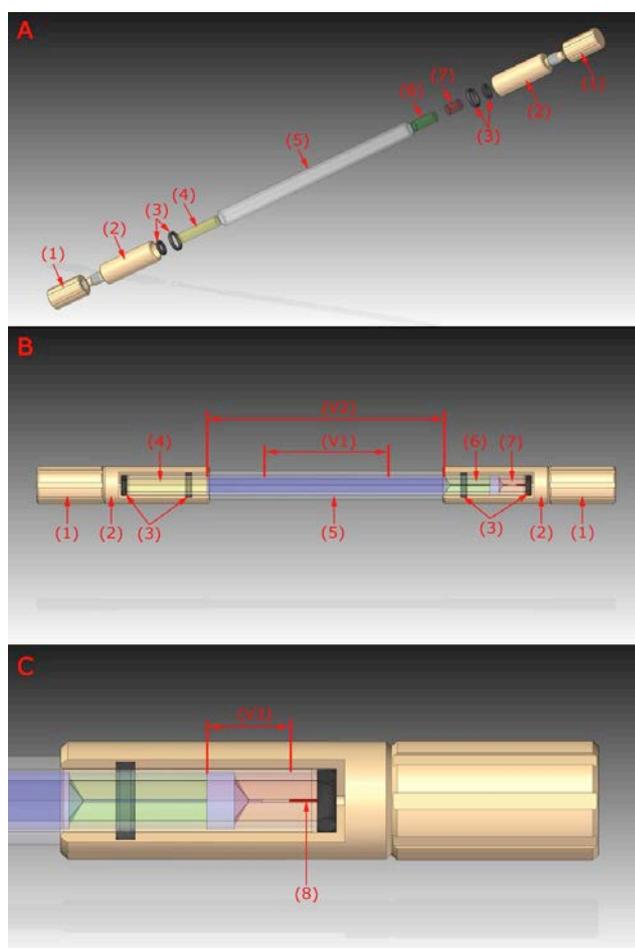


Figure 2.6: Schematic view of the shuttle container. A. Exploded view of all parts; B. Assembled shuttle container with glass parts and sample in blue; (1) End plug; (2) Glass connector; (3) Shock and vibration damper; (4) Bottom glass plug; (5) Shuttle glass tube; (6) Inner glass capillary; (7) Top glass capillary; (V1) active sample volume. (V2) Total sample volume. C. Details of the upper part of the shuttle container: (V3) Sample reservoir and bubble catcher; (8) Glue seal.

3.5 Mapping magnetic field

In order to know the relationship between the position in the magnet and the explored magnetic field we measured as a function of the height above the magnetic center in steps of 1 mm using a mapping device built by our collaborator Dimitris Sakellariou (CEA) with two calibrated triple-axes Hall probes (Senis) with a precision of 0.1%. A CH3A10mE3D transducer was used for measurements from 0.05 to 2 T, while a 03A05F-A20T0K5Q transducer was used between 1 and 13 T.

4. ^{15}N Relaxation measurements

The experiments were carried out on samples of 0.2 and 3 mM uniformly ^{15}N labeled human ubiquitin (Giotto) in 50 mM acetate buffer (pH 4.5) in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (90/10 v/v).

4.1 High-field relaxation measurements

A set of conventional nitrogen-15 relaxation experiments was recorded without shuttling at 14.1 T, 18.8 T, and 22.3 T (600, 800 and 950 MHz for ^1H , respectively) using state-of-the-art methods to cancel cross-correlation effects^{120,127,165} (Figure 2.7). The reader is referred to the first chapter for extensive details about pulse programs.

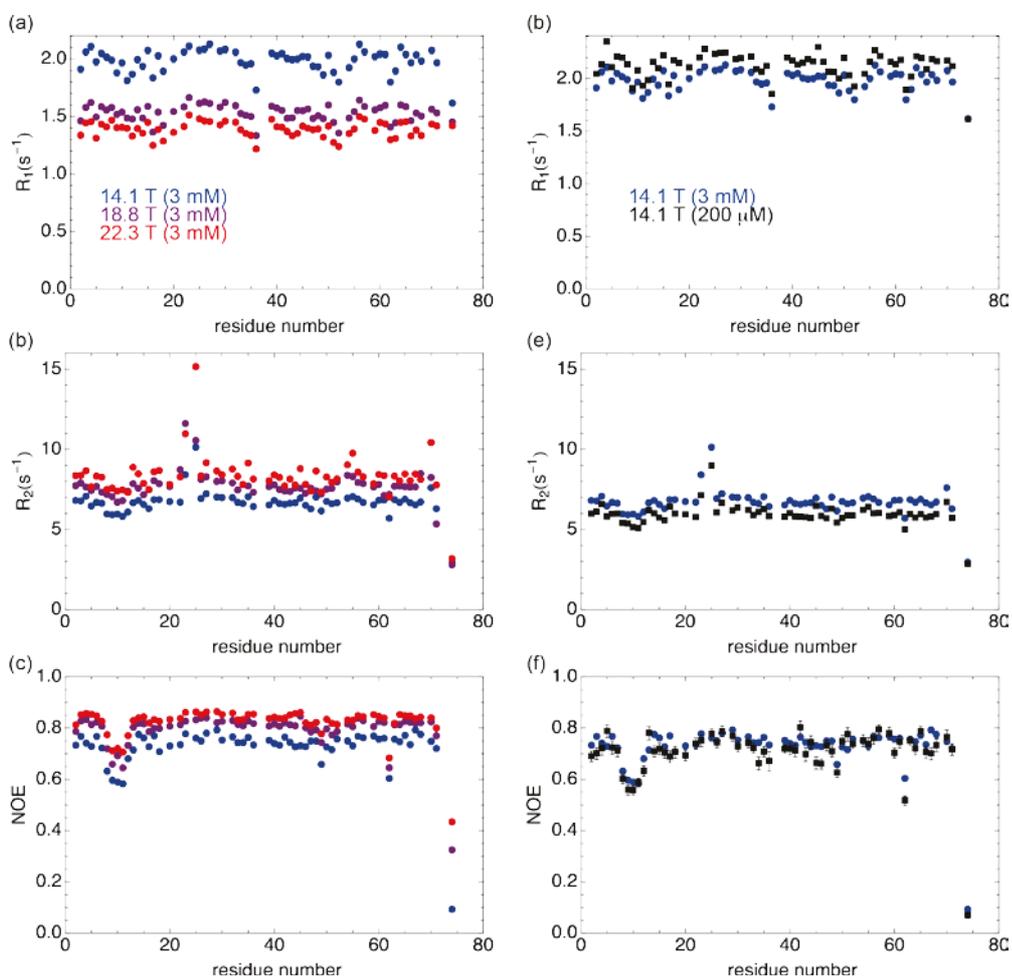


Figure 2.7: Backbone relaxation rates measurements in 3 mM ubiquitin measured at 600 MHz (blue), 800 MHz (purple), 950 MHz (red) and in 200 μM ubiquitin measured at 600 MHz (black). (a) Longitudinal relaxation rates R_1 of ^{15}N at 3 mM; (b) Transverse relaxation rates R_2 of ^{15}N at 3 mM; (c) ^{15}N - $\{^1\text{H}\}$ NOE ratios at 3 mM; (d) Longitudinal relaxation rates R_1 of ^{15}N at 3 mM (blue circles)

Cyril Charlier

5 T (at $z = 27$ cm above the magnetic center) (Figure 2.9-a) and $S/N = 24$ for $T_{\text{rel}} = 51$ ms at $B_0^{\text{low}} = 0.5$ T ($z = 46$ cm) (Figure 2.9-b) when 16 transients were recorded for each of 64 complex points in the indirect t_1 dimension (the experimental time was about 85 minutes for each interval T_{rel}). All signals were processed and analyzed with NMRPipe¹⁶⁶.

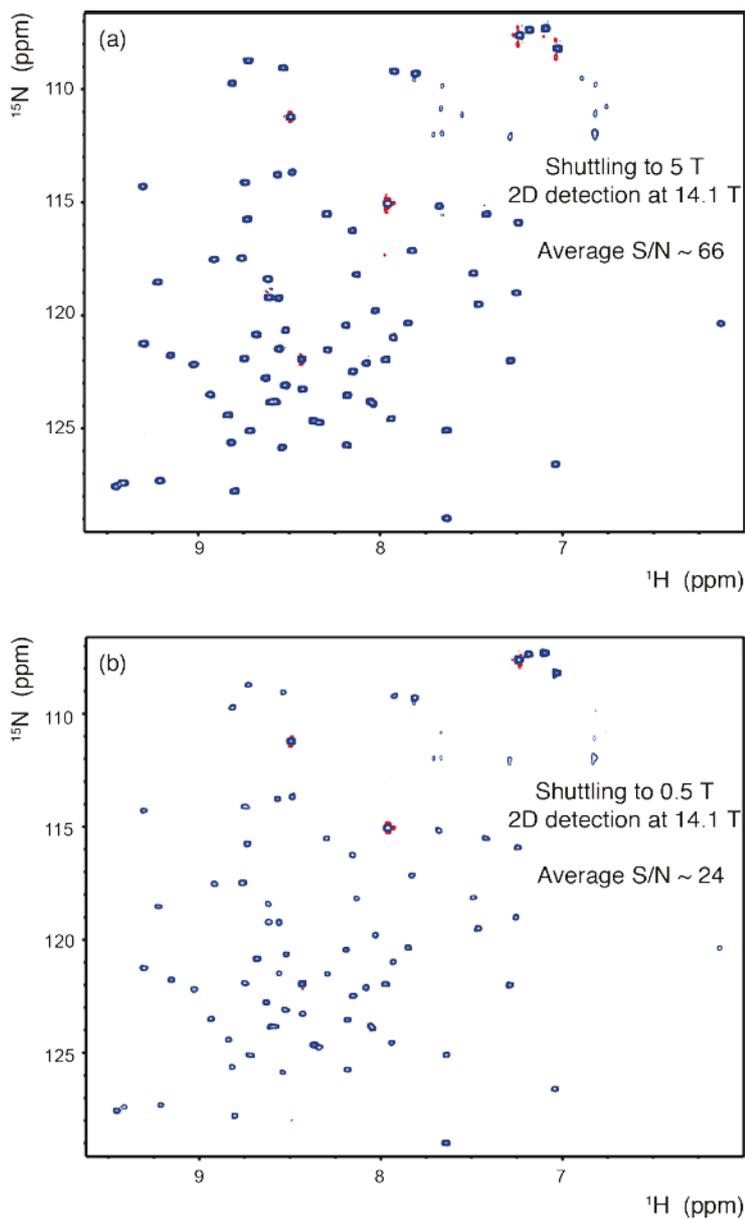


Figure 2.9: 2D ^{15}N - ^1H spectra of ^{15}N -labeled Ubiquitin detected at 14.1 T, with the shortest relaxation delays at (a) 5 T and (b) 0.5 T.

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

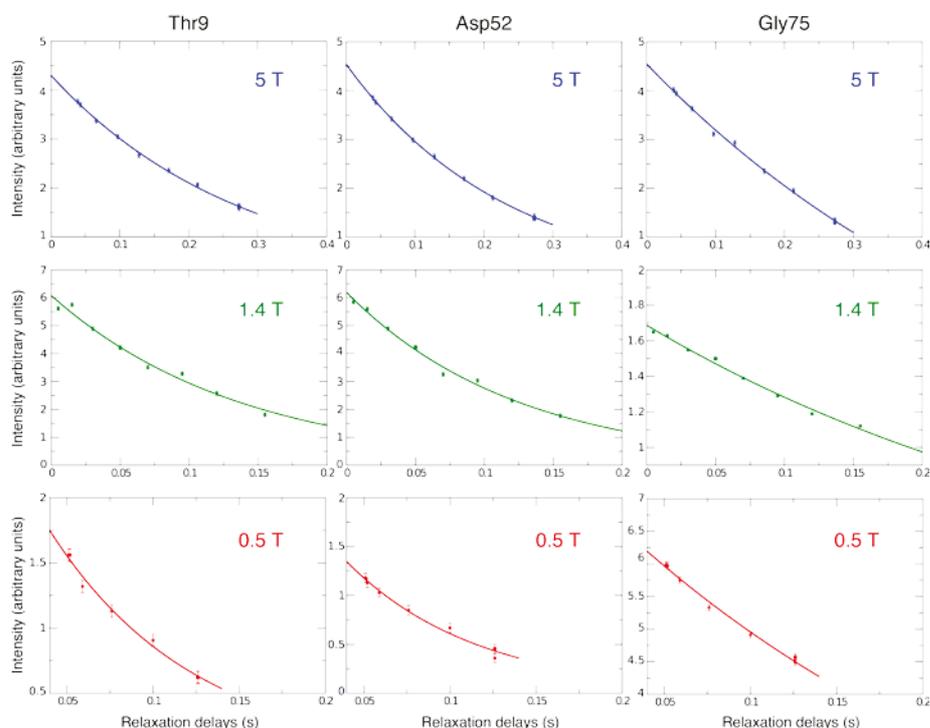


Figure 2.10: Relaxation decays and mono-exponential fits of a few signals at three representative magnetic fields. Results of a single relaxation experiment (total duration 11 to 13 hours) are shown for residues Thr9 (left); Asp52 (center); and Gly75 (right) at the following magnetic fields B_0^{low} : 5 T (top); 1.4 T (center); and 0.5 T (bottom). The solid lines show the results of mono-exponential fits of the relaxation decays.

The relaxation curves at low fields were fitted to mono-exponential decays (Figure 2.10). The longitudinal relaxation rates $R_1(B_0) = 1/T_1(B_0)$ of ^{15}N nuclei in the backbone of ubiquitin determined at 10 different fields are shown in Figure 2.11. Interestingly, variations along the sequence are more pronounced at low fields. In particular, the $\beta 1$ - $\beta 2$ turns, known as a dynamic region, shows a lower-than-average increase of R_1 values when the magnetic field decreases (Figure 2.11).

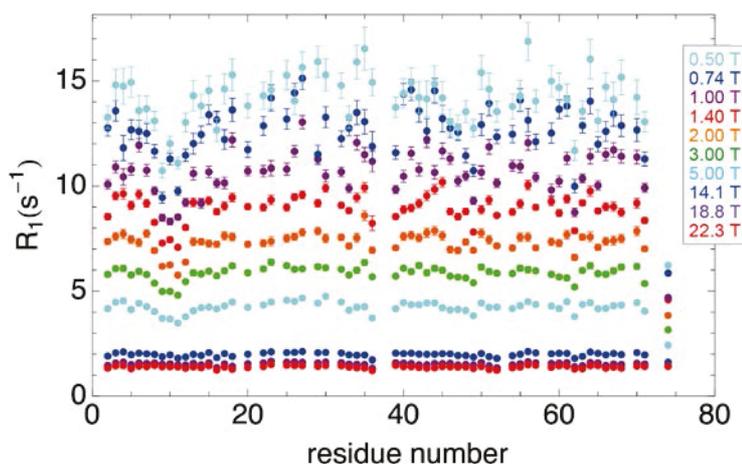


Figure 2.11: Experimental longitudinal relaxation rates $R_1 = 1/T_1$ of ^{15}N in backbone amide groups of ubiquitin as a function of magnetic fields $0.5 \leq B_0^{low} \leq 5$ T and $14.1 \leq B_0^{high} \leq 22.3$ T. From bottom to top: $B_0^{high} = 22.3, 18.8, 14.1$; $B_0^{low} = 5.0, 3.0, 2.0, 1.4, 1.0, 0.74$, and 0.5 T. Note that all rates increase with decreasing field; and the lower B_0^{low} , the greater the variations of the rates along the protein sequence.

4.3 Control of the sample temperature

Particular care has been taken to control the temperature in these experiments, since temperature regulation at low field B_0^{low} is not yet feasible in our prototype. The sample temperature was monitored using chemical shifts differences of a few selected pairs of residues in ^1H - ^{15}N HSQC spectra of ubiquitin in the manner of Salvi *et al*¹⁶⁷ (Figure 2.12): $\delta(^1\text{H}^{\text{N}} \text{L8}) - \delta(^1\text{H}^{\text{N}} \text{V5})$, $\delta(^1\text{H}^{\text{N}} \text{L8}) - \delta(^1\text{H}^{\text{N}} \text{I44})$, $\delta(^1\text{H}^{\text{N}} \text{L8}) - \delta(^1\text{H}^{\text{N}} \text{H68})$. These differences in chemical shifts were then fitted to a linear function of the temperature (Figure 2.13). Table 2.1 summarizes the temperatures of all experiments performed in our study.

Experiments at $B_0 = 22.3$ T (950 MHz for ^1H) were acquired at $T = 298.5$ K, about 2 K higher than all other experiments. In order to correct for this temperature difference we used experiments at 14.1 T as a reference. Using a full set of relaxation rates (R_1 , R_2 , NOE, longitudinal η_z and transverse η_{xy} cross-correlations) we estimated the overall correlation times at $B_0 = 14.1$ T and $B_0 = 22.3$ T. We assigned the variation of τ_c to the change of viscosity of water with temperature. Taking this effect into account, we corrected the experimental relaxation rates observed at 22.3 T prior to analysis of the relaxation rates in terms of dynamics.

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

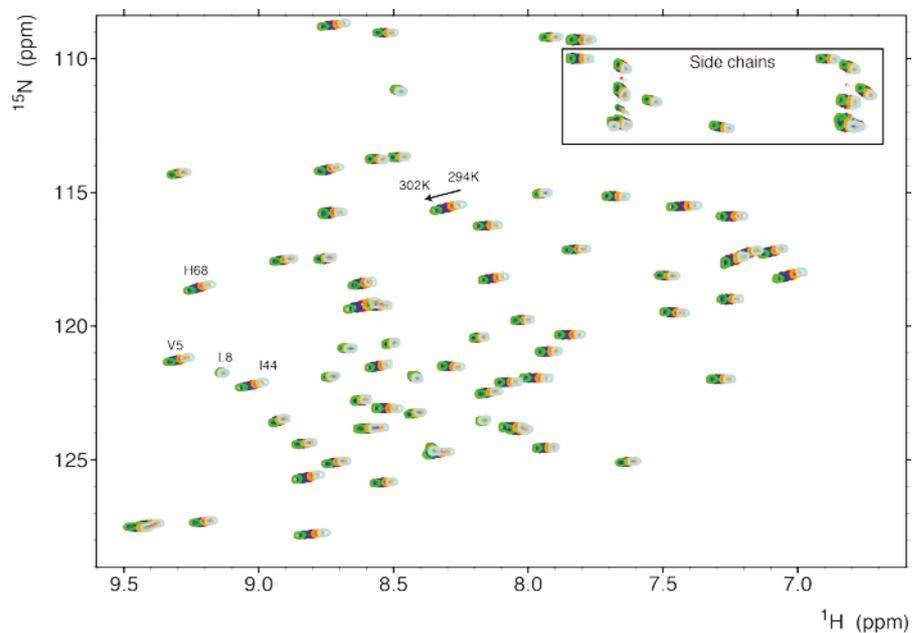


Figure 2.12: ^1H - ^{15}N -HSQC spectra of human ubiquitin as a function of the temperature

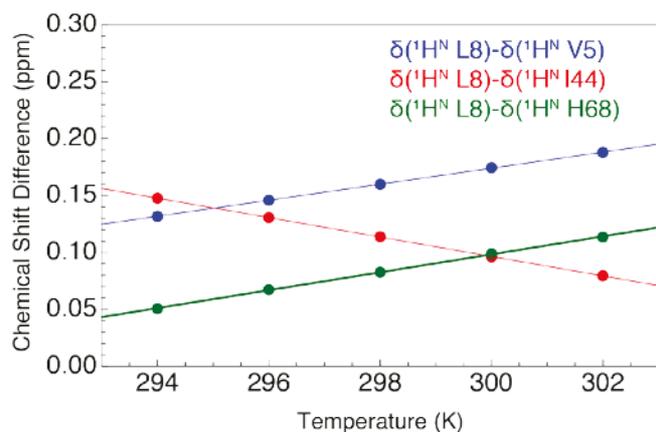


Figure 2.13: Evolutions of proton chemical shift difference in three pairs of residues.

Table 2.1: Sample temperature for different experiments

Experiment	Field (T)	Concentration (mM)	Temperature (K)
R ₁	0.5	3	296.33
R ₁	0.74	3	297.09
R ₁	1	3	297.20
R ₁	1.4	3	296.06
R ₁	2	3	296.26
R ₁	3	3	296.59
R ₁	5	3	296.95
R ₁	14.1	0.2	296.39
R ₂	14.1	0.2	296.45
NOE	14.1	0.2	296.52
η_{xy}	14.1	0.2	296.39
η_z	14.1	0.2	296.56
R ₁	14.1	3	296.62
R ₂	14.1	3	296.41
NOE	14.1	3	296.42
η_{xy}	14.1	3	296.64
η_z	14.1	3	296.64
R ₁	18.8	3	296.38
R ₂	18.8	3	296.46
NOE	18.8	3	296.38
η_{xy}	18.8	3	296.49
η_z	18.8	3	296.46
R ₁	22.3	3	298.53
R ₂	22.3	3	298.49
NOE	22.3	3	298.49
η_{xy}	22.3	3	298.52
η_z	22.3	3	298.51

5. Presentation of the ICARUS protocol

Cross-relaxation pathways lead to multi-exponential decays, which in high fields can usually be transformed into mono-exponential decays by suitable pulse sequences^{120,165}. In addition, cross-relaxation during the shuttle transfers makes that the initial spin terms cannot be pure. Since it is not possible to apply any *rf*-pulses at low fields, systematic deviations from mono-exponential decays must be taken into account in the analysis¹⁴⁴. We have developed a protocol dubbed Iterative Correction for the Analysis of Relaxation Under Shuttling (ICARUS). ICARUS is a MATLAB (MathWorks, Inc.) package, which permits a quantitative analysis in terms of local dynamics of longitudinal relaxation rates (¹⁵N R₁) recorded at a series of low magnetic fields. In order to achieve this, ICARUS is adapted from a well-known program developed for the analysis of high-field nitrogen-15 relaxation, using model-free or extended model-free spectral density functions. This package uses both ROTDIF¹⁶⁸ to obtain overall tumbling parameters and DYNAMICS¹⁶⁹ to fit microdynamics parameters (S^2 , τ_c , τ_{loc} , S^2_f , τ_f).

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

The full ICARUS procedure is summarized in the following flow chart (Figure 2.14). First, the analysis of relaxation rates was carried out (in terms of overall tumbling and microdynamics) at all three high fields 14.1, 18.8, and 22.3 T using the programs ROTDIF¹⁶⁸ and DYNAMICS¹⁶⁹. The parameters resulting from this initial step were then used to predict the deviations from simple exponential decays in a spin system that comprises one ¹⁵N-¹H pair and two remote protons (Figure 2.15). For each field B_0^{low} , we simulated the relaxation of each ¹⁵N-¹H pair in ubiquitin *during* shuttling, with constant velocity (Figure 2.16). The deviations between the calculated ‘apparent’ nitrogen-15 relaxation rates and the ‘true’ low-field relaxation rates were then used to correct for systematic errors in the experimental rates. In a second iteration, longitudinal relaxation rates at all 10 fields and ¹⁵N-¹H NOE’s at the three high fields 14.1, 18.8, and 22.3 T were used as input to the program DYNAMICS. This cycle was reiterated four times to achieve a satisfactory convergence for all residues. The typical corrections to low-field longitudinal magnetic fields varied from 4.5 to 13%. Effects of the cross-correlation of the fluctuations of nitrogen-15 chemical shift anisotropy and the ¹⁵N-¹H dipolar coupling are dominant above 3 T, with average corrections ranging from 5.1% at 3 T to 9.2% at 5 T, while effects of ¹⁵N-¹H dipolar cross-relaxation dominate below 2 T, with corrections on the order of 11% at fields below 1 T. In the following subsections we will discuss in details several points of the ICARUS procedure.

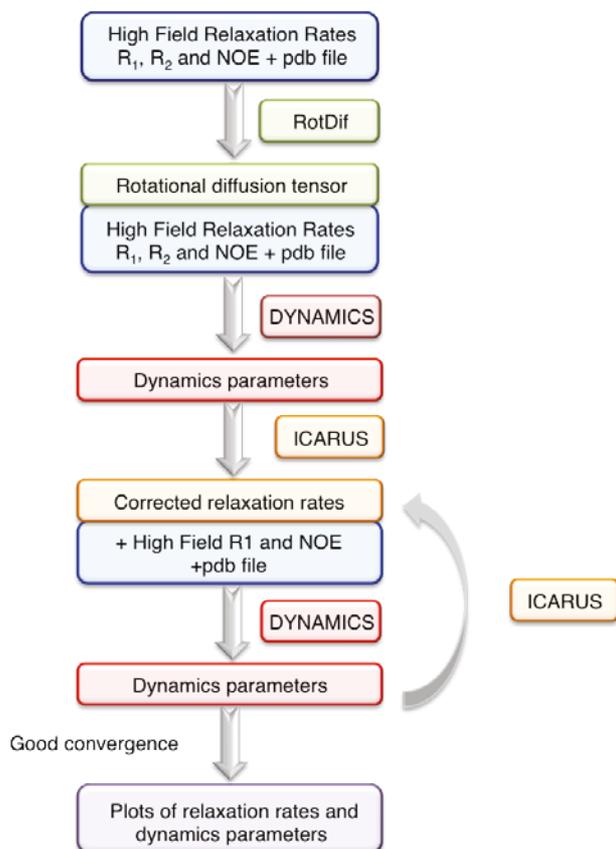


Figure 2.14: Flowchart of ICARUS

5.1 Spectral density functions

- The model-free spectral density function $J(\omega)$ is:

$$J(\omega) = \frac{S^2\tau_c}{(1 + (\omega\tau_c)^2)} + \frac{(1 - S^2)\tau_e}{(1 - (\omega\tau_e)^2)} \quad (2.1)$$

where S^2 is the generalized order parameter, τ_c the overall isotropic rotational correlation time of the molecule, and $\tau_e = \tau_c\tau_{loc}/(\tau_c + \tau_{loc})$, where τ_{loc} is a single effective correlation time that describes all internal motions.

- The spectral density function $J(\omega)$ used for the extended model-free approach is:

$$J(\omega) = \frac{S^2\tau_c}{(1 + (\omega\tau_c)^2)} + \frac{(1 - S_f^2)\tau_f^{eff}}{(1 + (\omega\tau_f^{eff})^2)} + \frac{(S_f^2 - S^2)\tau_s^{eff}}{(1 + (\omega\tau_s^{eff})^2)} \quad (2.2)$$

with $\tau_f^{eff} = \tau_c\tau_f/(\tau_c + \tau_f)$ and $\tau_s^{eff} = \tau_c\tau_s/(\tau_c + \tau_s)$ where τ_f (respectively τ_s) is the correlation time for fast (respectively slow) internal motions. The generalized order parameter can be expressed as a product $S^2 = S_f^2S_s^2$ where S_f^2 and S_s^2 are generalized order

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

parameters describing fast and slow motions. The models used in simulations of spin dynamics by ICARUS always correspond to the model selected in the previous step by DYNAMICS.

5.2 Spin system

All simulations of spin dynamics were carried out for the following spin system comprising one nitrogen-15 and 3 protons with the following distances and interactions:

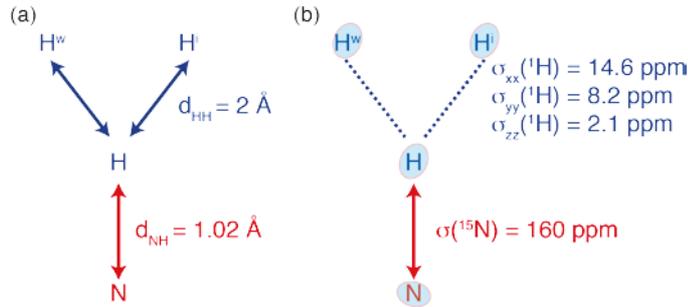


Figure 2.15: Spin system used in ICARUS protocol with (a) the distance and (b) the CSA values.

The relaxation matrix \hat{R} was approximated by:

$$\hat{R} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\theta_H & \rho_H & \sigma_{NH} & \delta_H & \sigma & 0 & \sigma & 0 \\ -\theta_N & \sigma_{NH} & \rho_N & \delta_N & 0 & 0 & 0 & 0 \\ -\theta_{NH} & \delta_H & \delta_N & \rho_{NH} & 0 & \sigma & 0 & \sigma \\ -\theta_H^i & \sigma & 0 & 0 & \rho_H^i & 0 & 0 & 0 \\ -\theta_{NH}^i & 0 & 0 & \sigma & 0 & \rho_{NH}^i & 0 & 0 \\ -\theta_H^w & \sigma & 0 & 0 & 0 & 0 & \rho_H^w & 0 \\ -\theta_{NH}^w & 0 & 0 & \sigma & 0 & 0 & 0 & \rho_{NH}^w \end{pmatrix} \quad (2.3)$$

with⁸⁴

$$\begin{aligned} \rho_H = & \frac{1}{10} c_{NH}^2 [J(\omega_H - \omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] \\ & + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ & + \frac{2}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)] \end{aligned} \quad (2.4)$$

$$\rho_N = \frac{1}{10} c_{NH}^2 [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H - \omega_N)] + \frac{2}{15} c_N^2 J(\omega_N) \quad (2.5)$$

$$\begin{aligned}\rho_{NH} &= \frac{3}{10} c_{NH}^2 J(\omega_N) + \frac{2}{15} c_N^2 J(\omega_N) + \frac{3}{10} c_{NH}^2 J(\omega_H) \\ &\quad + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ &\quad + \frac{2}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)]\end{aligned}\quad (2.6)$$

$$\begin{aligned}\rho_H^i &= \frac{1}{10} c_{NH}^2 [J(\omega_H - \omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] \\ &\quad + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ &\quad + \frac{1}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)]\end{aligned}\quad (2.7)$$

$$\begin{aligned}\rho_{NH}^i &= \frac{1}{10} c_{NH}^2 [3J(\omega_N)] + \frac{2}{15} c_N^2 J(\omega_N) + \frac{1}{10} c_{NH}^2 [3J(\omega_H)] \\ &\quad + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ &\quad + \frac{1}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)]\end{aligned}\quad (2.8)$$

$$\begin{aligned}\rho_H^w &= \frac{1}{10} c_{NH}^2 [J(\omega_H - \omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] \\ &\quad + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ &\quad + \frac{2}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)]\end{aligned}\quad (2.9)$$

$$\begin{aligned}\rho_{NH}^w &= \frac{1}{10} c_{NH}^2 [3J(\omega_N)] + \frac{2}{15} c_N^2 J(\omega_N) + \frac{1}{10} c_{NH}^2 [3J(\omega_H)] \\ &\quad + \frac{2}{15} [c_{Hxx}^2 + c_{Hyy}^2 - c_{Hxx}c_{Hyy}] J(\omega_H) \\ &\quad + \frac{1}{10} c_{HH}^2 [J(0) + 3J(\omega_H) + 6J(2\omega_H)]\end{aligned}\quad (2.10)$$

$$\sigma_{NH} = \frac{1}{10} c_{NH}^2 [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (2.11)$$

$$\delta_H = \frac{-4}{10} c_{NH} [c_{Hxx} J(\omega_H)_{xx} + c_{Hyy} J(\omega_H)_{yy}] \quad (2.12)$$

$$\delta_N = \frac{-4}{10} c_{NH} c_N J(\omega_N)_c \quad (2.13)$$

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

$$\sigma = \frac{1}{10} c_{HH}^2 [J(2\omega_H) - J(0)] \quad (2.14)$$

$$\theta_N = \theta_H = \theta_{NH} = \theta_H^i = \theta_H^w = \theta_{NH}^w = \theta_{NH}^i = 0 \quad (2.15)$$

$$J(\omega_H)_{xx} = \frac{[3(\cos(\beta_{Hxx}))^2 - 1]}{2} J(\omega_H)$$

$$J(\omega_H)_{yy} = \frac{[3(\cos(\beta_{Hyy}))^2 - 1]}{2} J(\omega_H) \quad (2.16)$$

where $c_{NH} = \mu_0 \gamma_H \gamma_N h / (8\pi^2 d_{NH}^3)$, $c_{HH} = (\mu_0 \gamma_H^2 h / 8\pi^2 d_{HH}^3)$, $c_N = \Delta\sigma_N \omega_N$, $c_{Hxx} = \omega_H (\sigma_{Hxx} - \sigma_{Hzz})$ and $c_{Hyy} = \omega_H (\sigma_{Hyy} - \sigma_{Hzz})$; μ_0 is the permittivity of free space γ_H (γ_N) is the gyromagnetic ratio of the proton (respectively of the nitrogen-15 nucleus); h is Planck's constant; ω_H (ω_N) is the Larmor frequency of the proton (respectively of the nitrogen-15); $\Delta\sigma_N$ (= 160 ppm) is the average value of the anisotropy of the ^{15}N chemical shift (CSA); σ_{Hxx} (= 14.6 ppm), σ_{Hyy} (= 8.2 ppm) and σ_{Hzz} (= 2.1 ppm) are the main components of the ^1H CSA tensors; d_{NH} (= 1.02 Å) is the internuclear nitrogen-hydrogen distance; d_{HH} (= 2.1 Å) is the effective distance between the proton H and the other two protons (H^i and H^w). β_{Hxx} (= $90\pi/180$) and β_{Hyy} (= $99\pi/180$) are the angles between the NH vector and the respective components of the ^1H CSA tensors.

Commentaire [CC2]: Here it is!

5.3 First step in ICARUS

In a first step, ROTDIF and DYNAMICS are used to obtain hydrodynamic and microdynamic parameters from high-field relaxation data (^{15}N R_1 , R_2 and NOE at 14.1 T, 18.8 T and 22.3 T) only (table 2.2). Using ROTDIF, the parameters of the overall rotational diffusion tensor for an axially symmetric model were estimated from relaxation rates at 14.1 T (0.2 mM or 3 mM), 18.8 T (3 mM) and 22.3 T (3 mM) (Table 2.2).

Table 2.2: Parameters for the overall rotational diffusion tensor

Magnetic Field	14.1 T	14.1 T	18.8 T	22.3 T*
Concentration	0.2 mM	3 mM	3 mM	3 mM
$\tau_c = 6\text{Tr}(D)$	4.22 ± 0.15 ns	4.89 ± 0.1 ns	4.84 ± 0.2 ns	4.84 ± 0.16 ns
$D_{\text{par}}/D_{\text{per}}$	1.22 ± 0.08	1.22 ± 0.04	1.18 ± 0.08	1.20 ± 0.07
θ	$112^\circ \pm 12^\circ$	$112^\circ \pm 7^\circ$	$120^\circ \pm 15^\circ$	$119^\circ \pm 12^\circ$
ϕ	$157^\circ \pm 29^\circ$	$157^\circ \pm 15^\circ$	$155^\circ \pm 36^\circ$	$157^\circ \pm 27^\circ$

* Analysis carried out with temperature-corrected rates as explained in text.

5.4 Iterative analysis with ICARUS

The microdynamic parameters obtained in the first step are used next to calculate the expected evolution of each spin system during the transfer between the high- and low-field positions as well as during the relaxation and stabilization delays (Figure 2.16). The elements of the relaxation matrix depend on the magnetic field. The transfer durations shown in table 2.3 and Figure 2.16 have been determined by using optical sensors. The position of the shuttle is described by a trajectory with a lag time at the starting position followed by motion assumed to be for simplicity at a constant speed of about 11 m/s. There is an additional pre-shuttling delay of 26 ms at high field. In the simulations, the time-dependence of the populations is obtained from the continuous integration of the evolution under the relaxation matrix along the trajectory from the high- to the low-field positions.

To simulate relaxation during the transfer, a time-dependent relaxation matrix is derived as a function of the position of the shuttle. The same approach is employed for relaxation during the back-transfer from the low- to the high-field position, with a somewhat lower speed, also assumed to be constant, of about 6.5 m/s. Relaxation during the delay at the low-field position is also simulated. An additional 40 ms delay accounts for the minimum duration of the stay at the low-field position. The last step is the calculation of the relaxation of the spin system during the 100 ms stabilization delay once the shuttle has returned to the high-field position.

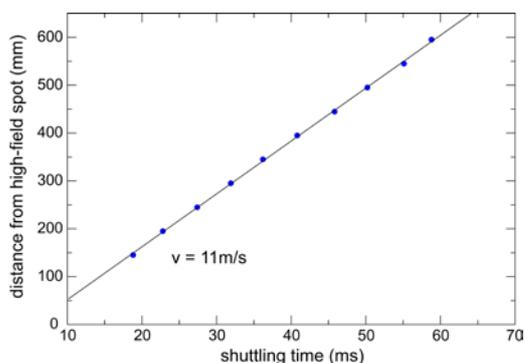


Figure 2.16: Correlation between the height above the high-field position and the shuttling time. The speed of the shuttle is higher for the upward (going up above the magnet, *i.e.* moving towards the low-field position) motion ($v = 11$ m/s) than for the downward motion ($v = 6.5$ m/s). The correlation confirms that the velocity is constant after an initial lag time. (In practice, relaxometry data were often recorded with slightly different pneumatic settings, which explains small variations between these data and those presented in Table 2.3)

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

Table 2.3: Position and shuttling time used for simulations

B ₀ (T)	Position (cm)	Shuttling time (ms)
5	27.1	41.1
3	31.1	44.1
2	34.4	46.4
1.4	37.35	48.3
1	39.8	49.9
0.75	42.55	51.6
0.5	46.2	53.7

All simulated relaxation decays were fitted to mono-exponential functions. The deviations between the fitted relaxation rates obtained from the simulated experiment and the longitudinal relaxation rates of nitrogen-15 nuclei calculated from the set of dynamic parameters were used to correct the experimental relaxation rates for the next iteration of ICARUS. The corrected experimental rates are used, along with high-field relaxation data (¹⁵N *R*₁ and ¹⁵N-¹H NOE) as input for DYNAMICS in the subsequent steps of the ICARUS analysis.

The convergence of the analysis is fast for most residues, although up to four steps may be required for all residues to converge as described in the flow chart (Figure 2.14).

5.5 Error evaluation

The determination of relaxation rates at low fields is more challenging than at high fields. Our analysis takes some of the complexity of the spin systems into account, with approximate evaluations of cross-relaxation pathways involving a defined set of interactions (proton-proton dipole-dipole couplings, CSA tensors...). It is unlikely that spectral noise is the main source of errors. In order to evaluate systematic errors, we implemented a jack-knife procedure. We have carried out the ICARUS analysis for 7 sets of data. Each data set included the longitudinal relaxation rates *R*₁ and NOE's at all high field (14.1, 18.8, and 22.3 T) as well as the longitudinal relaxation rates at 6 of the 7 low fields. All the results below show average of the values obtained in the 7 ICARUS analyses of the jack-knife procedure. The errors are equal to the standard deviations of these datasets multiplied by $\sqrt{(7 - 1)}$.

5.6 Effective distance between H^N-H^W and H^N-Hⁱ

The effective distances between the H^N amide proton and the two additional protons are critical to scale the corrections in the ICARUS procedure. In order to determine these distances, we measured longitudinal relaxation rates at 14.1 T with experiments similar to the

shuttling method, but where the longitudinal nitrogen-15 polarization is allowed to evolve during fixed delays before and after the relaxation delay and where no rf-pulses are applied during the relaxation delay. The optimal distances d_{HH} were found to vary between 1.6 and 2.7 Å with an average of 2.1 Å. This result is confirmed by a computation of the sum of dipolar interactions with all protons where the median value corresponds to an effective distance $d_{\text{HH}} = 2.07$ Å. Site-specific variations of d_{HH} obtained in the two approaches were only weakly correlated so that we decided to use an average $d_{\text{HH}} = 2.1$ Å for all residues. In order to evaluate the potential systematic errors of the resulting order parameters, we carried out a complete ICARUS analysis for a series of distances $1.7 < d_{\text{HH}} < 2.6$ Å.

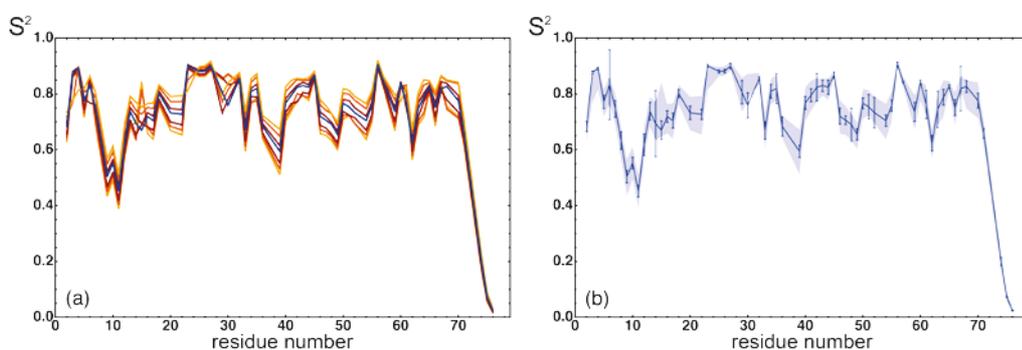


Figure 2.17: Dependence of order parameters upon the effective distance of neighboring protons employed in the ICARUS analysis. (a) Order parameters obtained with $d_{\text{HH}} = 2.1$ Å are shown in purple, other curves represent order parameters obtained with $d_{\text{HH}} = 1.7, 1.8, 1.9, 2.2, 2.4,$ and 2.6 Å. With few exceptions, the order parameters increase with increasing values of d_{HH} . (b) Order parameters obtained with $d_{\text{HH}} = 2.1$ Å are shown in blue, with error intervals obtained from the jackknife analysis of ICARUS. The grey region shows the interval between the maximum and minimum order parameters obtained in the ICARUS analysis with the following values of $d_{\text{HH}} = 1.7, 1.8, 1.9, 2.0, 2.2, 2.4,$ and 2.6 Å

Results are shown in Figure 2.17. Order parameters of some sites tend to be sensitive to the distance d_{HH} , but the main features of our analysis, in particular the low order parameters found in the $\beta 1$ - $\beta 2$ turn, remain stable regardless of the distance d_{HH} .

6. Application to Ubiquitin

The results of this analysis are shown in Figure 2.18. As expected from studies of ^{15}N and ^{13}C relaxation^{99,154}, residual dipolar couplings^{170–172} and molecular dynamics^{157,158,173}, we find that ubiquitin is fairly rigid. However, we determined the order parameters to be significantly lower than in earlier relaxation-based studies. We may compare (Figure 2.18-a) (Green) the order parameters resulting from the analysis of relaxation rates at all 10 fields,

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

(Red) those obtained from relaxation rates at 14.1 T only, and (Orange) those obtained at three high fields (14.1, 18.8 T and 22.3 T). With few exceptions, the order parameters resulting from our analysis of relaxation at ten magnetic fields are the lowest. In particular, the dynamics of the crucial β 1- β 2 turn (residues 7-12) can be best described by an extended model-free¹⁷⁴ spectral density function with similar timescales for all six residues (Figure 2.19). A global fit of these six residues gives a common effective timescale $\tau_{7-12} = 2$ ns. This is in a good agreement with the well-documented hypothesis of a collective motion¹⁷⁵.

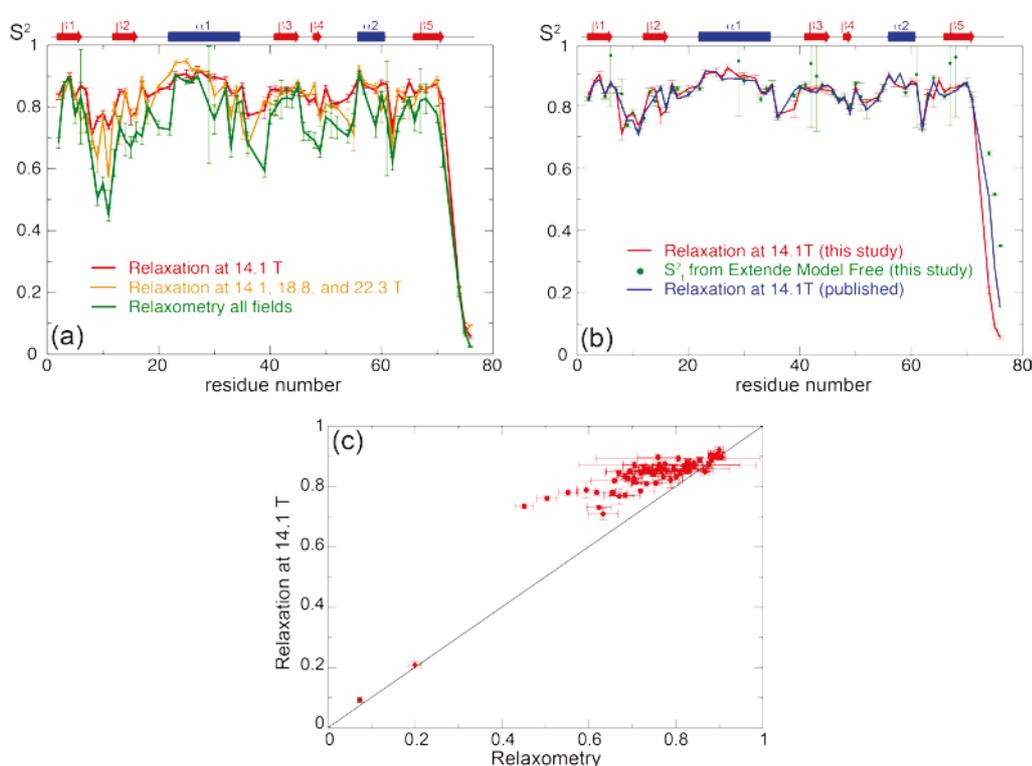


Figure 2.18: (a) Order parameters S^2 in ubiquitin obtained from the analysis of nitrogen-15 relaxation rates, taking into account: (red) relaxation rates at 14.1 T only; (orange) relaxation data at three fields 14.1, 18.8, and 22.3 T; (green) relaxation rates at all 10 fields from 0.5 to 22.3 T. (b) Comparison of order parameters in human ubiquitin. (Red) Order parameters S^2 obtained from our analysis of relaxation rates at 14.1 T only. (Blue) Order parameters derived from the analysis of a different set of nitrogen-15 relaxation rates at 14.1 T⁸⁰. (Green) Order parameters of fast motions for an extended model-free spectral density function in the analysis by DYNAMICS of relaxation rates recorded at 14.1 T only. (c) Comparison of the order parameters of backbone N-H vectors in ubiquitin obtained from the analysis of relaxation at 10 magnetic fields shown on the x-axis while the y-axis displays order parameters obtained by our analysis of relaxation at 14.1 T.

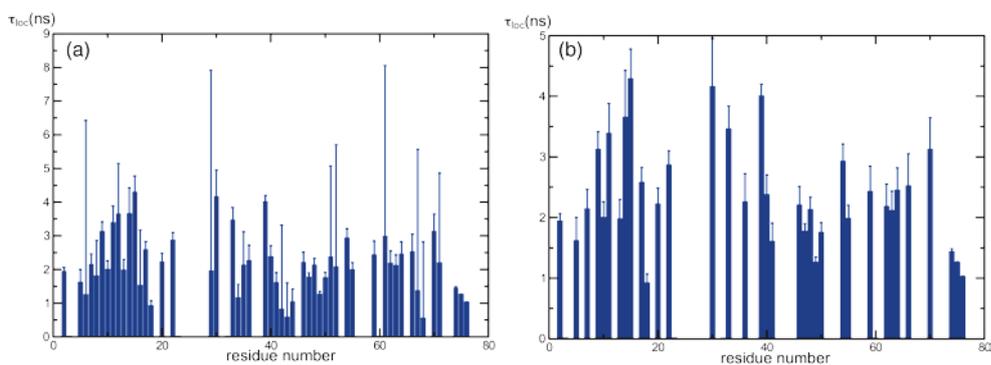


Figure 2.19: Timescales τ_{loc} of local motions determined by the analysis of relaxation rates measured at 10 different fields. The timescales displayed correspond either to τ_{loc} for basic model-free spectral density functions, or to τ_s for extended model-free spectral density functions. The large error bars reflect the instabilities of the selected model when applying a jack-knife analysis. (a) Timescales τ_{loc} of local motions determined by the analysis of relaxation rates measured at 10 different fields are shown for all residues while (b) only residues where τ_{loc} is significant are shown with $\Delta\tau_{\text{loc}} < \tau_{\text{loc}}/3$.

This motion was so far believed to occur at a slower, so-called supra- τ_c , timescale $\tau_{7-12} > \tau_c$, since RDC studies indicated large-amplitude motions while relaxation at a single high field failed to detect such motions. Interestingly, lower order parameters and motions on similar timescales are also found at the C-terminus of helix α_1 and loop α_1 - β_4 (residues 33 and 36), which participate, along with the β_1 - β_2 turn, in the principal mode of ubiquitin dynamics^{171,176}. The relaxation of this principal mode could be described *in silico* by a fit with two timescales, 0.4 and 13 ns¹⁷⁶. Our analysis assumes only one single timescale for the semi-local motions, and the fit leads to an intermediate value, $\tau_{7-12} = 2$ ns, which could well result from effective averaging between these two timescales. More complex models of spectral density functions should open the way to a better agreement between experimental rates and theory.

The inability of high-field relaxation studies to identify these motions is hardly surprising, since most studies have only been carried out at a single field. Therefore, the sampling of the spectral density function was insufficient and did not characterize sufficiently well motions on nanosecond timescales. The consequences of under-sampling of the spectral density function are exacerbated if one uses simple models that do not properly reproduce the actual spectral density functions. Strikingly, the order parameters for fast motions obtained by extended-model free analysis of all relaxation rates match very well order parameters obtained from the analysis of relaxation data at 14.1 T only (Figure 2.18-c). This point is also nicely illustrated by the analysis of a 1.2 μs molecular dynamics trajectory of ubiquitin¹⁷³. In

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

this study, order parameters (S^2) derived from the average orientations of NH vectors were low for the $\beta 1$ - $\beta 2$ turn. However, when the non-exponential correlation functions were forced to fit with a simple extended model-free correlation function, the order parameters became significantly higher, similar to those found in relaxation studies at 14.1 T¹⁵⁴.

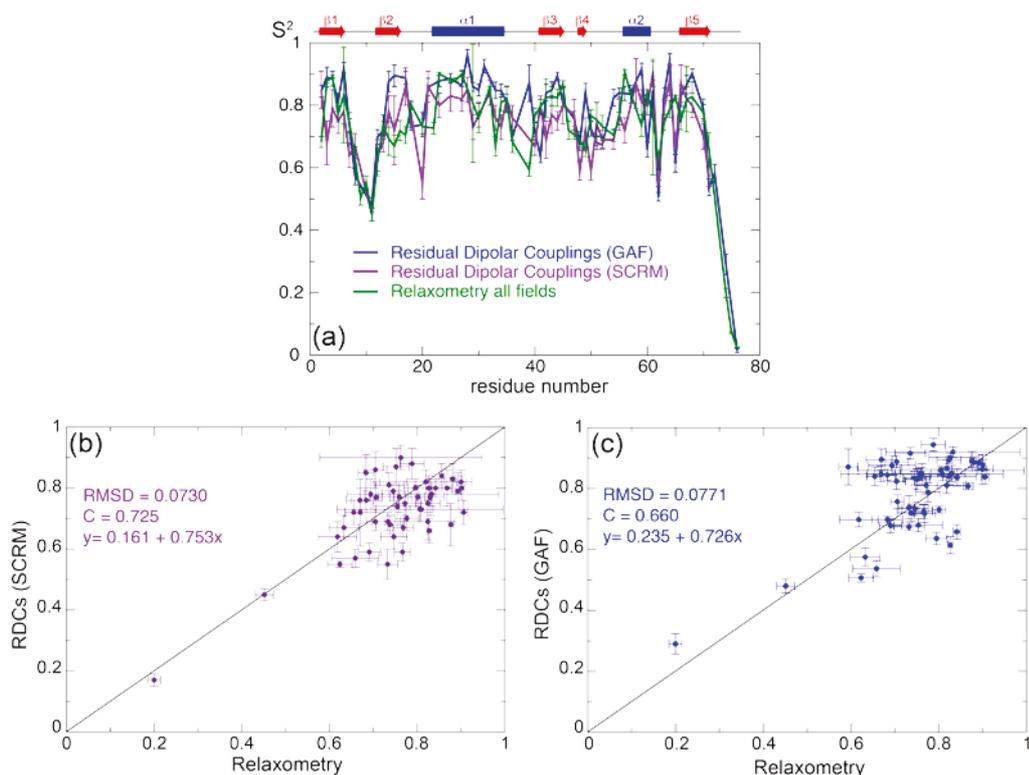


Figure 2.20: (b) Comparison of order parameters obtained from relaxation rates at all fields (green) and from analysis of residual dipolar couplings (RDCs) in large sets of orienting media, either by GAF (blue)⁸⁰ or SCRM (purple)¹⁷¹. Comparison of the order parameters of backbone N-H vectors in ubiquitin obtained from the analysis of relaxation at 10 magnetic fields shown on the x-axis while the y-axis displays order parameters obtained by the following methods: (b) the analysis of residual dipolar couplings measured in 23 different alignment media with the SCRM approach¹⁷¹; (c) the analysis of 36 sets of dipolar couplings with the SF-GAF approach⁸⁰;

Figure 2.20 presents the comparison of orders parameters obtained (Green) from our relaxometry analysis of relaxation at ten magnetic fields and from two independent analyses of residual dipolar couplings (RDCs) in large sets of oriented media using (Blue) the Gaussian Axial Fluctuations (GAF)⁸⁰ or (purple) the Self-Consistent RDC-based Model-free analysis (SCRM)¹⁷¹ approaches. The three profiles are similar. In particular, the order parameters in the $\beta 1$ - $\beta 2$ turn (residues 7-12) are almost identical, so that one expects the amplitudes of

motions in this turn that cannot be detected by relaxation to be very small. Significant correlations between slow supra- τ_c motions of the $\beta 1$ - $\beta 2$ turn and those of the β sheet are therefore unlikely, though correlated motions in the core of the β -sheet cannot be excluded¹⁷⁵. Our relaxometry data show that the whole $\beta 2$ strand, which lies at the edge of the β -sheet, is significantly dynamic. Similarly, studies of the third immunoglobulin binding domain of streptococcal protein G (GB3) have also shown the presence of enhanced motions in the last strand of an otherwise fairly rigid β -sheet¹⁷⁷. These results are different from those of a GAF analysis of dynamics in ubiquitin⁸⁰, where the $\beta 2$ -strand is found to be rigid. However, they are in better agreement with results from the SCRm analysis¹⁷¹.

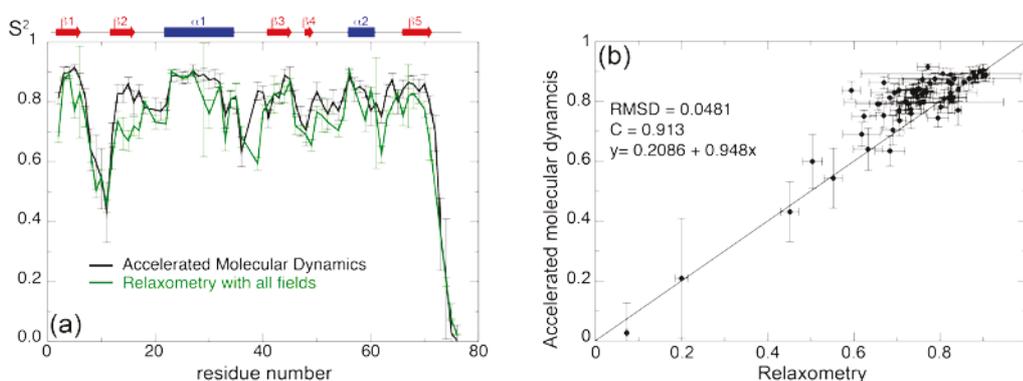


Figure 2.21: (a) Comparison of order parameters in human ubiquitin. (Black) Order parameters derived from the analysis of an accelerated molecular dynamics trajectory⁸⁰ (Green) Order parameters from the analysis of relaxation rates measured at 10 different fields. (b) Comparison of the order parameters of backbone N-H vectors in ubiquitin obtained from the analysis of relaxation at 10 magnetic fields are shown on the x-axis while the y-axis displays order parameters obtained by an accelerated molecular dynamics trajectory of ubiquitin¹⁵⁷.

In addition to this region, our relaxometry method allowed us to detect enhanced dynamics in several loops: $\beta 2$ - $\alpha 1$, $\beta 4$ - $\alpha 2$, and $\alpha 2$ - $\beta 5$ as well as in the $\beta 3$ - $\beta 4$ turn and the $\beta 4$ strand, which lies at the other edge of the β sheet. This is, again, in good qualitative agreement with both RDC studies (Figure 2.20-c-d). Furthermore, the agreement with accelerated molecular dynamics simulations is excellent, with a good correlation coefficient ($R = 0.91$) between the two datasets (Figure 2.21-b). Note, however, that the order parameters found by relaxometry are systematically (albeit only slightly) lower than those obtained by molecular dynamics.

Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

A graphical representation of order parameters in ubiquitin is shown in Figure 2.22. The edges of the β -sheet are occupied by the β 1- β 2 turn and the flexible C-terminal tail, at one end, and by the α 2- β 5 loop and β 3- β 4 turn at the other end, with the β 2 and β 4 strands on each side. All of these regions are found to be dynamic, albeit to different extents. The picture that emerges is a hierarchy of timescales¹⁷⁸ near the main binding interface of ubiquitin that consists of a β -sheet with a core that is flexible on a slow timescale of about 50 μ s^{167,179} while its edges are mobile on faster nanosecond timescales (we could use the image of “butterfly motions”). Between these two timescales, small correlated fluctuations of the β -sheet appear also to be allowed¹⁷⁵. The ability of the edges of the interface to undergo conformational rearrangements on nanosecond timescales would be compatible with an induced fit mechanism in the early stage of binding.

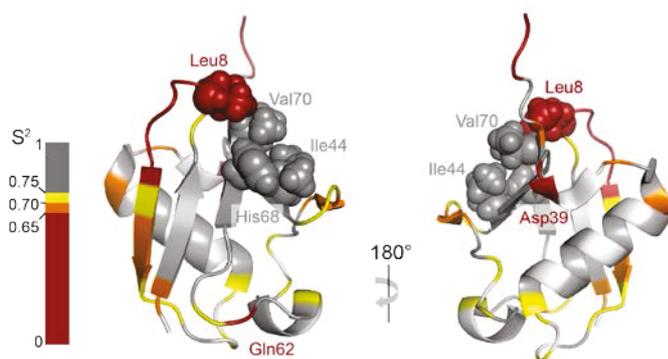


Figure 2.22: Dynamics in ubiquitin on fast nanosecond timescales. Rigid residues with high order parameters $S^2 > 0.75$ are shown in grey, mobile residues with intermediate and small order parameters are shown in yellow ($0.70 < S^2 < 0.75$), orange ($0.60 < S^2 < 0.70$), and red ($S^2 < 0.6$). Residues for which no data are available are shown in white. The main interface with binding partners comprises the side-chains of residues Leu8, Ile44, His68 and Val70 represented by space-filling models.

Figure 2.22 shows a few selected plots of longitudinal relaxation rates $R_1(B_0)$. Dramatic differences can be observed between mobile and rigid residues. A good agreement between experimental and theoretical profiles is observed for most residues. Some dispersion profiles feature systematic discrepancies, which highlight the limitations of current models of spectral density functions.

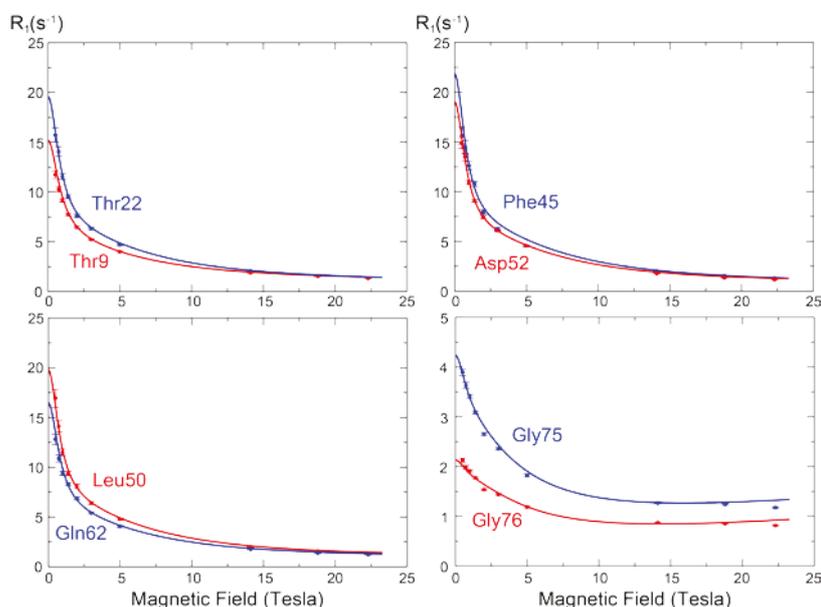


Figure 2.23: Longitudinal relaxation rates $R_1(B_0)$ as a function of the static field (so-called ‘relaxometry dispersion profiles’) for 8 selected residues in ubiquitin. Note that the vertical scale is expanded by a factor 5 for the two C-terminal glycines (bottom right). The blue and red dots show corrected longitudinal relaxation rates, adjusted to compensate for relaxation during shuttling, while the lines show dispersion profiles calculated from the microdynamic parameters obtained in our analysis.

Our experimental data call for the development of new, more sophisticated models. Some relaxation profiles, e.g. for the highly mobile C-terminal glycine residues G75 and G76, present deviations from the theoretical profiles at both high and low magnetic fields, even when postulating a spectral density function comprising a sum of three Lorentzian functions with five adjustable parameters, thus suggesting the presence of a distribution of timescales¹⁸⁰. In particular, with only two timescales for internal motions, the fitted spectral density function is rather flat at high frequencies. Hence, contributions of the chemical shift anisotropy to relaxation lead to an increase in the $R_1(B_0)$ curve between 14 and 23 T, contrarily to the experimental results. Similarly, a small but systematic underestimation of the spectral density $J(\omega = 0)$ in some of the most mobile regions, and hence of the back-predicted transverse relaxation rates, can be understood by postulating a rapid initial decay of the spectral density function at low frequencies. Interestingly, no significant contribution of chemical exchange to transverse relaxation R_{ex} could be detected^{125,131}. The analysis of relaxation at three high fields also leads to underestimate $J(\omega = 0)$ as illustrated by the need to unrealistically postulate R_{ex} contributions to fit *all* relaxation data. Nanosecond fluctuations of

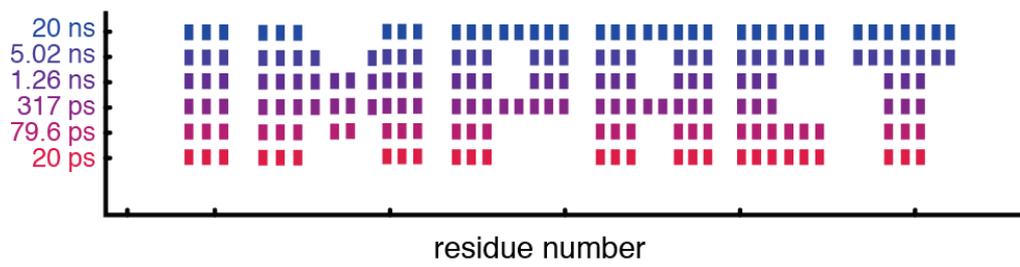
Chapter 2: Internal dynamics in ubiquitin explored by high-resolution relaxometry

the overall diffusion tensor associated with motions of the C-terminal tail, transient oligomerization¹³⁶, and mode coupling of local and global motions^{181–183} may be responsible for these unexpected features.

7. Conclusion

We have measured and analyzed residue-specific relaxation rates in a protein over a range of nearly two orders of magnitude of magnetic fields. Our high-resolution relaxometry approach reveals unexpected motions in the protein ubiquitin. In particular, the motion of the $\beta 1$ - $\beta 2$ turn appears to have larger amplitudes than could be previously identified by relaxation at high fields, in agreement with RDCs and AMD. Until now, discrepancies between high-field relaxation and RDC-based methods were attributed to the cut-off of internal motions by overall rotation. High-field relaxation studies have led to underestimate near- τ_c motions, because relaxation rates at high fields are not sufficiently sensitive to motions in the nanosecond frequency range. Although many proteins, and ubiquitin in particular, are mobile on slow supra- τ_c timescales (slower than overall rotational diffusion), a mere comparison of order parameters obtained from high-field relaxation and RDCs is likely to overestimate the amplitude of such slow motions. This study shows that high-resolution relaxometry with fast sample shuttling allows one to map the spectral density functions in exquisite detail and offers unprecedented information about local motions in proteins on timescales that are faster than or comparable to their overall tumbling.

Chapter 3: Interpretation of motions in disordered proteins



1. *Introduction*

Intrinsically disordered proteins play important roles in biological processes (see General Introduction). An IDP/IDR cannot be described by a single conformation but rather by an ensemble of interconverting conformations. As mentioned above, NMR is a versatile tool for the characterization of the conformational space of proteins and is a leading biophysical technique for the study of IDP/IDRs. Various methods have been developed to study protein dynamics, in particular from nuclear spin relaxation^{76,107}

Typically, two approaches are possible for the interpretation of relaxation data depending on knowledge or assumptions about the spectral density function. In the first case, the functional form of the spectral density function is assumed and the parameters of the motions can be extracted. In contrast, in the original approach of spectral density mapping no assumption is made on the form of spectral density function and more generally on the dynamics of the protein^{101,184}. This second approach has been widely used in the past to characterize protein dynamics^{102,185,186}. Latter, modifications of the original version were introduced leading to so called reduced spectral density mapping^{102,187,188}. Spectral density mapping is based on a simple idea: relaxation rates are expressed as a linear combination of the values of the spectral density function at the eigenfrequencies of a spin system ($0, \omega_I, \omega_S, \omega_S \pm \omega_I$) in the relaxation theory. These relationships can be inverted so that the spectral density function at a few eigenfrequencies of the spin system (or effective frequencies) can be expressed as linear combinations of relaxation rates. Spectral densities or relaxation rates can be further exploited by assuming a form of the spectral density function. Most commonly, protein dynamics is characterized using model-free or extended model-free approaches (see Chapter 1)^{96,97,174}. The original approach of model-free was designed⁹⁵⁻⁹⁷ considering two independent motions: one global (overall rotational tumbling) and one local. As the name indicates, the model-free approach is not based on any particular model of motions but relies on simple assumptions on the form of the correlation functions. The main questionable points in the case of IDPs is the statistical independence of overall and internal motions and the existence of a single overall diffusion tensor. One should keep in mind that the model-free data analysis could lead to misinterpretations of the relaxation data for cases where the assumptions are not suitable⁸⁵. To overcome this problem local motions can be analyzed independently for each residue using “local model-free”¹⁸⁹.

More generally disordered proteins have to be described by a large ensemble of conformations. The many types of motions possible in this conformational space suggest the

Chapter 3: Interpretation of motions in disordered proteins

existence of a continuum of timescales of motions in the nanosecond range. These properties lead to the development of spectral density function including distribution function $f(\tau_c)$ for the correlation times:

$$J(\omega) = C(0) \int_0^{\infty} f(\tau_c) \frac{\tau_c}{1 + (\omega\tau_c)^2} d\tau_c \quad (3.1)$$

To date, two examples of analytical distribution have been used in the literature: the Cole-Cole distribution^{180,190,191} and the Lorentzian distribution¹⁹². These analytical functions exhibit fundamental restrictions such as diverging when $\omega \rightarrow 0$ for instance, which is suppressed by the use of an arbitrary cut-off in the case of the Lorentzian distribution function and an analogous manipulation in the case of the Cole-Cole distribution. Similar treatment can be found in polymer sciences with the interpretation of only R_1 and NOE data. Besides the limits of a particular approach, the choice of the mathematical function describing the distribution function for correlation times introduces a physical bias in the analysis. This can be problematic when appropriate physical models are not known yet. In the case of the Model-Independent Correlation time distribution (MIC approach)¹⁹³, the statistical independence of three types of motions is not required since extended model-free results are considered as a simplified representation of a continuous distribution of correlation times.

Commentaire [FF3]: They use $J(0) = J(1 \text{ rad/s})$

The aim of our work was to develop a model for the interpretation of multiple-fields relaxation data on IDPs with a minimum of assumptions on either protein dynamics or mathematical aspects. We studied a fragment of the Homeoprotein Engrailed 2 comprising both ordered and disordered regions. It leads to the development of a new approach, called *IMPACT*, for the Interpretation of Motions by a Projection onto an Array of Correlation Times. After a short description of Engrailed 2, we will discuss the principle of the *IMPACT* approach as well as the results obtained on this protein.

2. *Engrailed*

Homeoproteins constitute a large class of transcription factors highly conserved amongst species¹⁹⁴. Engrailed 2 is a homeoprotein with multiple roles such as transcriptional and translational regulation, secretion and internalization¹⁹⁵. Engrailed 2 possesses a well-folded DNA binding domain also called homeodomain (200-259) and a long, ~200 residues, disordered N-terminal domain. The homeodomain of Engrailed 2, as some others homeoprotein described in the literature, consists of three helices. Helices 1 and 3 adopt an antiparallel conformation and helix 2 makes an angle of ~60° with the two other helices (Figure 3.1). This fold was found very similar in both *drosophilia* and *chicken*^{196,197}.

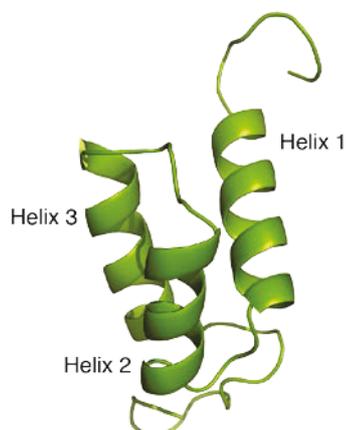


Figure 3.1: Structure of the homeodomain of chicken Engrailed 2 (pdb code: 3ZOB)¹⁹⁷.

To date, the structure of the homeodomain is well known but the dynamics of the protein have not been characterized yet. Preliminary results obtained with Paramagnetic Relaxation Enhancements (PREs) indicate that transient contacts exist between the N-terminal disordered extension and the homeodomain¹⁹⁸. In addition, it has been shown that the disordered region modifies the binding affinity of the homeodomain for DNA. The characterization of the conformational space of Engrailed will be very helpful for understanding the regulation of its functional activity but is beyond the scope of this thesis.

For this purpose we studied, a construct, which consists of amino acids 146 to 259 (MW = 13.5 kDa). It contains the well-folded homeodomain (200-259) and a N-terminal 54-residues disordered region (146-199) (Figure 3.2)^{199,200}. The full assignment of backbone and side-chain resonances was obtained prior to our study. An assigned ¹H-¹⁵N HSQC spectrum is shown in Figure 3.3²⁰⁰. Typically NH resonances from the N-terminal region are poorly dispersed in the proton dimension while those from the homeodomain residues are well dispersed. The secondary structure propensity (SSP) scores²⁰¹ (Figure 3.4) confirm the disordered nature of the N-terminal extension and detect the presence of some residual structure for the so-called hexapeptide (169-174). The three helices are clearly identifiable (Figure 3.4).

Chapter 3: Interpretation of motions in disordered proteins

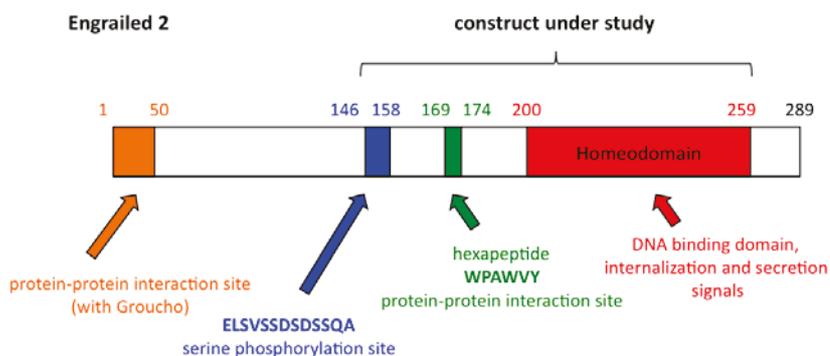


Figure 3.2: Functionally important regions of Engrailed 2 homeoprotein. The DNA binding domain (homeodomain), which contains also the internalization and secretion signals is marked in red. The serine-rich fragment phosphorylated by CK2 kinase is marked in blue. Protein-Protein interaction sites are marked in orange (Groucho protein) and in green (PBX and FoxA2 proteins)²⁰⁰.

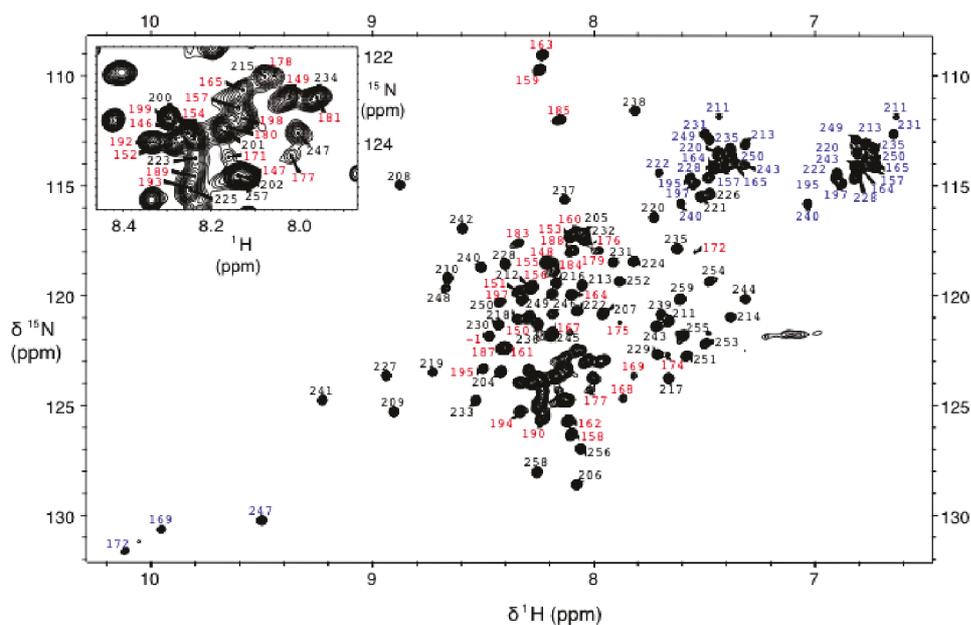


Figure 3.3: Assigned 2D ^1H - ^{15}N spectra of the ^{15}N labeled 146-259 fragment of Engrailed 2 protein acquired at 500 MHz. A zoom of the central crowded region of the spectrum is shown in upper-left. Backbone resonances corresponding to residues in the homeodomain (200-259) and in the N-terminal extension (146-199) are marked in black and red, respectively. The side-chain resonances of Trp, Asn and Gln are indicated in blue²⁰⁰.

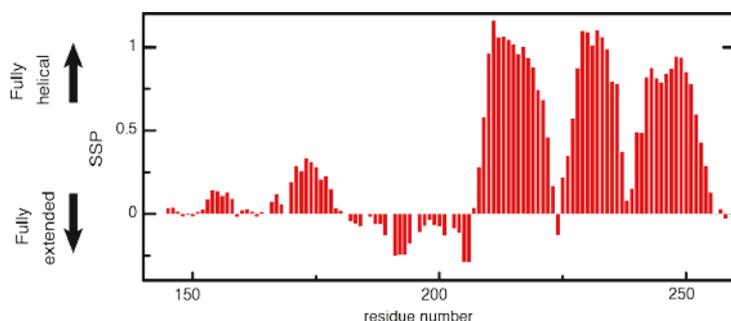


Figure 3.4: Secondary structure propensity (SSP) calculated from the chemical shifts of carbonyl, alpha, and beta carbon-13 nuclei²⁰⁰.

3. High-field relaxation measurements

High-field ^{15}N relaxation measurements were carried out at 5 magnetic fields, with proton Larmor frequencies of: 400, 500, 600, 800 and 1000 MHz, including longitudinal ^{15}N relaxation rates (R_1), the steady-state ^{15}N - $\{^1\text{H}\}$ Nuclear Overhauser Effects (NOEs) and the longitudinal (η_z) and transverse (η_{xy}) cross-relaxation rates due to fluctuations of the nitrogen-15 chemical shift anisotropy (CSA) and the dipolar interaction with the amide proton. Transverse ^{15}N relaxation rates under Carr-Purcell-Meiboom-Gill trains of echoes (R_2^{cpmg}) were measured at 800 MHz. All experiments were recorded on Bruker Avance spectrometers. Experiments at 500 MHz, 800 MHz, 1 GHz and the NOE at 600 MHz have been recorded on triple resonance indirect detection cryogenic probes equipped with z-axis pulsed field gradients. Experiments at 600 MHz were recorded on an indirect detection triple resonance probe with triple axis gradients with radiofrequency coils at room temperature. Experiments at 400 MHz were recorded on a liquid-nitrogen cooled cryogenic probe (Prodigy BBO) equipped with z-axis gradients. More details about the experiments and the pulse sequences used have been described in the first chapter. All the experiments were performed on a uniformly ^{15}N -labeled chicken Engrailed (residues 146-260) sample at 0.6 mM in 40 mM succinate buffer at pH 6^{198,200}. All the experiments were processed using NMRPipe²⁰² and the peak intensities were obtained from nlinLS. The signal intensities of R_1 and R_2 experiments were fitted to a single exponential decays using Curvefit.

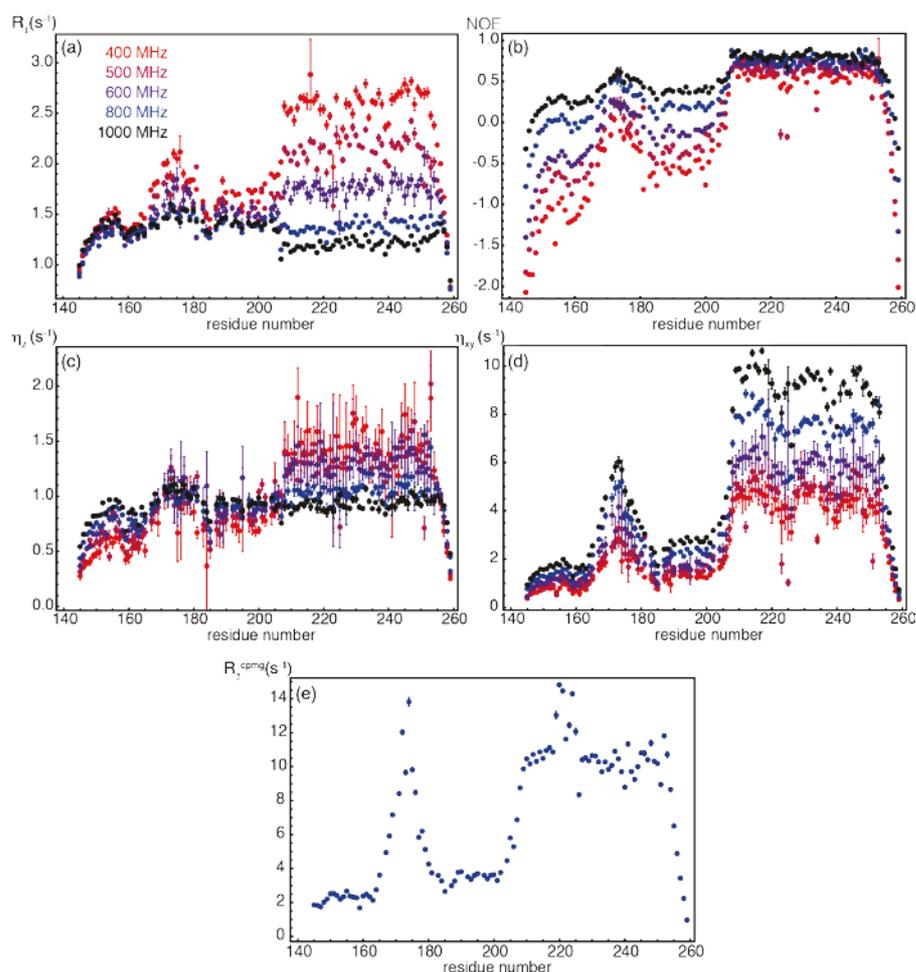


Figure 3.5: Backbone ^{15}N relaxation measurements on Engrailed at 400 MHz (red), 500 MHz (burgundy), 600 MHz (purple), 800 MHz (blue) and 1000 MHz (black). (a) Longitudinal relaxation rates R_1 ; (b) ^{15}N - $\{^1\text{H}\}$ NOE ratios; (c) CSA-DD longitudinal cross-relaxation rates (η_z); (D) CSA-DD transverse cross-relaxation rates (η_{xy}); (e) Transverse relaxation under a CPMG train (R_2^{cpmg}).

The R_1 and NOE values show expected magnetic field dependence. The uniform decrease of the longitudinal relaxation rates R_1 with increasing magnetic field B_0 in the 200-259 homeodomain (Figure 3.5-a) indicates motions in the nanosecond range, resulting from overall rotational diffusion. The R_1 rates are decreasing with the magnetic field while the NOE ratio is increasing (Figure 3.5-a-b). The variations in R_1 with the magnetic field are more pronounced in the homeodomain than in the disordered regions except for the hexapeptide (169-174). Nuclear Overhauser effects are sensitive markers of local order in IDPs and IDRs and have been used as such for many years⁷². At moderate magnetic fields

(400 to 600 MHz), $^{15}\text{N}\{-^{1}\text{N}\}$ steady-state NOE measurements led to a negative signal intensities in the saturated experiments. The ratio saturated versus nonsaturated resulted in a ratio equal to 0 or negative, which is a typical marker for disordered regions (Figure 3.5-b)²⁰³. The variations with the magnetic field are pronounced in disordered regions. In contrast, the NOE values obtained for the homeodoamin are only weakly residue- or field-dependent. Thus, the profiles of NOEs at high-fields, for instance at 23.5 T (or 1 GHz), are more flat than at lower fields, so that the NOE may not be the most appropriate marker of order in IDPs at very high magnetic field. Transverse relaxation rates exhibit high values in the hexapeptide and in the homeodomain. In the hexapeptide, such high values are due both to slower motions on ps-ns timescales and the contributions of chemical exchange (Figure 3.5-e). Longitudinal cross-relaxation rates η_z (Figure 3.5-c) increase with B_0 in the IDR. This reflects the very slow decay, slower than $1/\omega$, of the spectral density function in the range 40-100 MHz (i.e., the range of ^{15}N Larmor frequencies between 9.4 and 23.5 T) as the increase of the amplitude of the CSA interaction counterbalances the decay of the spectral density function with increasing frequency. Significant error bars for the longitudinal cross-correlation rates illustrate the low sensitivity of this experiment. On the other hand, transverse cross-correlation rates η_{xy} (Figure 3.5-d), which depend primarily on $J(\omega = 0)$, exhibit sharp variations at all fields that are strongly correlated with SSP scores. This suggests that transverse cross-correlation rates η_{xy} should become the method of choice to characterize order in IDPs and IDRs.

All together, these rates are in qualitative agreement with SSP scores and provide extensive experimental data, which distinguish clearly the four regions in the construct: the disordered regions (residues 146-168 & 175-199), the hexapeptide (169-174) and the homoeodomain (200-259).

4. Reduced spectral density mapping

R_1 and R_2 relaxation rates as well as $^{15}\text{N}\{-^{1}\text{N}\}$ NOE can be derived from the fluctuations of the Chemical Shift Anisotropy (CSA) of the ^{15}N nucleus and the Dipole-Dipole (DD) interaction between the ^{15}N nucleus and its bond proton (see Chapter 1). The relaxation rates of ^{15}N nuclei are related to the spectral density function $J(\omega)$ that described reorientation motions of the $^{15}\text{N}\text{-}^1\text{H}^{\text{N}}$ vector as⁹³:

$$R_1 = \frac{d^2}{4} [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad (3.2)$$

Chapter 3: Interpretation of motions in disordered proteins

$$R_2 = \frac{d^2}{8} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_N + \omega_N)] + \frac{c^2}{6} [3J(\omega_N) + 4J(0)] \quad (3.3)$$

$$NOE = 1 + \left(\frac{d^2}{4}\right) \left(\frac{\gamma_H}{\gamma_N}\right) (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)) / R_1 \quad (3.4)$$

with $d = \frac{\mu_0 h \gamma_N \gamma_H}{8\pi^2} \left(\frac{1}{r_{NH}^3}\right)$ and $c = \frac{\omega_N}{\sqrt{3}} \Delta\sigma$ where μ_0 is the permittivity of the free space; h is Planck's constant; γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N nuclei, respectively; ω_H and ω_N are the Larmor frequency of ^1H and ^{15}N nuclei, respectively; $\Delta\sigma$ is the difference between the parallel and perpendicular components of the assumed axially symmetric chemical shift tensor of ^{15}N and $r_{NH} = 1.02\text{\AA}$ is the N-H bond length.

In equations (4.1-4.3), $J(\omega)$ is sampled at five frequencies ω_H , ω_N , $\omega=0$, $\omega_H + \omega_N$, $\omega_H - \omega_N$ but, usually, only three rates are measured. Thus, it is not possible to extract the values of the spectral density functions directly from relaxation measurements without making assumptions on the shape of the spectral density function^{96,97,174}. To avoid this problem, Peng and Wagner introduced the measurement of 6 relaxation rates and an analysis they called spectral density mapping^{101,184}. Three years latter a second approach was published to have accurate description of the spectral density function by simplifying the equations for relaxation rates^{102,204}. Indeed, the three frequencies $\omega_H + \omega_N$, ω_H , and $\omega_H - \omega_N$ are close so that the terms $J(\omega_H + \omega_N)$, $J(\omega_H)$ and $J(\omega_H - \omega_N)$ can be replaced by a single term at an effective frequency $J(\varepsilon\omega_H)$, whereas $J(0)$ and $J(\omega_N)$ differ significantly.

First of all, the effective spectral density at high frequency $J(0.87\omega_H)$ (Figure 3.6-a) can be derived from $^{15}\text{N}\{-^1\text{H}\}$ NOE and longitudinal nitrogen-15 relaxation rates (R_1)²⁰⁴

$$J(0.87\omega_H) = \frac{4\gamma_N R_1 (NOE - 1)}{5d^2 \gamma_H} \quad (3.5)$$

In our case, the availability of relaxation rates measured at several fields allowed us to fit the spectral density function at high frequency $J(0.87\omega_H)$ to the following expression:

$$J(\omega) = \lambda + \frac{\mu}{\omega^2} \quad (3.6)$$

in analogy to an earlier study of carbon-13 relaxation¹⁰⁹. The parameters λ and μ are real positive parameters. The results of the fit of the spectral density were used to evaluate contributions to the spectral density at higher frequencies (at $\omega_H \pm \omega_N$) to derive $J(\omega_N)$ from the R_1 rate (Equation 3.2) according to the equation:

Cyril Charlier

$$J(\omega_N) = \frac{\left(R_1 - \frac{d^2}{4} (J(\omega_H - \omega_N) - 6J(\omega_H - \omega_N)) \right)}{\left(\frac{3d^2}{4} + c^2 \right)} \quad (3.7)$$

Overall, contributions of high-frequency terms to R_1 are small²⁰⁵ so that the deviations between the values of $J(\omega_N)$ obtained from conventional approximations¹⁰² and the current method are limited to about 2%, which is commensurate with the estimated precision (Data not shown). The values of $J(\omega_N)$ derived at five magnetic fields are shown in Figure 3.6-b.

The final step of the reduced spectral density mapping evaluates the spectral density function $J(0)$. $J(0)$ is usually determined from R_2 , $J(0.87\omega_H)$ and $J(\omega_N)$ using (3.2). The dependence of $J(0)$ upon the transverse relaxation rate R_2 introduces a bias for residues showing line broadening due to chemical exchange. Significant chemical exchange contributions to $R_2(^{15}\text{N})$ can be observed in the hexapeptide region of the disordered region and in the homeodomain (Figure 3.5-e). Such contributions preclude the proper derivation of $J(0)$ from $R_2(^{15}\text{N})$ rates, in particular at high magnetic fields. To avoid this problem $J(0)$ can be derived from $J(\omega_N)$ by using the longitudinal (η_z) and transverse (η_{xy}) CSA/DD cross-correlated relaxation rates¹³¹:

$$J(0) = \frac{3}{4} \left[2 \frac{\eta_{xy}}{\eta_z} - 1 \right] J(\omega_n) \quad (3.8)$$

Due to the low sensitivity of the longitudinal CSA/DD cross-correlated cross-relaxation experiments, $J(0)$ was estimated only from data at the highest fields (18.8 T and 22.3 T) (Figure 3.6-c).

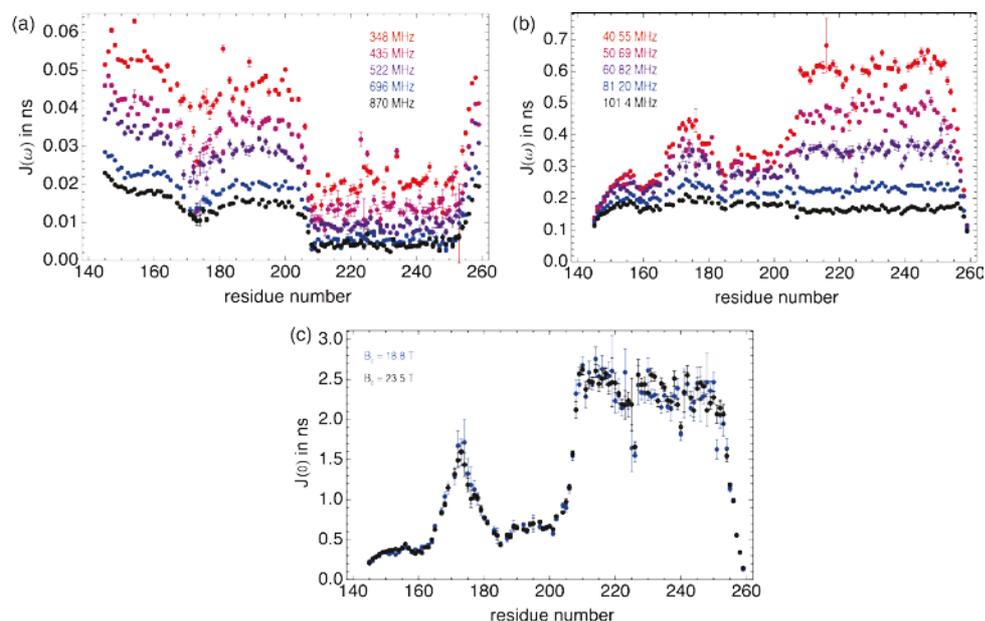


Figure 3.6: Spectral density functions for backbone NH vectors in Engrailed 2. (a) Effective spectral density near the proton Larmor frequency $J(0.87\omega_H)$ (ns). (b) Spectral density at the Larmor frequency of nitrogen-15 $J(\omega_N)$ (ns). (c) Spectral density at zero frequency $J(0)$ (ns). All the data are color-coded as a function of the magnetic field at which the relaxation rates were recorded with the same code as in Figure 3.5.

5. Principles of IMPACT

The limitations of conventional approaches to the analysis of relaxation rates in IDPs and IDRs result from the complexity of their dynamics. Significant motions are expected throughout the tens of ps to tens of ns range. It is likely that they cannot be accurately described by a single distribution of correlation times, or by a small number of correlation times. However, the scarcity of relaxation rates limits the number of adjustable parameters that can be determined, and thus limits the sophistication of spectral density functions that can be postulated. Here, we significantly increase the number of correlation times by defining an *array of n fixed correlation times*. Only the relative coefficient of each correlation time in the distribution is fitted to experimental data, so that the number of adjustable parameters is reduced. Thus, our only assumption is that the correlation function can be approximated by a sum of exponentials. The physical content of the IMPACT model is thus limited to a minimum. IMPACT can be described as a mathematical approach, which converts experimental relaxation rates (or, equivalently spectral density mapping results) into a distribution of correlation times that is more amenable to physical interpretation than the raw experimental data.

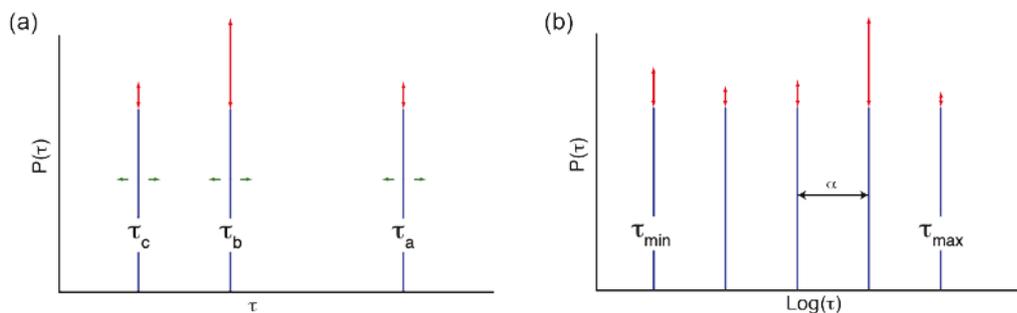


Figure 3.7: Principle of IMPACT analysis. (a) In the three correlation times (3CT) analysis, both the value and the relative weight of each correlation time is adjusted. (b) In IMPACT, only the relative weight is adjusted while the values of correlation times is fixed and equally spaced on the logarithmic scale.

The array of n correlation times is defined as a geometric series, so that correlation times are equally spaced on a logarithmic scale (Figure 3.7):

$$\tau_i = \alpha^{i-1} \tau_{max} \quad \alpha = \left(\frac{\tau_{min}}{\tau_{max}} \right)^{\frac{1}{n-1}} \quad (3.9)$$

Thus $J(\omega)$ is a sum of Lorentzian functions:

$$J(\omega) = \sum_{i=1}^n J_i(\omega) = \frac{2}{5} \sum_{i=1}^n \left(\frac{A_i \tau_i}{1 + (\omega \tau_i)^2} \right) \quad (3.10)$$

where A_i is the coefficient of the correlation time τ_i in the spectral density function. The coefficients A_i must be positive and fulfill the following normalization constraint:

$$\sum_{i=1}^n A_i = 1 \quad (3.11)$$

Thus, the number of adjustable parameters is reduced to $n-1$.

A preliminary step of the IMPACT analysis is the optimization of the three parameters τ_{min} , τ_{max} , and n . The first step is to define the range of correlation times that are probed by relaxation rates. A series of correlation times could be chosen as the inverse of the Larmor frequencies at which the spectral density is mapped in analogy to the study by LeMaster²⁰⁶. Considering that the range of frequencies where a Lorentzian function varies extends beyond the inflection point, we chose a slightly different approach. We first define the range of correlation times that are sampled by various ¹⁵N relaxation rates. We consider that the lowest magnetic field adapted to protein studies is 400 MHz, while the highest accessible field currently is 1 GHz. Thus, the lowest non-zero frequency at which the spectral density is

Chapter 3: Interpretation of motions in disordered proteins

sampled is $\omega_N/2\pi = 40$ MHz, and the highest $0.87\omega_H/2\pi = 870$ MHz. A Lorentzian function with a correlation time $\tau_c = 40$ ns drops to 1% of $J(0)$ at $\omega_N/2\pi = 40$ MHz. A Lorentzian with $\tau_c = 18$ ps merely decreases to 99% of $J(0)$ at $0.87\omega_H/2\pi = 870$ MHz. The resulting range spans slightly more than three orders of magnitude. Therefore, we have decided to limit the range to three orders of magnitude

$$\tau_{max}/\tau_{min} = 10^3 \quad (3.12)$$

In order to define the optimal values of τ_{min} and τ_{max} , we have carried out a series of IMPACT analyses for $[\tau_{min}, \tau_{max}] = [1 \text{ ps}, 1 \text{ ns}]$ to $[100 \text{ ps}, 100 \text{ ns}]$ as well as for $4 < n < 9$. By contrast to the approach of LeMaster²⁰⁶, the number of correlation times is adjustable. The statistical relevance of each combination of parameters was evaluated from the resulting AIC (Akaike's Information Criteria)²⁰⁷⁻²¹⁰:

$$AIC = n_{exp} \ln \left(\frac{1}{n_{exp}} \times \sum_{k=1}^{nres} \chi_k^2 \right) + 2n_{model} + C \quad (3.13)$$

$n_{exp} = n_J * nres$ is the total number of experimental data, with $n_J = 11$ points at which the spectral density function $J(\omega)$ is sampled, when relaxation data at five magnetic fields are used; $nres = 108$ is the number of residues included in the analysis; $n_{model} = (n-1)*nres$ is the number of free parameters in each model. Here, the constant is $C = 0$. AIC are shown in Figure 3.8-a. Two local minima were found for $[\tau_{min}, \tau_{max}] = [34 \text{ ps}, 34 \text{ ns}]$ with $n = 5$ and for $[\tau_{min}, \tau_{max}] = [21 \text{ ps}, 21 \text{ ns}]$ with $n = 6$ (Figure 3.8-b). The likelihood of the latter array of correlation times is 10^3 times higher than that of the former. Here, we will thus present the IMPACT analysis with $[\tau_{min}, \tau_{max}] = [21 \text{ ps}, 21 \text{ ns}]$ and $n = 6$. This result is dominated by the diverse dynamic properties of the IDR of Engrailed. Indeed, if we exclude the rigid residues of the homeodomain, the optimal parameters change slightly (residues 146 to 207) $[\tau_{min}, \tau_{max}] = [42 \text{ ps}, 42 \text{ ns}]$ with $n = 5$; while the optimal set of parameters for the rigid part of the homeodomain alone (residues 208-259) is significantly different $[\tau_{min}, \tau_{max}] = [10 \text{ ps}, 10 \text{ ns}]$ and $n = 4$.

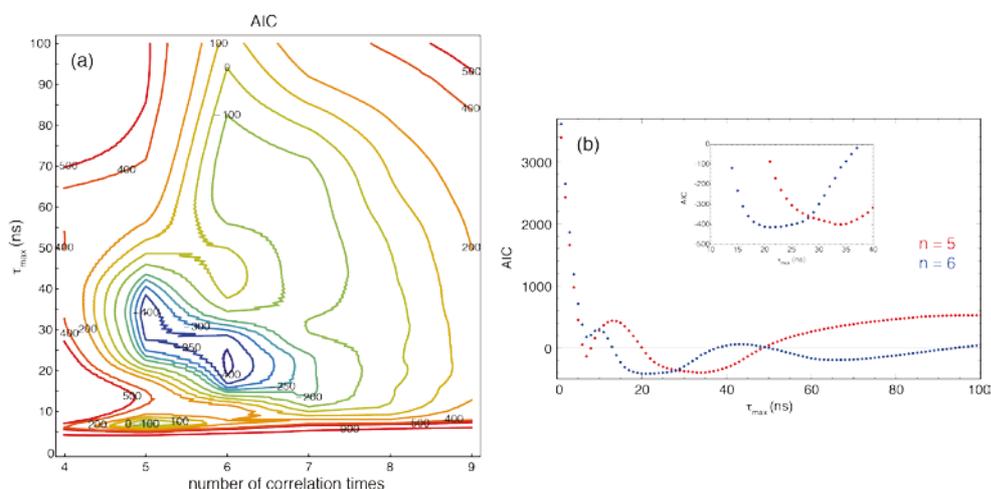


Figure 3.8: (a) Optimization of IMPACT by considering Akaike’s Information Criteria (AIC). The range $[\tau_{min}, \tau_{max}]$ of correlation times characterized by IMPACT was varied from [1 ps, 1 ns] to [100 ps, 100 ns] and the number of correlation times was varied from $n = 4$ to 9. Despite the continuous lines shown in the contour plot, the number of correlation times is an integer. (b) 1D AIC plots for $n = 5$ or $n = 6$.

6. Application of IMPACT to Engrailed

The optimal parameters $n = 6$ and $[\tau_{min}, \tau_{max}] = [21 \text{ ps}, 21 \text{ ns}]$ were employed to analyze the spectral density function in Engrailed. Note that the coefficients A_i were fitted to spectral density mapping results. In principle, relaxation rates could also be used directly as input for the IMPACT analysis. Figure 3.8 illustrates the remarkable variety of dynamics found in Engrailed. In the homeodomain, the second correlation time τ_2 lies just below the correlation time for overall rotational diffusion, which is close to 7 ns, as can be seen from the analysis based on only two correlation times (*vide infra*). Thus, the second coefficient is, by far, the most important in the homeodomain. A small amplitude A_1 of τ_1 corrects for the fact that τ_2 is shorter than the actual correlation time for the motion of the whole domain. Note that the correlation time for overall diffusion τ_m is well approximated by:

$$\tau_m \approx (A_1 \tau_1 + A_2 \tau_2) / (A_1 + A_2) \quad (3.14)$$

The average value over the homeodomain is $\langle \tau_m \rangle = 7.19 \text{ ns}$. Small but significant and mostly uniform contributions A_3 of the correlation time τ_3 in the three alpha helices, which are also obtained in a conventional “model-free” analysis *vide infra*, may be attributed to fluctuations of the overall diffusion tensor¹⁸², likely due to conformational fluctuations of the IDR (residues 146-199). Enhanced values of A_3 in the loops may reflect the flexibility of these regions³⁸⁻⁴⁰. The very small coefficients A_4 and A_5 demonstrate the presence of a gap in the distribution of correlation times, as was also observed in ubiquitin²¹¹. Finally, the

Chapter 3: Interpretation of motions in disordered proteins

coefficients A_6 for the shortest correlation time τ_6 indicate the presence of fast motions in the tens of ps range. Note that the Lorentzian function $J_6(\omega)$ drops by about 1% of $J_6(0)$ at the highest frequency explored in this analysis (i.e. $\omega/2\pi = 870$ MHz). Thus, this last term in the spectral density function can be approximated to a constant that effectively represents all fast motions:

$$J_6(\omega) \approx \frac{2}{5}A_6\tau_6 \approx \frac{2}{5}\int_0^{\tau_5} p(\tau)d\tau \quad (3.15)$$

where $p(\tau)$ is the probability function of correlation times, containing little information on the complexity of such motions²¹².

Results obtained in the disordered region of Engrailed will be discussed with the help of figure 3.9 but also with the ‘IMPACT barcode’ shown in figure 3.10. In the latter, for each residue, the width of each histogram represents the coefficient A_i associated with the correlation time τ_i that can be read on the y-axis. This graph appears to be a convenient way to display the results of the IMPACT analysis in a single figure.

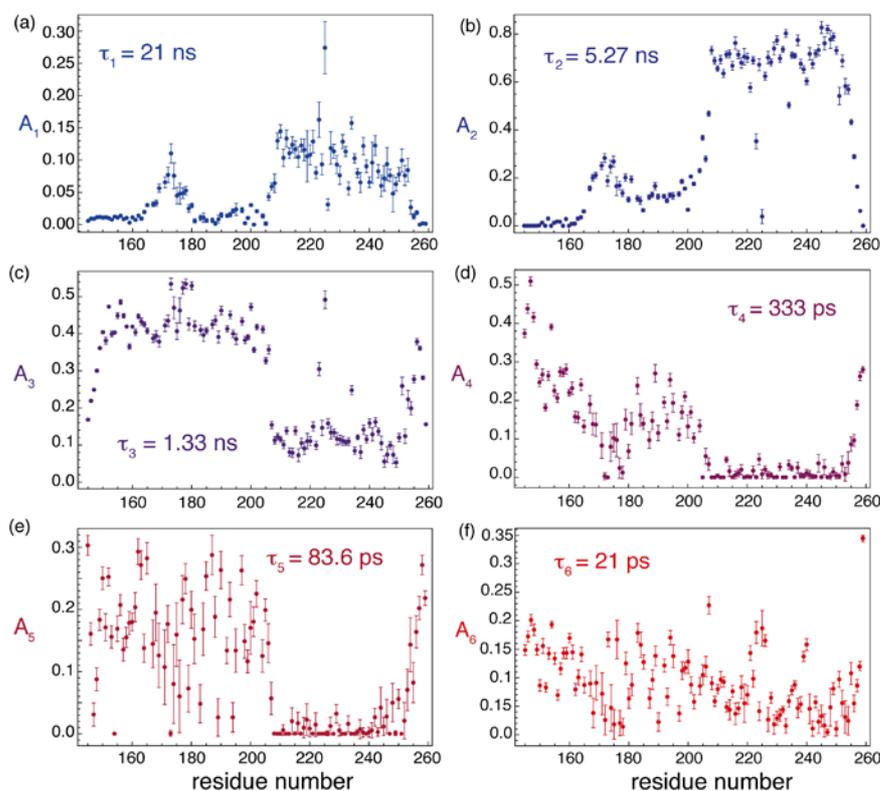


Figure 3.9: (a-f) Plot of the 6 coefficients A_i ($i = 1, 2, \dots, 6$) of the $n = 6$ correlation times τ_i in the range $[\tau_{min}, \tau_{max}] = [21 \text{ ps}, 21 \text{ ns}]$ determined by the IMPACT analysis of Engrailed.

For the first residues at the N-terminus and the last residues at the C-terminus, significant coefficients A_{3-6} are found for the four shortest correlation times. This seems to indicate the presence of motions that are broadly distributed over all time scales up to 1 ns. On the other hand, the two disordered regions just at the N-terminus and the C-terminus of the hexapeptide display a high density of motions around τ_3 . The coefficients for the correlation time τ_4 decrease almost linearly with the distance to the N- or C-termini of the polypeptide chain in disordered regions and reach different plateaus in each disordered segment. A notable difference between the disordered region at the N-terminus and the one in between the hexapeptide and the homeodomain is the slight decrease of the coefficient A_3 and a significant increase of A_2 . It is difficult to assign this change to a particular process without a better characterization of the conformational space of the protein. Nevertheless two effects may contribute to the presence of some orientational order beyond 1 ns: first, this IDR is short and located between a folded domain and a small hydrophobic cluster, thus its dynamics is likely influenced by the overall diffusion of both structured elements; second, this IDR contains a

Chapter 3: Interpretation of motions in disordered proteins

majority of residues that favor extended conformations, as confirmed by SSP scores: 3 proline residues, 9 positively charged, and only 1 negatively charged residues between positions 177 and 198; which should restrict the conformational space and possibly slow down reorientational dynamics.

The barcode representation of IMPACT coefficients nicely illustrates variations of the ensemble of correlation times between successive structural elements. Thus for instance, the decrease of motions in the sub-ns range in the hexapeptide is accompanied by an increase in the supra-ns range. Similarly, variations of the coefficients for correlation times at the N- and C-termini of the homeodomain (residues 200-210 and 254-260) illustrate the smooth transition of motional properties along the sequence of the polypeptide. Finally, even in the homeodomain, where a classical analysis of relaxation data should be most appropriate, the dynamic transitions between helices and the two loops are clearly visible and quantitatively characterized by the IMPACT approach. Loop $\alpha 1$ - $\alpha 2$ features enhanced dynamics in both the 1 ns and tens of ps ranges, as expected from the motions demonstrated by paramagnetic relaxation enhancement studies¹⁹⁷, while loop $\alpha 2$ - $\alpha 3$ shows a significant but more moderate enhancement of motions. Overall, the IMPACT representation offers an elegant view of the correlation of structural and dynamic features, as can be seen from the SSP scores (Figure 3.10-b).

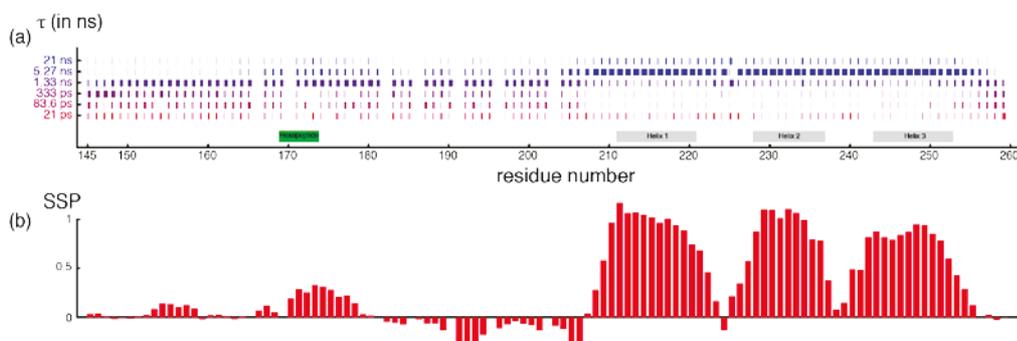


Figure 3.10: (a) Graphical representations of IMPACT analyses of the spectral density function in Engrailed 2. Histogram is drawn for all residues with the following rules: for each of the six correlation times τ_i ($i = 1, 2, \dots, 6$) considered in IMPACT, a rectangle represents the dynamics of the amino-acid residue. The width of each rectangle is proportional to the corresponding weights A_i in IMPACT. Grey rectangles indicate rigid alpha helices, and a green rectangle shows the rigid hydrophobic hexapeptide sequence. (b) The secondary structure propensity (SSP) (as in figure 3.4) is shown to guide the comparison between structural and dynamic features.

7. Comparison with 2CT and 3CT analyses

For the sake of comparison, we also fit two simple models where the spectral density function is assumed to consist of a sum of two and three Lorentzians in the manner of the familiar ‘model-free’ and ‘extended model-free’ approaches. However, as discussed in the introduction, the core hypotheses of the model-free formalism cannot be fulfilled *a priori*, since the longest correlation time is probably an effective correlation time rather than the correlation time of overall rotational diffusion. The spectral density J_{2CT} assuming two correlation times (2CT) can be written:

$$J_{2CT}(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_a}{1 + (\omega \tau_a)^2} + \frac{(1 - S^2) \tau'_b}{1 + (\omega \tau'_b)^2} \right] \quad (3.16)$$

with $\tau'^{-1}_b = \tau_a^{-1} + \tau_b^{-1}$, τ_a is the long correlation time and τ_b is the short effective correlation time, while S^2 is similar to the model-free order parameter. The spectral density function J_{3CT} assuming three correlation times (3CT) can be defined as follows:

$$J_{3CT}(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_a}{1 + (\omega \tau_a)^2} + \frac{(S_f^2 - S^2) \tau'_b}{1 + (\omega \tau'_b)^2} + \frac{(1 - S_f^2) \tau'_c}{1 + (\omega \tau'_c)^2} \right] \quad (3.17)$$

with $\tau'^{-1}_c = \tau_a^{-1} + \tau_c^{-1}$; $\tau_a > \tau_b > \tau_c$; S_f^2 is equivalent to the extended model-free order parameter for fast processes with a correlation time τ_c . The two functions were fitted to the experimental spectral density, and a simple model selection was based on the comparison of AICc (Equation 3.18-3.19), with $n_J = 11$ and $n_{model} = 3$ for the 2CT analysis, and $n_{model} = 5$ for the 3CT analysis.

$$AICc = AIC + \frac{2n_{model}(n_{model} + 1)}{n_{exp} - n_{model} - 1} \quad (3.18)$$

with

$$AIC = n_{exp} \ln \left(\frac{\chi^2}{n_{exp}} \right) + 2n_{model} + C \quad (3.19)$$

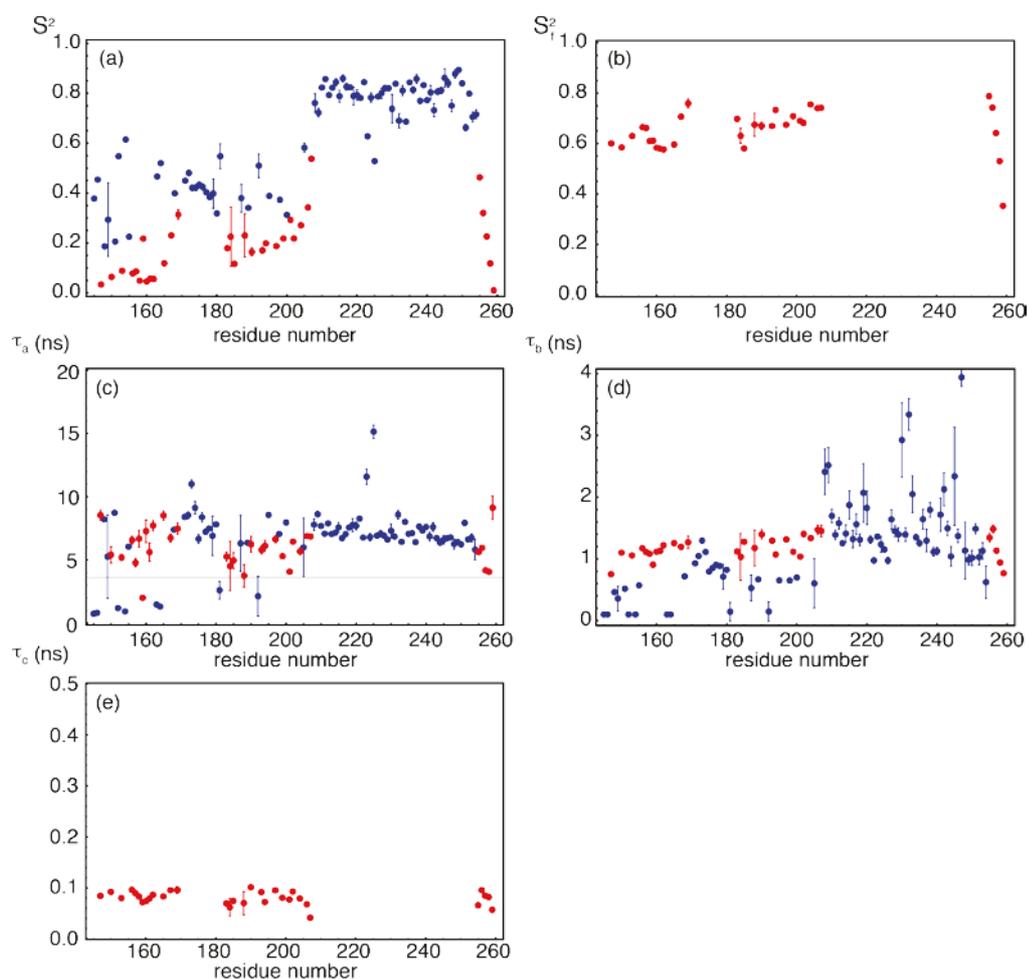


Figure 3.11: Results obtained for a conventional analysis with two correlation times (2CT, blue) or three correlation times (3CT, red). (a) Order parameter S^2 ; (b) order parameter S_f^2 for the fastest motion in the 3CT analysis; (c) longest correlation time τ_a ; (d) intermediate correlation time τ_b ; and (e) shortest correlation time τ_c . Either 2CT or 3CT model was selected based on the lowest AIC \square

Results of this analysis are shown in Figure 3.11. From a statistical point of view, the 2CT analysis is found sufficient to describe the motions in the mostly rigid homeodomain (residues 200-259), except at the flexible N- and C-termini. With few exceptions, the longest correlation time yields a reliable measure of overall rotational diffusion, the average value over the helices of the homeodomain is $\langle \tau_a \rangle_{\text{hom}} = 7.08$ ns, in good agreement with the IMPACT analysis, with $\langle \tau_m \rangle = 7.19$ ns. Interestingly, the second correlation time τ_b found for most residues in the rigid homeodomain lies in the range $0.9 < \tau_b < 1.8$ ns. This is in agreement with the IMPACT analysis and is possibly due to fluctuations of the overall diffusion tensor resulting from conformational transitions in the disordered N-terminal region

on timescales between 1 and 100 ns^{182,211}. In the disordered region (residues 146-199), the 2CT and 3CT analyses seem equally probable with no particular pattern along the sequence, except in the hydrophobic hexapeptide cluster (residues 169-174) where the 2CT analysis is more satisfactory. The random patterns of 2CT vs. 3CT selection seems to point to some instability of the model-selection step in the fit procedure. A 2CT or 3CT analysis can be performed with no model selection (Figure 3.12). The built-in absence of site-specific model selection in IMPACT shields this analysis from such a drawback. Low order parameters S^2 are found throughout the IDR, with a significant increase in order in the hydrophobic cluster. The long correlation times in the disordered regions have a broad distribution (standard deviation of 2.5 ns) but the average value $\langle \tau_a \rangle_{\text{IDR}} = 5.9$ ns is similar to what was found in unfolded²¹³ and intrinsically disordered proteins²¹⁴, and very close to the correlation time for overall tumbling of the homeodomain. The intermediate correlation time, which corresponds to the dominant term in the spectral density function, lies in the range $0.1 < \tau_b < 1.6$ ns, in agreement with the IMPACT analysis. The shortest correlation time lies in the range $40 < \tau_c < 120$ ps.

One should be careful with the physical interpretation of these observations in the IDR of Engrailed. The three correlation times obtained are clearly separated in time domain, which indicates a broad range of dynamic processes. The results should not be considered *a priori* as actual correlation times of particular motions, but rather as the best rendition of experimental results using two or three correlation times. This is illustrated by the jumps of order parameters and correlation times observed between the 2CT and 3CT models in the IDR, which illustrates the effective character of the fitted correlation times in this region, at least in the 2CT analysis. For instance, it is difficult to assign the longest correlation time to any particular dynamical process in the absence of complementary experimental or computational information. Such a process could be a single well-defined type of motion, such as the rotational diffusion of an IDR segment. Alternatively, the longest correlation time might account for the tail of a continuous distribution of correlation times and reflect slower motions in parts of the conformational space of the IDR. Interestingly, the correlation times obtained in a 3CT analysis often correspond to reciprocal frequencies ($\omega = 1/\tau$) that lie outside regions where the spectral density function can be adequately sampled (*i.e.*, below 40 MHz, between 100 MHz and 348 MHz and above 870 MHz). This is particularly true in the flexible region between the rigid hexapeptide and the rigid homeodomain and at the C-terminus of the protein. The regions where the spectral density function is most sensitive to

Chapter 3: Interpretation of motions in disordered proteins

the choice of correlation times correspond to ranges where we lack experimental constraints. This would be expected in the presence of a broad distribution of correlation times that would lead to a smooth decay of the spectral density function.

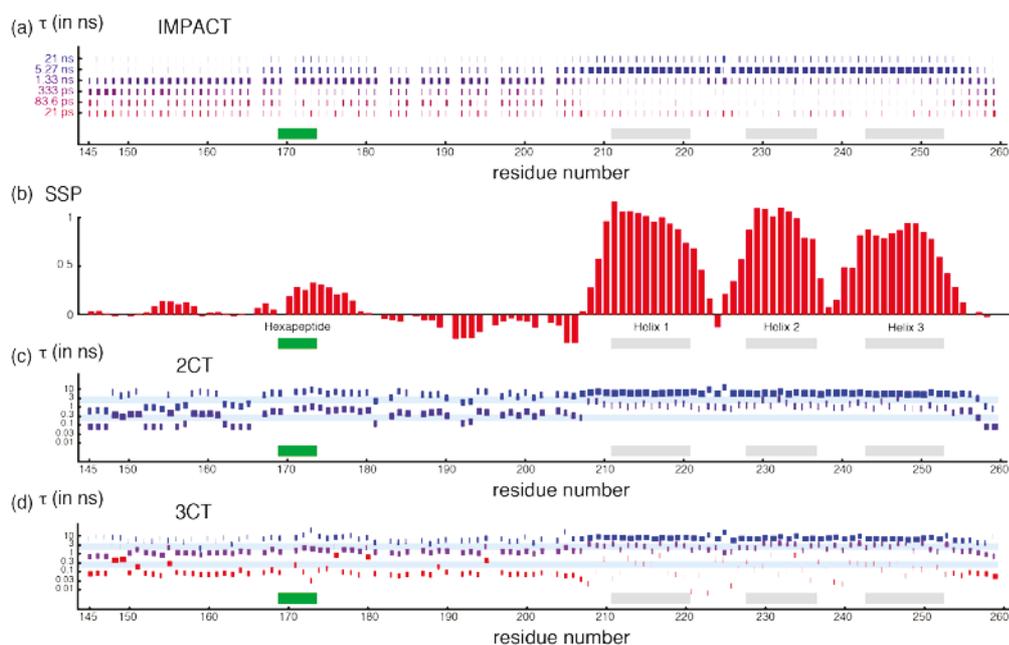


Figure 3.12: Graphical representations of (a) IMPACT, (c) 2CT, and (d) 3CT analyses of the spectral density function in Engrailed 2. Histograms are drawn for all residues, which represents the contributions of: (a) each of the six correlation times τ_i ($i = 1, 2, \dots, 6$) considered in IMPACT; (c) each of the two correlation times $\tau_{a,b}$ determined by the 2CT analysis; (d) each of the three correlation times $\tau_{a,b,c}$ determined by the 3CT analysis. The width of each rectangle is proportional to the corresponding weights (a) A_i in IMPACT; (c) $B_{a,b}$ in 2CT; (d) $B_{a,b,c}$ in 3CT. In (c,d), light blue horizontal bars represent the ranges of correlation times τ for which reciprocal frequencies lie in the constrained regions between $40 < 1/(2\pi\tau) < 100$ MHz or between $348 < 1/(2\pi\tau) < 870$ MHz. Grey rectangles in (a), (c) and (d) indicate rigid alpha helices, and a green rectangle shows the rigid hydrophobic hexapeptide sequence. (b) The secondary structure propensity (SSP) (as in Figure 3.4) is shown to guide the comparison between structural and dynamic features.

A direct comparison between the results of the 2/3CT analysis and our IMPACT approach is illustrated in Figure 3.12. For the sake of comparison, we define in Figure 3.12-c-d the coefficients $B_{a,b,c}$ associated with the correlation times $\tau_{a,b,c}$ as:

$$A = S^2, B_b = S_f^2 - S^2, B_c = 1 - S_f^2 \quad (3.20)$$

The statistical significance of the fit resulting from our IMPACT analysis is often better than with either the 2CT or the 3CT analysis in the disordered regions, while it is somehow comparable to that of the 2CT model in the rigid homeodomain (Data not shown).

Cyril Charlier

In particular, the almost complete absence of abnormally elevated values of the χ^2 in the IDR of Engrailed shows that a faithful set of fitted A_i parameters can be obtained with diverse dynamical features. Interestingly, as can be seen from the schematic representation of correlation times that correspond to reciprocal frequencies that can be determined by spectral density mapping, the inflexion points of most of the Lorentzian functions often lie in frequency regions where spectral density mapping does not yield any results. This is particularly true in the IDR between the hexapeptide and the homeodomain and at the C-terminus of the protein. This seems to indicate that the decay of the spectral density function is smoother than can be described by a sum of three Lorentzian functions. The fit pushes the inflexion points of individual contributions to the spectral density function beyond the areas that benefit from rich experimental constraints. In addition, the absence of a residue-specific model selection in IMPACT provides results that are directly comparable, residue by residue, which allows for a better qualitative description of dynamic properties along the protein sequence. Admittedly, model selection can be omitted in the 2CT or 3CT analysis, as in Figure 3.12.

A potential concern of the IMPACT analysis lies in the fact that the correlation functions $J_i(\omega)$ are not independent since they suffer from significant overlap. Hence, it is possible that different ensembles of coefficients A_i can describe the same experimental data. In order to test the sensitivity of our analysis to this potential flaw, we have plotted correlations of consecutive coefficients A_i (i.e. A_i as a function of A_{i+1}) for all 510 steps of the Monte Carlo procedure employed in the fit. Typical results are shown in Figure 3.13. There is a small anti-correlation between consecutive coefficients in several instances. This will give rise to a broader distribution of individual coefficients and thus lead to a decrease of the precision of these coefficients. In the worst case, a potential decrease of accuracy due to the interdependence of consecutive coefficients will be accompanied with a decrease in precision.

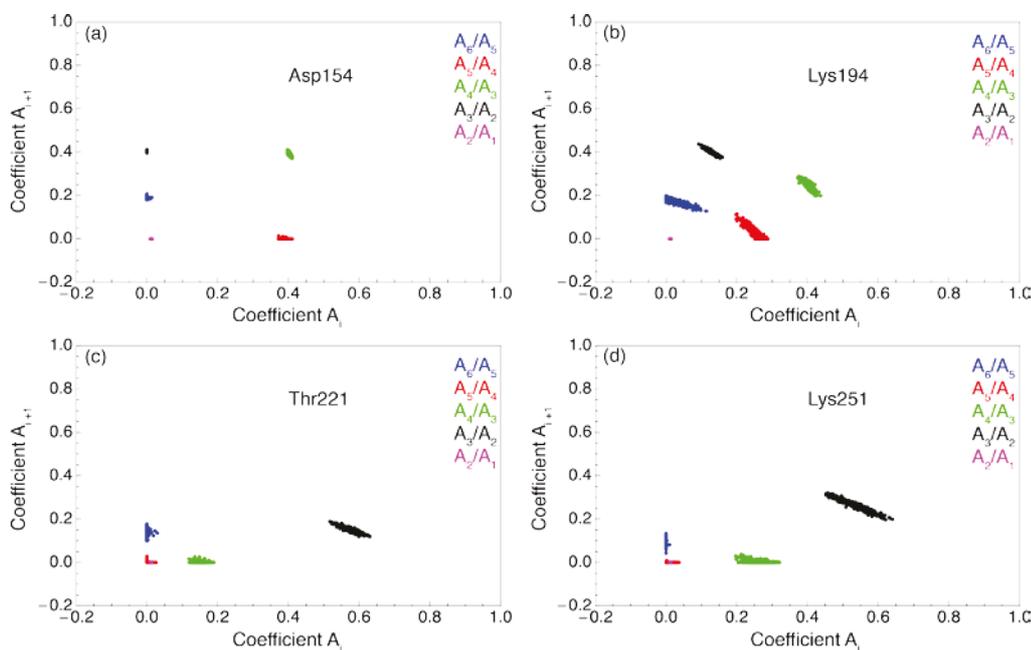


Figure 3.13: Correlation of consecutive IMPACT coefficients. A_i coefficients are displayed on the y-axis as a function of A_{i+1} coefficients displayed on the x-axis: A_5 as a function of A_6 (blue); A_4 as a function of A_5 (Red); A_3 as a function of A_4 (green); A_2 as a function of A_3 (black); A_1 as a function of A_2 (magenta). Typical results are shown for residues in different regions of the protein: (a) Asp154; (b) Lys194; (c) Thr221; (d) Lys251.

A potential concern of the IMPACT analysis is the risk of over-interpretation of the results. Here, we should be clear and provide a set of rules, which should be followed by users of this approach:

- (i) The correlation times τ_i are not physical correlation times of the system *a priori*. The range of correlation times is defined by the experimental observables but the individual values τ_i are derived from a statistical analysis, not a physical analysis.
- (ii) A non-zero coefficient A_1 , with τ_1 the longest correlation time, does not mean that some motions with a correlation time τ_1 were detected but that the distribution of correlation times is larger than zero for some correlation times larger than τ_2 .
- (iii) Similarly, as mentioned in Equation (3.15), the coefficient A_n of the shortest correlation time τ_n , is an effective representation of all fastest motions.
- (iv) If the coefficient A_i is larger than zero, the distribution of correlation times is larger than zero somewhere between τ_{i+1} and τ_{i-1} .

Cyril Charlier

- (v) If the coefficients A_i and A_{i+1} are both zero, the distribution of correlation times is expected to be zero at least between τ_{i+1} and τ_i .
- (vi) Finally, very few relaxation studies have compared rates at five or more magnetic fields^{83,215}. We have tested if an IMPACT analysis of relaxation rates recorded at only three fields could give meaningful results, using various combination of fields. In either case this requires about two weeks of experimental time. The results of the analysis of relaxation rates, shown in Figures 3.14 and 3.15 are remarkably similar to those presented in Figures 3.9 and 4.10. When the range of magnetic fields is restricted, with relaxation rates measured at 11.7, 14.1 and 18.8 T, some significant changes of IMPACT coefficients can be observed, but the overall description of the distribution of correlation times is very similar to what is obtained with relaxation rates at five magnetic fields. Hence, IMPACT can be applied to many proteins at a moderate cost in experimental time, and does not necessarily require exceptionally large data sets or magnetic fields as high as 23.5 T.

Chapter 3: Interpretation of motions in disordered proteins

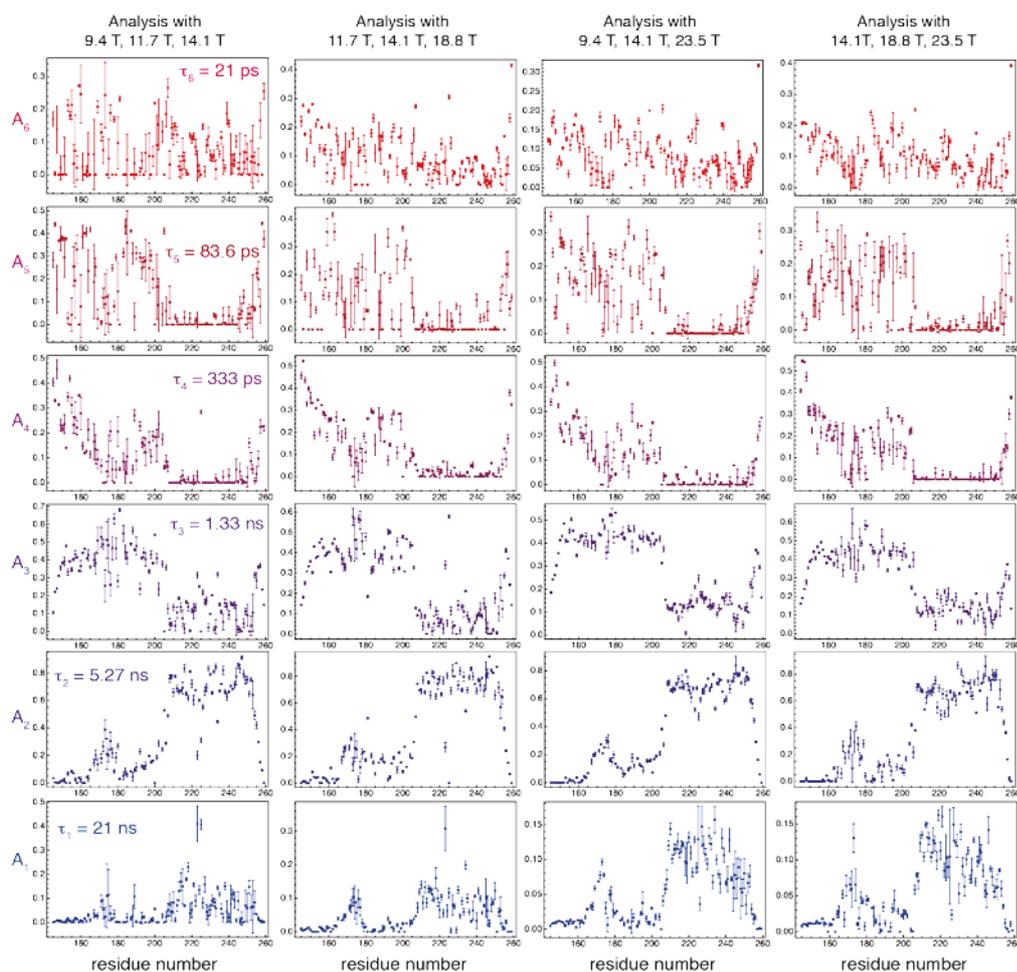


Figure 3.14: IMPACT results with various combinations of relaxation rates measured at three magnetic fields. As for the results presented in Figure 4.9, the number of correlation times was $n = 6$ and the range of correlation times was [21 ps, 21 ns]: A₁ with $\tau_1 = 21$ ns; A₂ with $\tau_2 = 5.27$ ns; A₃ with $\tau_3 = 1.33$ ns; A₄ with $\tau_4 = 333$ ps; A₅ with $\tau_5 = 83.6$ ps; A₆ with $\tau_6 = 21$ ps. (Readers should notice that the scale on the y-axis is not identical for all the plots)

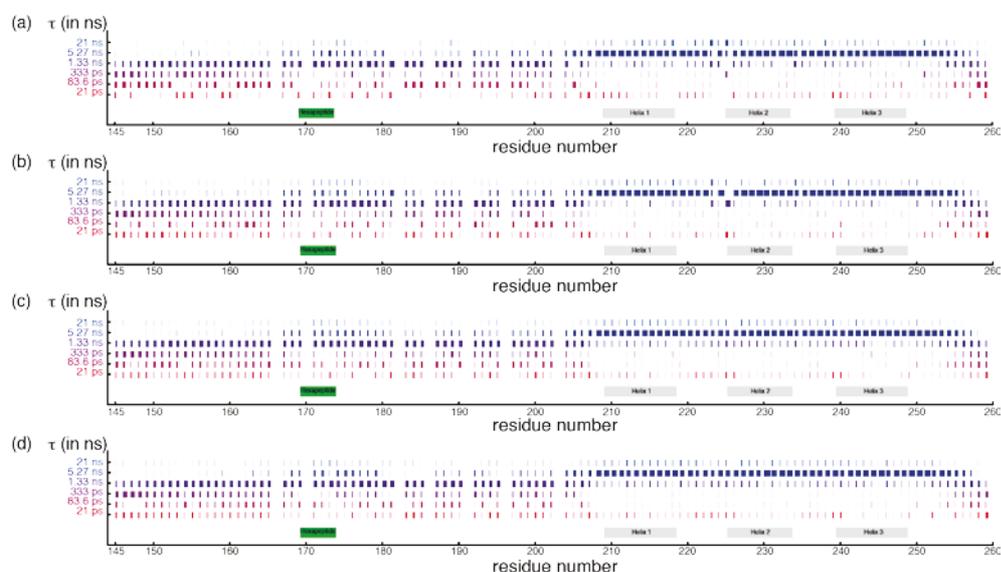


Figure 3.15: Bar-code representation of the IMPACT analysis of the spectral density function in Engrailed 2 with (a) 9.4 T, 11.7 T and 14.1 T; (b) 11.7 T, 14.1 T and 18.8 T; (c) 9.4 T, 14.1 T and 23.5 T; (d) 14.1 T, 18.8 T and 23.5 T. Histograms are drawn for all residues with the following rules: for each correlation time obtained or used in the analysis of the spectral density function τ_i , a rectangle is represented at the corresponding position along the y-axis, with a logarithmic scale; the width of each rectangle is proportional to the corresponding weight, A_i . The main structural features are illustrated by grey rectangles for alpha helices and a green rectangle for the location of the hydrophobic hexapeptide.

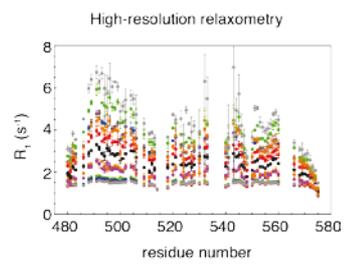
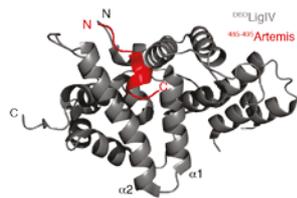
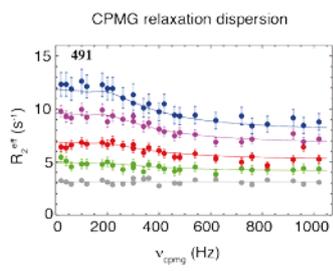
8. Conclusion

We have presented a large set of nitrogen-15 relaxation rates in the 115-residue partially disordered protein Engrailed 2 recorded at five different magnetic fields. The transverse cross-correlated rate η_{xy} was shown to be the most sensitive to the extent of order and disorder at all magnetic fields. The analysis validates the reduced spectral density mapping approach originally developed for folded proteins and already extensively applied to IDPs and IDRs. The spectral density functions can be fitted reasonably well with two or three correlation times although such results may be difficult to interpret. We have introduced a novel approach to the analysis of spectral density functions called ‘interpretation of motions with a projection onto an array of correlation times’ (IMPACT). This provides a better quantitative description of spectral density functions in IDRs as found in Engrailed than an analysis with three correlation times with the same number of adjustable parameters. We also introduce a “barcode” representation of IMPACT, which provides a condensed graphical representation of large amounts of data in a single figure. This representation lends itself to a qualitative discussion of order and disorder in proteins. IMPACT can also be useful to analyze

Chapter 3: Interpretation of motions in disordered proteins

a smaller set of relaxation rates recorded at only three magnetic fields. IMPACT provides a unique framework for the description of the timescales of motions in IDPs and IDRs. Our approach is complementary to the determination of conformational ensembles^{201,216}. Insight into the dynamics of IDPs and IDRs should greatly benefit from a combined analysis.

Chapter 4: Interaction and dynamics of the C- terminal region of Artemis



1. **Biological context**

1.1 *V(D)J recombination*

The immune system is a combination of cells, tissues and organs protecting bodies from infection. This system enables to recognize and destroy foreign invaders. The immune cells, T (thymus-derived) cells and B (bursa-derived) cells, all together called lymphocytes, exhibit a large variety of antigen receptors and defend the body against invaders. The adaptive immune system, a subsystem of the overall immune system, is in charge of responding to an unlimited number of pathogens. Such adaptability to any potential hazardous pathogen is made possible by the rearrangement of preexisting gene segments which leads to the formation of a vast number of different antigen-recognition regions at the surface of the T- and B-cells receptors (called TCR and BCR respectively). These receptors are composed of two distinct protein segments: α & β chains for TCR and heavy & light chains for BCR such as immunoglobulin (Figure 4.1). The overall structures of TCR and BCR are formed by heterodimers connected by disulfide bonds. Both heavy and light (or α β) chains include one variable segment and one constant segment. The variable region is responsible for the high diversity of TCR and BCR. The variable region of α chain of TCR and heavy chain of BCR is composed by V (Variable), D (Diversity) and J (Joining) segments whereas the β chain of TCR and the light chain of BCR consist of only V and J segments²¹⁷.

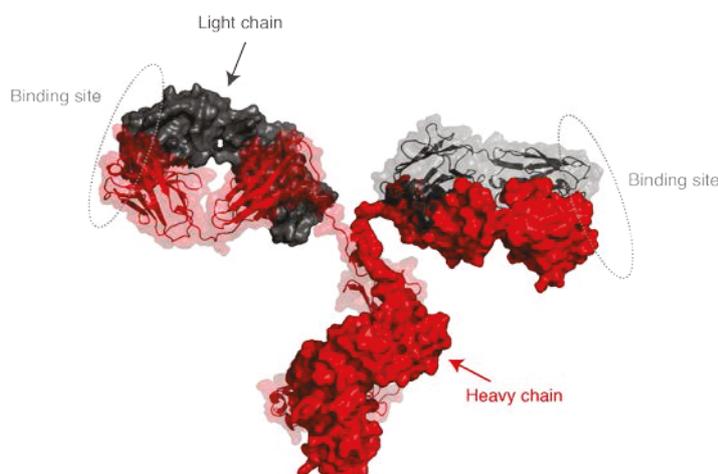


Figure 4.1: Structure of immunoglobulin, an example of BCR. The full structure is a heterodimer with one heavy chain (red) and one light chain (gray). Two binding sites are available through heterodimer formation and are represented by gray rings (pdb code 1IGT)²¹⁸. The structure is identical for TCR.

The adaptability of TCR and BCR receptors is coming from the combination of V, D and J segments in order to recognize the specific antigens (Figure 4.2-a). This mechanism, called V(D)J recombination, is a two-step process that could be defined as a DNA “cut-and-paste” (Figure 4.2-c). The initial step is mediated by the RAG protein complex encoded by the recombination activating gene 1 and 2 (*RAG1 – RAG2*). RAG binds the recombination signal sequences (RSSs) that flank each V, D and J gene segment on DNA (Figure 4.2-b)²¹⁹. RSS sequences are composed of conserved heptamers and nonamers separated by variable sequences of either 12 or 23 base pairs. Upon binding to RSSs, the RAG complex introduces a nick at the heptamer border and cleaves the DNA between the first base of RSS and the coding segment, resulting in the generation of a double-strand break at the signal end and a hairpin at the coding end (Figure 4.2-c). The second step consists in the processing and the repair of DNA ends thanks to the NHEJ (*Non Homologous End Joining*) repair pathway (described in the next section). Through this pathway, the hairpin coding ends are joined imprecisely to form a coding joint while signal ends are joined in a precise manner to form signal joints. The imprecise joining of coding ends generates additional antigen receptor diversity in T and B lymphocytes.

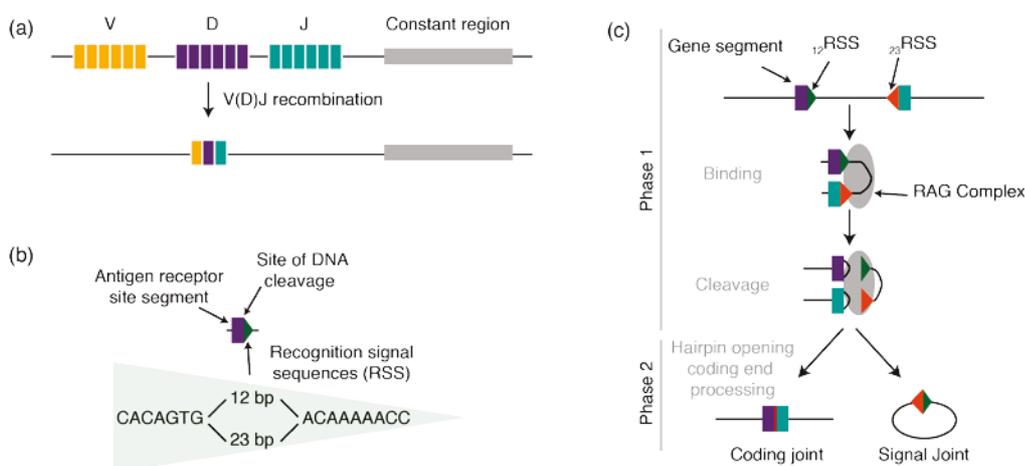


Figure 4.2: (a) General Principle of V(D)J recombination. (b) Description of one Recognition Signal Sequences (RSS). (c) Basic description of the V(D)J recombination two-step process^{217,220}.

1.2 NHEJ DNA repair system

A proper DNA double-strand break (DSB) repair is critical for the stability of the genome. DSBs occur frequently in cells and are caused by both normal metabolic activities and environmental factors. In the first case, DSBs are produced when replication forks encounter blocking lesions or during genome rearrangements, as in the V(D)J recombination

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

process. Typically, the second class of DSBs are generated by DNA damaging agents such as ionizing radiation (IR), chemical agents or UV light²²¹. Cells have developed a complex signaling network in order to sense and adopt a proper response to the damage. The DNA damage response uses different DNA damage checkpoints depending on the cell cycle stage. Cells heavily damaged or seriously deregulated are eliminated by apoptosis²²². The failure to repair DSBs leading to DNA damages may result in mutation, cancer, and cell or organism death.

Two mechanisms to repair DSBs have been developed in eukaryotic cells: NHEJ (Non-Homologous End Joining) and HR (Homologous Recombination). The choice between the HR and NHEJ repair pathways is usually given by the stage of the cell cycle. The cell cycle, the mechanism by which a cell leads to two daughter cells, is divided in 4 phases: G1, S, G2 and M phases. HR is a direct repair mechanism which requires proximity and availability between the two sister chromatids²²¹. The mechanism of HR, not detailed here, uses the sister chromatid as template to repair the broken DNA in an accurate way, and is restricted to S and G2 phases of the cell cycle²²³. In contrast, NHEJ repair pathway occurs throughout the cell cycle. NHEJ is a multi-step mechanism in which the essential components have been identified. First, the Ku70/Ku80 heterodimer binds the free ends of DNA, followed by the recruitment of DNA-PKcs (the catalytic subunit of the DNA-dependent protein kinase DNA-PK) and Artemis. The binding of DNA-PKcs to DNA ends activates its kinase activity leading to both autophosphorylation and phosphorylation of Artemis (Figure 4.3).

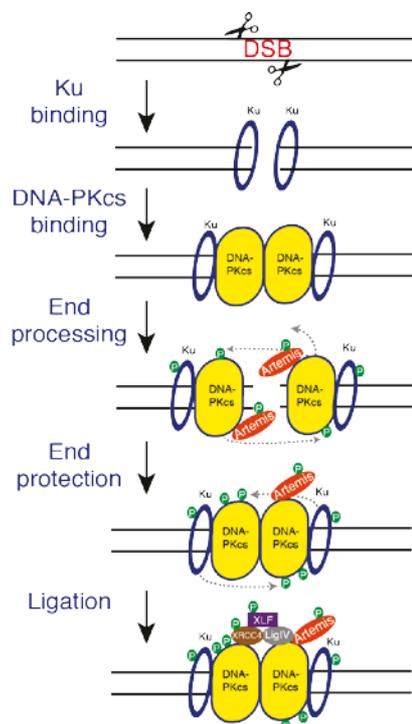


Figure 4.3: Description of the canonical Non Homologous End Joining (NHEJ) pathway. Following DSB, the recruitment of Ku promotes NHEJ. Sequential phosphorylation events on multiple DNA-PKcs amino acids clusters favors the initial step processing of DNA ends by Artemis, followed by DNA-PKcs-dependent protection of DNA ends required for DNA ligation²²⁴.

Autophosphorylation of DNA-PKcs induces a conformational change that allows the processing enzymes to access the ends of the double-strand break²²⁵. In particular, we will see in the following part that the phosphorylation of Artemis by DNA-PKcs turns out to be critical to acquire its endonuclease activity. The Artemis/DNA-PKcs complex is able to open and nick eventual hairpins formed on DNA ends. The removal of damaged nucleotides may require resynthesis by DNA polymerases (Pol μ and Pol λ) unless the DNA ends are already compatible. The last step of NHEJ consists in the ligation of the two ends thanks to a ligation complex formed by DNA Ligase IV, XRCC4 (X-ray cross-complementing protein 4) and XLF (XRCC4-like factor)^{225,226}. In the entire repair pathway, Artemis was the last factor to be identified.

1.3 Artemis

1.3.1 Discovery and identification

Severe Combined Immune Deficiency (SCID) is a severe disease leading to death within the first year of life, unless threatened with transplantation. SCID is characterized by the

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

absence or dysfunction of T-cells. In addition, the absence or presence at abnormal high levels of B-cells or Natural Killer (NK) have been shown to play a role in SCID (Natural Killer cells provide rapid response to infections and start to operate ~3 days after infection)^{227,228}. Clinical studies have revealed that about one third of T^BNK⁺ SCID patients (absence of T-cells, absence of B-cells, and abnormal high presence of NK-cells) carry mutations in the *RAG* gene. Some of the T^BNK⁺ SCID patients without *RAG* gene mutations are sensitive to ionizing radiations with a defect in DSBs repair pathway leading to *RadioSensitive-Severe Combined Immune Deficiency* (RS-SCID). Interestingly, no specific mutation responsible for the RS-SCID phenotype was found in any of the NHEJ recombination factors identified in 1998 (Ku70, Ku80, DNA-PKcs, XRCC4, or Ligase IV)²²⁹. Further genetic and clinical studies have located a new gene on the short arm of human chromosome 10, named *Artemis*, involved in DNA repair pathway as well as V(D)J recombination²³⁰. The identification and cloning of this gene was performed a year later²²⁸. Using genomic comparison, it was found that the *Artemis* gene is composed of 14 exons with sizes from 52bp to 1160bp, leading to a putative protein of 692 amino acids. The *Artemis* gene, ubiquitously expressed, was found mutated in several RS-SCID patients^{227,228}. In addition, overexpression of wild-type Artemis into fibroblasts of RS-SCID patients by transient transfection restored completely the V(D)J recombination process, highlighting the direct implication of Artemis in DNA repair pathway and V(D)J recombination.

Database searches using the protein sequence of Artemis further revealed significant similarities with several well-established members of the metallo- β lactamase superfamily²²⁸. Enzymes within this family are able to use water as a nucleophile to break the amide bond in β -lactame rings (catalyzed hydrolysis)²³¹. Overall, three distinct regions have been defined within Artemis: the β -lactamase homology domain (residues 1-164), the β -CASP domain (residues 165-385), and the COOH-terminal region (residues 386-692) (Figure 4.5-a). The first two domains carry the catalytic activity responsible for the nuclease function of Artemis in V(D)J recombination²³².

1.3.2 The β -lactamase and β -CASP domains

Artemis belongs to the large superfamily of proteins that share a metallo β -lactamase domain as their catalytic domain (Figure 4.4)²²⁸. Among this superfamily, a subset of enzymes use nucleic acids as substrate and are involved in RNA processing (CPSF-73) and DNA repair (SNM1, PSO2)²³¹. The metallo- β -lactamase fold consists of a four layered β -sandwich with two mixed β -sheets flanked by α -helices, the metal-binding sites being located

at one edge of the β -sandwich^{231,233}. The dinuclear Zn(II) center, used to perform the cleavage reaction, is located at the bottom of a wide shallow groove. The entire superfamily is characterized by the presence of a highly conserved signature HxHxDH in which histidines coordinate the zinc ion. Although the high-resolution structures of the β -lactamase and β -CASP domains of Artemis were not elucidated yet, sequence comparisons showed high similarities with proteins of this family specially acting on nuclei acids²³¹.

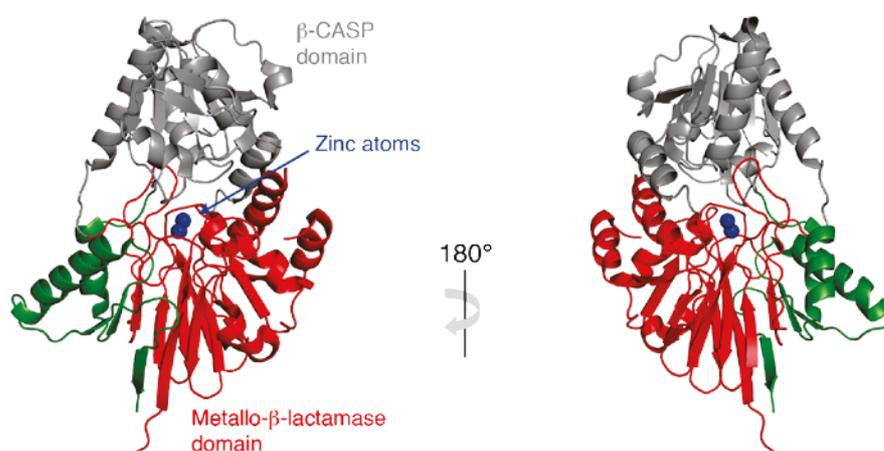


Figure 4.4: Structure of human CPSF-73.

1.3.3 Role of Artemis

The first prediction regarding the function of Artemis suggested the use of its hypothetical hydrolase activity to participate in opening the hairpin at the coding ends during the V(D)J recombination process. In particular, a potential relation between Artemis and DNA-PKcs factor was proposed from the observation that human cells deficient for Artemis have the same phenotype as murine DNA-PKcs mutants, including a 1000-fold decrease of coding joint formation while the level of signal joint formation remained unaffected^{228,234}. In addition, an absence of coding joint formation associated with an increased sensitivity to ionizing radiations (characteristic of DNA-PKcs-deficient phenotype) was also reported with Artemis knockout in mice²³⁵. The highly similar phenotype observed with cells defective for Artemis or DNA-PKcs introduced the idea that these two proteins may be part of a complex. The first experimental evidence for a stable complex between Artemis and DNA-PKcs, *in vitro* and *in vivo*, was published in 2002 by the group of Lieber¹⁸. DNA-PKcs was shown to interact with Artemis and regulate its activity via a mechanism which involves multi-phosphorylation events. In the absence of DNA-PKcs, Artemis possesses a 5' to 3' single-strand-specific exonuclease activity and is unable to open hairpins. Upon complex formation,

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Artemis is phosphorylated by DNA-PKcs and acquires an endonuclease activity to open hairpins generated by the RAD complex^{234,236}.

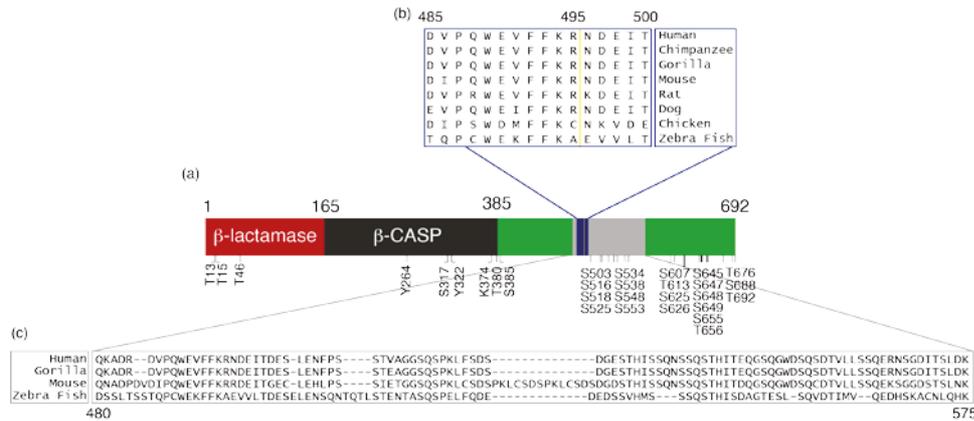


Figure 4.5: (a) Domain organization of the protein Artemis delineating the β -lactamase (red), β -CASP (dark gray) and C-terminal (green and light gray) regions. The region under study in this work is shown in light gray. The blue region highlights residues involved in the interaction with the DNA Binding Domain of Ligase IV (485-500). The yellow line separates the minimal interaction site (485-495) and the additional residues identified in our work (495-500) for the interaction with Ligase IV. (b) Protein sequence alignment among species of Artemis for amino acids 485-500 and (c) 480-575.

Since 2002, other kinases were shown to be involved in the regulation of Artemis activity through phosphorylation such as ATM (ataxia telangiectasia mutated) and ATR (ataxia telangiectasia and Rad3-related protein)²³⁷⁻²⁴⁰. Overall, 30 phosphorylation sites were identified *in vitro* and *in vivo* within the C-terminal region of Artemis, including 3 basal sites which role remains indeterminate (Figure 4.5-a). These different phosphorylation sites are likely to contribute to the regulation of Artemis endonuclease activity or interaction with its biological partners^{237,241}. To date, 14 of the 30 phosphorylation sites have been specifically attributed to DNA-PKcs kinase activity and the remaining sites were presumably attributed to ATM, ATR or DNA-PKcs kinases. Although binding to DNA-PKcs was shown to be essential to activate the endonucleolytic activity of Artemis, the functional significance of most phosphorylation sites remains unclear. Several studies have revealed that the phosphorylation of S516 and S645 by ATM plays an important role in the cell cycle regulation in response to DNA damage induced by ionizing radiation²⁴⁰. The phosphorylation of the same two sites by ATR was also demonstrated in response to other forms of DNA damage and cellular stress such as UV irradiation and hydroxyurea²⁴². Moreover, Artemis was proposed to act as a negative regulator of p53 in response to oxidative stress produced by

mitochondrial respiration^{238,243,244}. Overall, these findings indicate that, in addition to its nuclease activity, Artemis is a multifunctional protein which plays divergent roles in cellular metabolism. Even though the N-terminal catalytic core of Artemis confers its exo- and endonuclease activities, the identification of numerous phosphorylation sites specifically recognized by DNA-PKcs, ATM, or ATR in the C-terminal region clearly demonstrates the importance of the C-terminal domain for the regulation of the function of Artemis.

1.3.4 Artemis C-terminal region

The sequence of the C-terminal region of Artemis (residues 386-692) was submitted to protein databases to search for sequence similarities. The absence of sequence homology with others proteins led to define the C-terminal domain of Artemis as a new protein domain with unknown structure. Nevertheless, the high degree of conservation in the C-terminal domain amongst species (Figure 4.5-a-c) suggests a key role for this region in the function of Artemis^{232,245}. With the recent advances on the prediction of disordered regions in proteins (see the general introduction) the absence of known structure predicted in the C-terminal region can be re-interpreted. Indeed, the sequence analysis of Artemis using PreDisorder²⁴⁶ suggests a high degree of disorder in the C-terminal region 386-692 whereas the N-terminal domains are ordered (Figure 4.6).

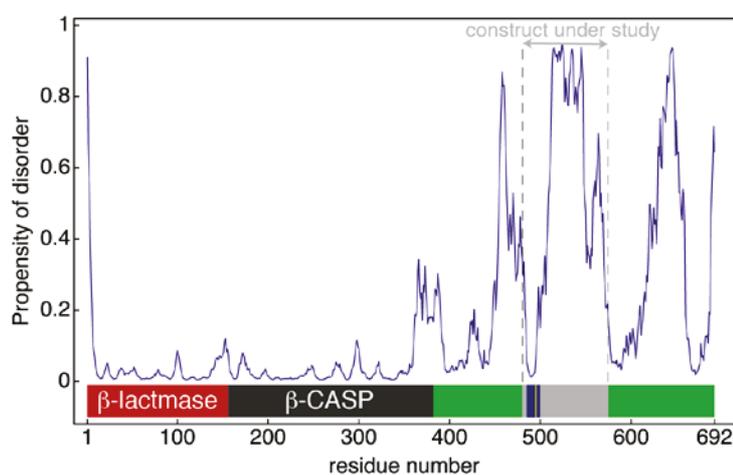


Figure 4.6: Disorder Prediction for the full Artemis protein using *PreDisorder*²⁴⁶.

The finding that a construct of Artemis lacking the disordered C-terminal region is able to fully rescue defects in V(D)J recombination pathway has led to a first model for the regulation of Artemis nuclease activity by DNA-PKcs^{232,237}. In this model, the C-terminal region is proposed to have an inhibitory role in the binding of Artemis to its DNA substrate.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Such inhibition would avoid genetic instability that could be induced by improper control of the nuclease activity. As shown in Figure 4.7, it is suggested that Artemis may adopt at least two distinct conformational states, “open” and “closed”, depending on the presence of DNA-PKcs. In the absence of DNA-PKcs, the C-terminal domain of unphosphorylated Artemis covers the entrance of the catalytic core, preventing Artemis to process DNA. The phosphorylation of Artemis by DNA-PKcs may induce a major conformational change that would result in higher exposition of the catalytic domain and activation of the nuclease activity. Overall, this model suggests that the transition between the “closed” and “open” states of Artemis is facilitated by the flexibility of the disordered C-terminal region (Figure 4.7). As mentioned above, most phosphorylation sites are located in the C-terminal region (Figure 4.5) and the favorable access of disordered regions to kinases could explain the higher level of phosphorylation in IDPs/IDRs in comparison with ordered regions.

The model for the regulation of the nuclease activity of Artemis based only on multi-phosphorylation within its C-terminal region was however challenged by recent findings. In particular, it was reported that multisite mutations of Artemis phosphorylation sites do not affect the nuclease activity *in vitro* or the ability of Artemis to rescue defects in the V(D)J pathway *in vivo*^{242,247}. These data are not compatible with a model in which phosphorylation of Artemis regulates its nuclease activities in NHEJ or V(D)J pathways. Interestingly, it was shown that autophosphorylation of DNA-PKcs is mandatory for the activation of the nuclease function of Artemis, suggesting that a conformational modification in DNA-PKcs is required to induce the nucleolytic activity of Artemis²⁴⁷. Overall, these new findings suggest that the entire disordered C-terminal region of Artemis is dispensable for its nuclease activities. Instead, phosphorylation of the C-terminal region would serve to regulate separate functions of Artemis in cell cycle regulation²⁴².

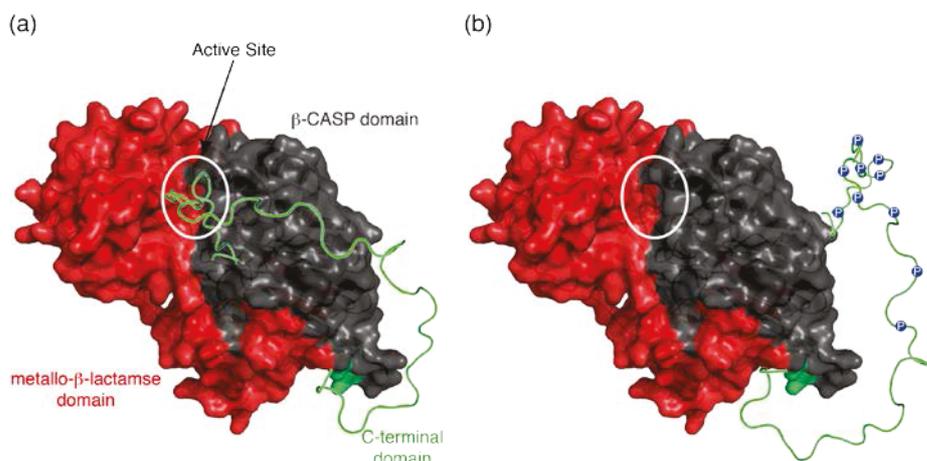


Figure 4.7: Putative model of Artemis activation by DNA-PKcs. The transition between (a) the “closed” state (without endonuclease activity) and (b) the “open” state (with endonuclease activity) is mediated by a conformational change of the C-terminal (green) domain upon phosphorylation. This transition makes the active site of the active core accessible (white ellipse). This active core consists of the β -lactamase (red) and the β -CASP (gray) domains (pdb code: 2I7V)²³⁷.

1.4 Ligase IV

More recently, the C-Terminal domain of Artemis was shown to be involved in the interaction with a second NHEJ factor, DNA Ligase IV²⁴⁸. The loss of interaction between Artemis and Ligase IV might be a new important factor for elucidating the molecular basis of the RS-SCID phenotype.

DNA ligase IV was cloned and sequenced in 1995²⁴⁹. It belongs, with DNA ligases I and III, to the ATP-dependent DNA ligases subfamily. This family is involved in a wide range of critical DNA processes such as DNA replication, repair, and recombination. DNA ligase IV is highly conserved in eukaryotes and ortholog forms have been found in yeast, plants and vertebrates²⁵⁰. The sequence of Ligase IV is characterized by the presence of a DNA Binding Domain (DBD) fused to a Catalytic Domain (CD), as observed in the sequences of ligases I and III (Figure 4.8-a). In addition, Ligase IV contains two repeats of BRCT (Breast and ovarian cancer susceptibility protein 1C-terminal) domain in the C-terminal region²⁵¹. Although the DBD domain of Ligase IV shares low sequence similarity with ligases I and III DBDs (sequence identity of 13% and 15%, respectively), the overall structure of this domain is similar in the 3 ligases (Figure 4.8-b)²⁵². It is composed of two helical subdomains connected by long loops, each subdomain containing 5 to 6 α -helices. The major difference between Ligase IV and Ligases I and III concerns the length of the helix α 2 that is respectively 2 and 3 turns longer in Ligase IV DBD.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

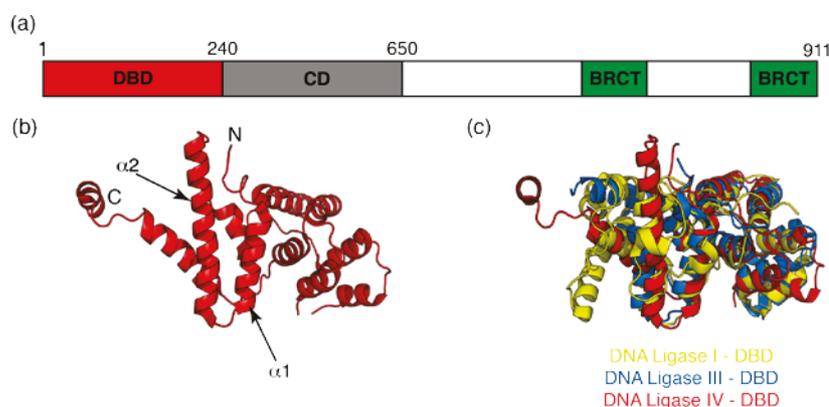


Figure 4.8: Structure of Ligase IV: (a) Domain organization of the full length DNA Ligase IV with the DNA Binding Domain (DBD), the Catalytic Domain (CD) and the repeats of BRCT domain; (b) Structure of DNA Ligase IV (PDB ID: code 4HTO) with two important helices labeled; (c) Structural alignment of DNA Ligase I-DBD (PDB ID: code 1X9N), DNA Ligase III-DBD (PDB ID: code 3L2P) and DNA Ligase IV-DBD (PDB ID: code 4HTO).

1.5 Interaction Artemis/Ligase IV

Artemis was found to interact with the complex Ligase IV-XRCC4 independently of its interaction with DNA-PKcs²⁴⁸. In particular, the implication of the C-terminal region of Artemis in the interaction with Ligase IV was revealed by association-binding experiments²⁴⁸. Further mutagenesis experiments have identified in the C-terminal region of Artemis three highly conserved residues, W489, F492, and F493 as crucial for the interaction, and an 11 amino acids peptide (485-DVPQWEVFFKR-495) as the minimal interaction site. Indeed, when mutated to alanine the three amino acids W489, F492, and F493 were shown to disrupt formation of the Artemis-Ligase IV complex but did not modify the interaction with DNA-PKcs²⁵².

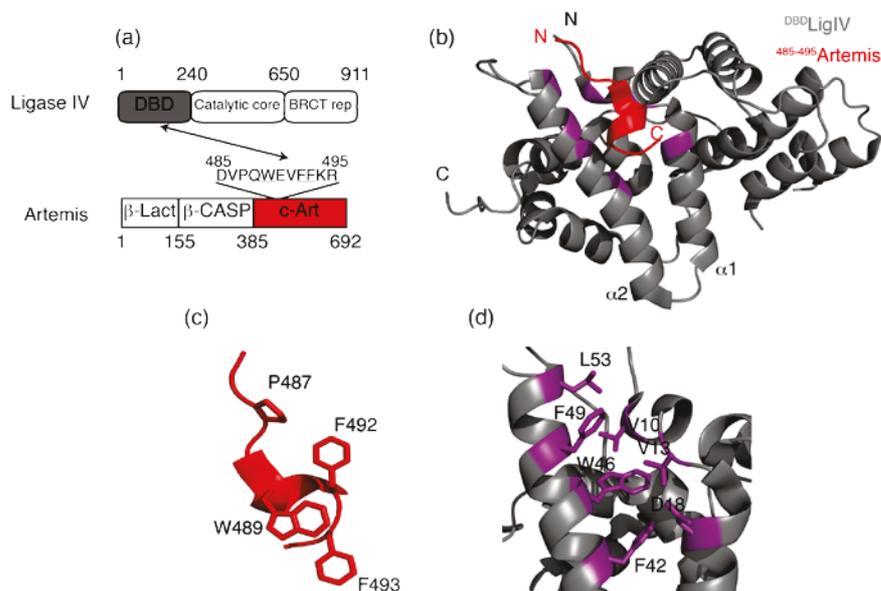


Figure 4.9: Crystal structure of ^{DBD}Ligase IV/sPepArt complex. (a) A schematic view of interacting domains and regions of Ligase IV and Artemis. (b) Overall structure of the ^{DBD}Ligase IV/sPepArt complex. ^{DBD}Ligase IV is shown in gray and sPepArt is shown in red. Important amino acids are highlighted for (c) sPepArt and (d) ^{DBD}Ligase IV²⁵². Residues colored in red (b, c) or purple (b, d) are identical.

The structure of the DBD of Ligase IV bound to a peptide corresponding to the region 485-495 of Artemis was determined using X-ray crystallography in two distinct studies^{252,253}. Both studies showed that Artemis interacts with $\alpha1$ and $\alpha2$ helices of Ligase IV-DBD through a large hydrophobic surface. Remarkably, the $\alpha2$ helix which is two turns longer in Ligase IV compared to Ligases I and III facilitates the formation of the hydrophobic pocket. In both crystal structures, the C-terminal residues of the Artemis peptide (W489-F493) adopt an α -helical conformation while the N-terminal residues (D485-Q488) adopt an extended β -strand conformation. The presence of a short helical motif is supported by the formation of two intra-molecular hydrogen bonds: N488(O)-V491(N) and W489(O)-F492(N). The Artemis-Ligase IV complex is stabilized by extensive van der Waals interactions between the 3 aromatic residues of the Artemis peptide that lie on the same side of the helical portion (W489, F492, and F493) and hydrophobic residues exposed at the surface of Ligase IV-DBD (V10, V14, F42, W46, and F49). Additional van der Waals contacts between P487 on the N-terminal extension of the Artemis peptide and L53 from the extended portion of helix $\alpha2$ of Ligase IV DBD contribute to the stabilization of the complex. Isothermal titration calorimetry

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

(ITC) was used to investigate the interaction between the Artemis 485-495 peptide and Ligase IV DBD in solution. A dissociation constant of 4.8 μM was found as well as a 1:1 binding stoichiometry²⁵². The finding that mutations of F42 or F49 to alanine in Ligase IV-DBD resulted in a complete loss of interaction has confirmed the crucial role of the hydrophobic surface formed by helices $\alpha 1$ and $\alpha 2$ in the binding to Artemis.

1.6 Aims of the study

The first objective of this work was to better understand the role of the C-terminal region of Artemis in the interaction with the DNA-binding domain of Ligase IV. X-ray crystallography and mutagenesis studies have allowed to elucidate, at least partly, the molecular basis for the interaction between Artemis and Ligase IV-DBD. However, the highly disordered C-terminal region in which the interacting site 485-495 is located could influence the binding process, just as this region was proposed to inhibit the binding to DNA-PKcs kinase. Because X-ray crystallography cannot be used to get structural information at atomic resolution on IDPs/IDRs, we undertook the structural characterization of the interaction between Artemis and Ligase IV by liquid-state NMR spectroscopy. For this study, we selected a 96-residues fragment from the disordered C-terminal region of Artemis (residues 480-575) that contains the Ligase IV minimal interaction region at the N-terminus. In parallel, taking advantage of the expected disordered properties of the fragment 480-575, we used it to apply and develop methodology based on nuclear spin relaxation for the study of IDPs/IDRs. The first step of this work consisted in the expression and purification of the Artemis fragment 480-575 and the DNA-binding domain of Ligase IV. For that purpose, our collaborators, Patricia Cortes and Aneel Aggarwal from the Mount Sinai School of Medicine in New York, kindly provided the plasmids for the expression system of both proteins. The Artemis fragment 480-575 and the DNA-binding domain of Ligase IV will be now referred as $\text{Art}^{\text{LigIV}}$ and $^{\text{DBD}}\text{LigIV}$, respectively.

2. Sample preparation of $\text{Art}^{\text{LigIV}}$ and $^{\text{DBD}}\text{LigIV}$

2.1 Expression and purification of $\text{Art}^{\text{LigIV}}$

2.1.1 General strategy

The construct $\text{Art}^{\text{LigIV}}$, corresponding to the region 480-575 of human Artemis (Swissprot accession number Q96SD1), was expressed as a Glutathione S-transferase (GST) fusion protein. A pGEX plasmid containing $\text{Art}^{\text{LigIV}}$ cDNA was transformed into an *E. coli* BL21 (DE3) pLysS expression strain. As shown in Figure 4.10-a, the pGEX- $\text{Art}^{\text{LigIV}}$ vector

Cyril Charlier

encodes a GST domain, a PreScission protease cleavage site, and the Art^{LigIV} fragment. Because of the cloning strategy, the Art^{LigIV} construct resulting from the cleavage by PreScission contains 8 additional residues at the N-terminal extremity, including 2 from the protease cleavage site (Figure 4.10-b) and a six residue-long spacer. The expected molecular weights are: 37.5 kDa for the fusion protein GST-Art^{LigIV}, 26.4 kDa for the cleaved GST and 11.1 kDa for Art^{LigIV}.

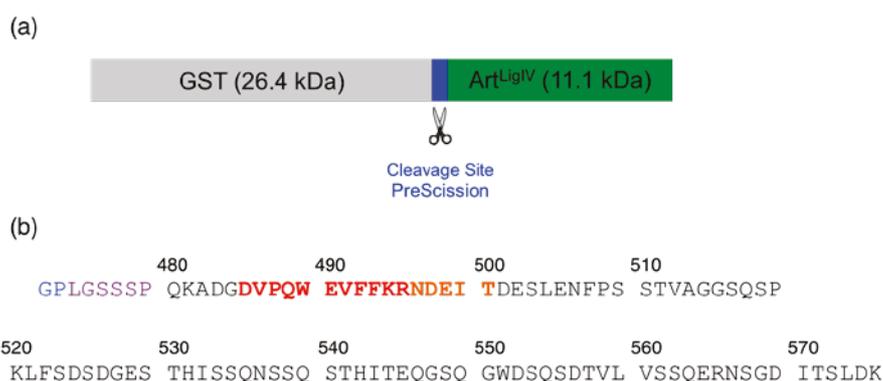


Figure 4.10: (a) Description of the fusion protein considered in this work, with the Glutathione-S-transferase (GST) in gray, the PreScission cleavage site in blue, and the Artemis fragment, Art^{LigIV} (480-575) in green. (b) Amino acid sequence of the 104-residue construct of Artemis obtained after cleavage by PreScission. Our construct contains 8 additional residues at the N-terminal extremity including 2 from the cleavage site (blue) and 6 from the cloning procedure (purple). The region marked in red corresponds to the known minimal interaction site with Ligase IV. Additional residues interacting with ^{DBD}LigIV identified in our work (see below) are marked in orange.

Unlabeled GST-Art^{LigIV} was produced in LB rich medium first. Uniformly labeled ¹⁵N and ¹⁵N/¹³C proteins were overexpressed in M9 minimum media using the procedure developed by Marley *et al.* for preparing labeled recombinant proteins with high yields²⁵⁴. Protein expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) in the cell cultures followed by an overnight incubation at 18°C under agitation. Cells were harvested by centrifugation and lysed by sonication and chemical treatment. Purification of Art^{LigIV} was carried out using a 2-steps protocol summarized in Figure 4.11. The fusion protein GST-Art^{LigIV} was first purified by glutathione-affinity chromatography and digested on-resin at 4°C with recombinant PreScission. A second step of purification was performed using size-exclusion chromatography to remove the PreScission protease and impurities (Figure 4.11). Pure fractions containing Art^{LigIV} were finally buffer exchanged and concentrated by ultrafiltration using a 3-kDa cutoff membrane. The reader is referred to the section “Materials and Methods” for extensive details about the sample preparation of Art^{LigIV}.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

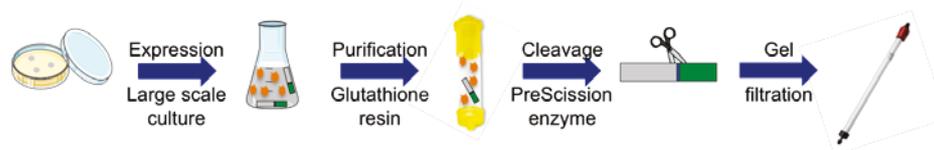


Figure 4.11: General strategy for the sample preparation of Art^{LigIV}

2.1.2 Results

We first attempted to prepare unlabeled Art^{LigIV} from cells cultured in LB medium. Protein expression and purification was followed by standard SDS-PAGE analysis using commercial gradient gels (Bio-Rad). Coomassie Brilliant Blue was used to stain proteins in the gels. As shown in Figure 4.12-a, SDS-PAGE analysis of the total cell lysate after IPTG induction revealed a high-intensity band at the expected molecular weight of GST-Art^{LigIV}. This high-intensity band at ~ 37 kDa was also observed for samples containing proteins bound to the glutathione resin before cleavage (Figure 4.12-b), indicating a proper expression of the fusion protein. Overnight cleavage by PreScission resulted in an efficient split of GST-Art^{LigIV}, as shown by the significantly weaker intensity of the band at ~ 37 kDa and the appearance of a strong band at a position corresponding to the predicted size of cleaved GST (~ 25 kDa). Surprisingly, the presence of the 25 kDa band was however not accompanied by the appearance of another strong band at the molecular weight of Art^{LigIV} (Figure 4.12-b). Only low-intensity bands could be detected below 15 kDa which suggested proteolytic degradation of the Art^{LigIV} fragment. A second intriguing observation was the presence of a high-intensity band at ~ 25 kDa in the “flow through” fraction containing proteins that no longer bound the glutathione resin after cleavage. The latter finding suggested that cleaved GST is not capable of binding the glutathione resin after digestion of the fusion protein. We further attempted to optimize the digestion process by performing PreScission cleavage in solution, with variable incubation time, on fractions containing GST-Art^{LigIV} eluted with reduced glutathione. Unfortunately, SDS-PAGE analysis led to the same observations with, in particular, an absence of band at the expected molecular weight of Art^{LigIV} (data not shown).

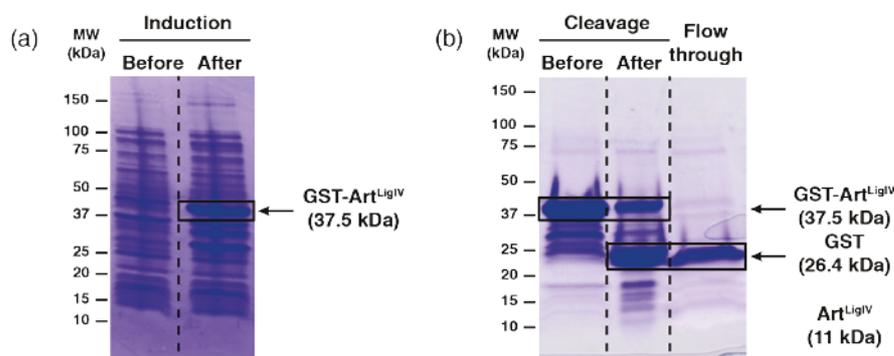


Figure 4.12: SDS-PAGE analysis of the expression (a) and purification (b) of unlabeled GST-Art^{LigIV}. (a) Lanes “Before” and “After” correspond to cell lysate before and after IPTG induction, respectively. The presence of a stronger band at ~ 37 kDa in the second lane indicates overexpression of the GST-Art^{LigIV} fusion protein (37.5 kDa). (b) Purification and PreScission cleavage of GST-Art^{LigIV} on glutathione resin. Lane “Before”, proteins bound to the GST resin after loading and washing; lane “After”, total fraction containing the GST resin after overnight cleavage by PreScission under agitation; lane “Flow through”, proteins that no longer bound the GST resin after PreScission cleavage. Migration bands were assigned based on expected molecular weights of GST-Art^{LigIV}, cleaved GST (26.4 kDa), and Art^{LigIV} (11.1 kDa).

Although we could not identify Art^{LigIV} after the PreScission cleavage, we went further in the purification process and loaded the “flow through” fraction from Figure 4.12-b into a gel filtration column. Elution of proteins containing aromatic residues, in particular tryptophans, was monitored by measuring UV absorbance at 280 nm. In the chromatogram presented in Figure 4.13-a, only one elution fraction (A3) was shown to significantly absorb UV at 280 nm. While SDS-PAGE analysis of the A3 fraction revealed a highly pure protein with an apparent molecular weight of ~ 25 kDa (Figure 4.13-b), mass spectrometry analysis of the same fraction rather showed one intense peak at 11.1 kDa (Figure 4.13-c), which perfectly matched the expected molecular weight of Art^{LigIV}. In addition, a 1D ¹H NMR spectrum recorded on the A3 fraction was found to display thin linewidths, consistent with the presence of a 11 kDa disordered protein rather than a globular and dimeric 26 kDa protein (GST is dimeric in solution). All together, our combined analysis using SDS-PAGE, mass spectrometry, and NMR demonstrated the presence of intact Art^{LigIV} in the purification fractions and highlighted an abnormal electrophoretic mobility of this disordered fragment during SDS-PAGE. Despite their significantly different molecular weights, GST and Art^{LigIV} exhibit similar migration rates during electrophoresis, which was particularly confusing and misleading. Abnormal migration of proteins containing a high content of polar residues, as Art^{LigIV}, has been previously reported and was correlated with a low affinity for SDS detergent, resulting in slower migration rates during electrophoresis²⁵⁵.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

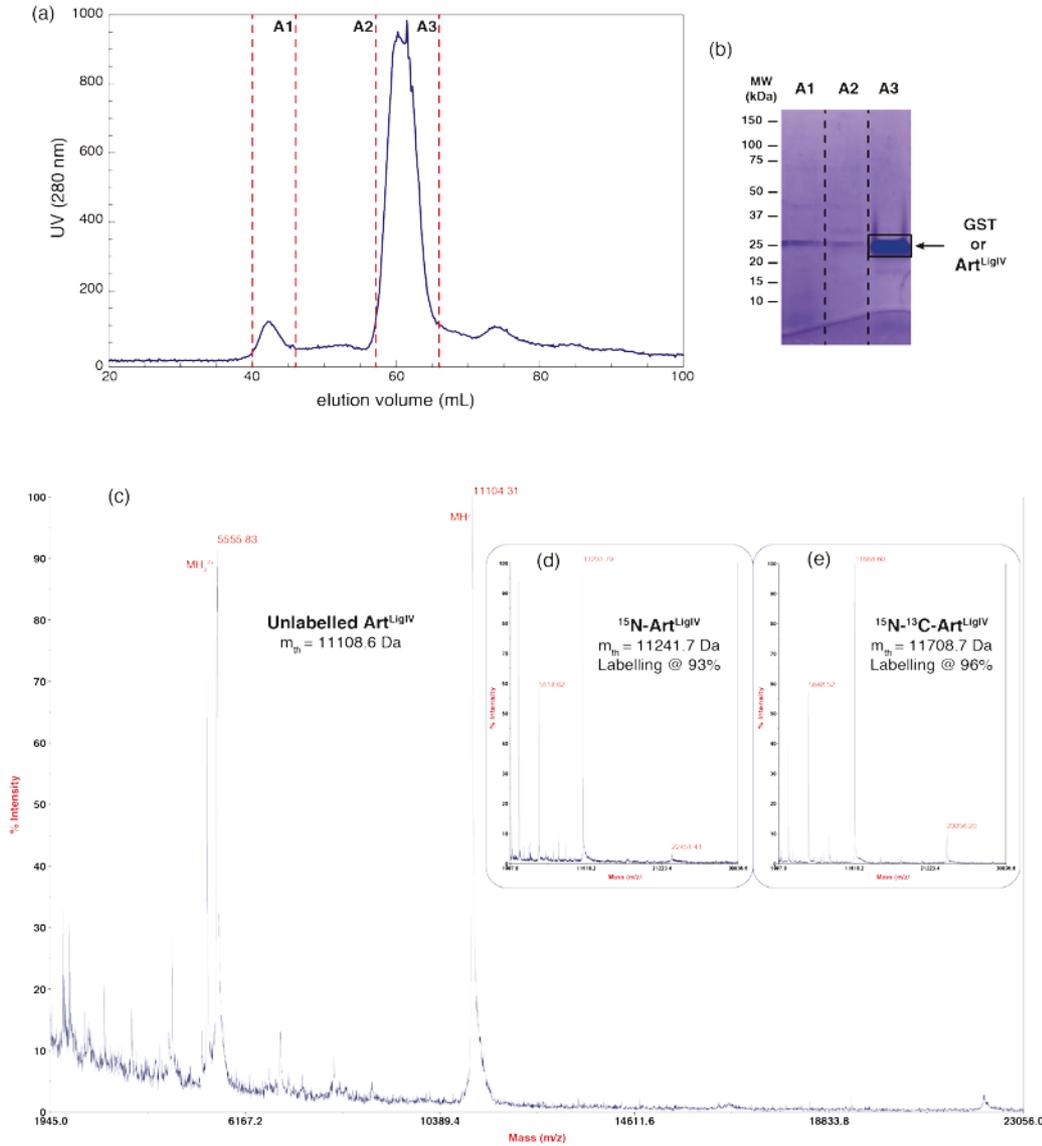


Figure 4.13: Purification and characterization of Art^{LigIV}. (a) Size-exclusion purification of the “flow through” fraction of the GST resin collected after PreScission cleavage. Based on UV absorbance at 280 nm, 3 new fractions (A1, A2, and A3) were collected from the Superdex column and analyzed by SDS-PAGE (b). A unique migration band was observed in lane A3 with an apparent molecular weight of ~ 25 kDa. (c) Mass spectrometry analysis of the A3 fraction rather indicated a molecular weight of 11107 +/- 5 Da, which perfectly matched the expected molecular weight of Art^{LigIV}. The two peaks with the highest intensity correspond to the single-charged (11104.3 Da) and doubly-charged (5555.8 Da) ions for Art^{LigIV}. Mass spectroscopy was also used to evaluate incorporation of ¹⁵N and ¹⁵N/¹³C isotopes in uniformly labeled ¹⁵N (d) and ¹⁵N/¹³C (e) Art^{LigIV}.

Finally, we could observe a slight difference in the migration rates between cleaved GST and Art^{LigIV} by reducing the protein concentration in the analyzed samples. As shown in Figure 4.14, on-resin digestion of ¹⁵N-GST-Art^{LigIV} resulted in the apparition of a thick band with an apparent molecular weight of ~ 25 kDa. A fine comparison of the SDS-PAGE profiles obtained for the “flow through” and elution fractions revealed a splitting of the 25 kDa band into two nearby but distinct bands, with the migration rate of Art^{LigIV} being slightly faster.

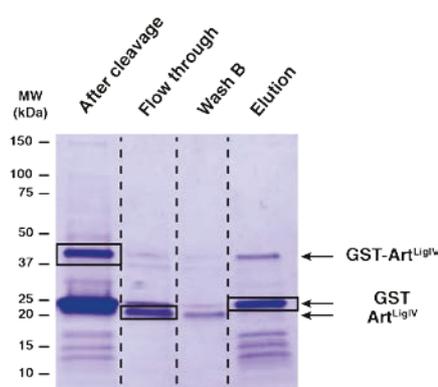


Figure 4.14: Purification and cleavage of uniformly labeled ¹⁵N GST-Art^{LigIV} monitored by SDS-PAGE. Lane “After cleavage”, total fraction containing the GST resin after overnight cleavage by PreScission under agitation; lane “Flow through”, proteins that no longer bound the glutathione resin after PreScission cleavage; lane “Wash B”, wash of the resin after cleavage; lane “Elution”, elution of the glutathione resin with reduced glutathione.

It should be stressed that an abnormal migration of Art^{LigIV} in the gel filtration column was also observed from the UV chromatogram in Figure 4.13-a. Indeed, injection of a sample containing the 14 kDa lysozyme protein onto the same Superdex column led to the elution of a single peak at ~ 80 minutes, which was 20 minutes later than Art^{LigIV}. Early elution of disordered proteins during size-exclusion chromatography is a well-known observation and can be expected from their higher Stokes radius compared to globular proteins with the same molecular weight.

The purification strategy summarized in Figure 4.11 was used to prepare highly pure, uniformly labeled ¹⁵N and ¹⁵N/¹³C Art^{LigIV}. Mass spectrometry indicated proper incorporation of the ¹⁵N or ¹⁵N/¹³C isotopes into Art^{LigIV}, with labeling yields of 93% and 96% for the ¹⁵N and ¹⁵N/¹³C samples, respectively (Figure 4.12-d-e).

2.2 Expression and purification of ^{DBD}LigIV

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

The DNA-binding domain of human Ligase IV (^{DBD}LigIV, region 1-240) was expressed as a (His)₆-tagged fusion protein. A pET-15b plasmid containing ^{DBD}LigIV cDNA was transformed into an *E. coli* BL21 (DE3) pLysS expression strain. The resulting expression vector encodes a 29.1 kDa fusion protein that contains the wild-type ^{DBD}LigIV sequence preceded by a thrombin cleavage site and a N-terminal (His)₆-tag. Unlabeled fusion protein was produced from cells cultured in LB rich medium. Purification of ^{DBD}LigIV was carried out using a four-step protocol. The fusion protein was first purified by nickel-affinity chromatography and digested in solution with thrombin protease. The cleaved ^{DBD}LigIV domain was further purified by cation-exchange chromatography to remove thrombin and the short (His)₆-tag. A last purification step was carried out using size-exclusion chromatography to obtain a highly pure protein sample. ^{DBD}LigIV was finally buffer exchanged and concentrated by ultrafiltration using a 3-kDa cutoff membrane. A more detailed protocol is presented in the “Materials and Methods” section.

As shown in Figure 4.15, SDS-PAGE analysis of purified ^{DBD}LigIV indicated a highly pure protein with an apparent molecular size that corresponds to ^{DBD}LigIV molecular weight.

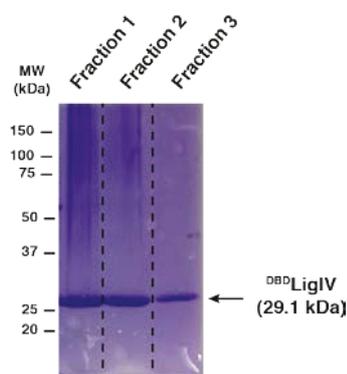


Figure 4.15: Purification of unlabeled ^{DBD}LigIV monitored by SDS-PAGE. The final sample obtained after size-exclusion chromatography was analyzed at three different concentrations (fractions 1, 2, and 3).

2.3 Materials and Methods

2.3.1 *Art*^{LigIV} samples

Transformation of BL21 cells

E. coli BL21 (DE3) pLysS competent cells were transformed with a pGEX plasmid containing *Art*^{LigIV} cDNA. First, 1 μ L of the pGEX plasmid was mixed with 50 μ L of BL21 competent cells and the mixture was incubated 30 minutes at 4°C. A thermal shock was applied at 42°C during 90 seconds. The mixture was transferred into 1 mL LB medium and

Cyril Charlier

incubated at 37°C for 45 minutes under agitation at 220 rpm. Cells were successively centrifuged, resuspended with 200 µL LB, spread on a LB-agar plate containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol, and incubated at 37°C for 16 hours. Stocks of transformed cells were prepared by adding 15% glycerol to single-colony cultures grown at 37°C in LB, and were conserved at -80°C.

Protein expression

Expression of isotopically labeled Artemis was performed using a procedure derived from the protocol developed by Marley *et al.*²⁵⁴ This expression protocol started with a 100 mL LB preculture inoculated with 250 µL of bacterial glycerol stock and supplemented with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. The preculture was incubated overnight at 37°C under agitation at 220 rpm. The next day, the preculture was diluted in 8 different flasks, each containing 500 mL LB medium and antibiotics (100 µg/mL AMP and 30 µg/mL CAM), to start with an OD₆₀₀ (optical density measured at 600 nm) of 0.05. The resulting 4 L culture was incubated at 37°C with shaking at 220 rpm. When OD₆₀₀ reached 0.6, cells were harvested by centrifugation at 2900xg for 15 minutes at 20°C, and pellets were resuspended in a total volume of 1 L M9 medium. After 40-minutes incubation at 37°C, ¹⁵NH₄Cl and glucose (¹²C or ¹³C) were added to the culture medium as required for ¹⁵N or ¹⁵N/¹³C labeling. After a new incubation period of 35 minutes at 37°C, protein expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) in the culture, followed by an overnight incubation at 18°C. The next day, cells were harvested by centrifugation at 2900xg for 20 minutes at 4°C, and pellets were resuspended in a total volume of 60 mL glutathione resin equilibration buffer containing 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM EDTA, 0.1% Triton, and 5% Glycerol.

Purification

Cell pellets were lysed by 10 sonication cycles interleaved with 1 minute breaks, and subsequently centrifuged at 30000xg for 40 minutes at 4°C. 4 mL of glutathione resin (GE Healthcare) were equilibrated with 50 mL of equilibration buffer. The supernatant containing soluble GST-Art^{LigIV} was mixed to the resin solution. After 2-hours incubation at 4°C under agitation, the resin was packed into a glass column. The flow through of the GST column was discarded and the resin was successively washed with 50 mL equilibration buffer and 50 mL PreScission protease cleavage buffer (25mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA, and 5% Glycerol). 1 mg of preScission protease was added to the resuspended resin to

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

perform an overnight, on-resin cleavage at 4°C under gentle agitation. The next day, the resin was packed into the glass column and the flow through fraction was collected. The column was washed with 5 mL precision protease cleavage buffer, and proteins bound to the GST resin were eluted with a buffer containing 10 mM reduced glutathione. Art^{LigIV} was further purified by size-exclusion chromatography, using a Superdex 75 column (GE Healthcare) previously equilibrated with a buffer containing 20 mM Bis-Tris pH 6.5, 150 mM NaCl, and 1 mM EDTA. Elution fractions were analyzed by SDS-PAGE, and fractions containing highly pure Art^{LigIV} were concentrated by ultrafiltration using a 3-kDa cutoff membrane (Millipore). A cocktail of anti-proteases (Roche) was added to the NMR samples.

2.3.2^{DBD} LigIV samples

E. coli BL21 (DE3) pLysS competent cells were transformed with a pET-15b plasmid containing ^{DBD}LigIV cDNA using the protocol described above for Artemis. A preculture containing 100 mL LB supplemented with 100 µg/mL ampicillin and 30 µg/ml chloramphenicol was inoculated with 250 µL of bacterial glycerol stock. After overnight incubation at 37°C, the preculture was diluted in a 1 L LB to start with an OD₆₀₀ of 0.05. Cells were grown at 37°C under agitation at 220 rpm. When OD₆₀₀ reached 0.6, protein expression was induced by adding 1 mM IPTG, and the culture was further grown for 3 hours at 37°C. Cells were harvested by centrifugation at 2900xg for 15 minutes at 4°C, and pellets were resuspended in 20 mL lysis buffer containing 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM Imidazole, 5% glycerol, 1 mM β-mercaptoethanol, 0.2 % Triton X-100, 1 mM PMSF, and 0.1% Lysozyme. Cell pellets were lysed by 10 sonication cycles interleaved with 1 minute intervals, followed by a centrifugation at 30000xg for 40 minutes at 4°C. The supernatant was loaded on a nickel column (HisTrap HP, GE Healthcare) previously equilibrated with binding buffer (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5% glycerol, and 1 mM β-mercaptoethanol). The fusion protein was eluted with an imidazole gradient using an elution buffer containing 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole, 5% Glycerol, and 1 mM β-mercaptoethanol. Fractions containing (His)₆-^{DBD}LigIV were dialyzed against the thrombin buffer containing 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% Glycerol, and 1 mM β-mercaptoethanol. Thrombin cleavage was performed overnight at 4°C under gentle agitation. ^{DBD}LigIV was further purified by cation-exchange chromatography using a SP Sepharose column (GE Healthcare), and by size-exclusion chromatography using a Superdex 75 gel filtration column. Fractions containing highly pure ^{DBD}LigIV were finally concentrated and buffer exchanged in 25 mM Tris-HCl pH 7.5,

Cyril Charlier

150 mM NaCl, and 1 mM TCEP.

3. Backbone Assignment of Artemis

3.1 Concept of backbone sequential assignment

The starting point of any protein study by Nuclear Magnetic Resonance consists in the assignment of backbone resonances. Proton resonances of small non-labeled proteins (<100 amino acids) can be assigned by a combination of 2D homonuclear experiments such as COSY (Correlation Spectroscopy), TOCSY (Total Correlation Spectroscopy), and NOESY (Nuclear Overhauser Effect Spectroscopy) or ROESY (Rotating frame Overhauser Effect Spectroscopy)²⁵⁶. For proteins containing more than one hundred residues, 2D ¹H-¹H spectra become too crowded to distinguish and assign each correlation. Double-resonance ¹H, ¹⁵N 3D experiments, carried out on ¹⁵N isotopically enriched proteins, allowed assignment of larger proteins^{174,257–259}. This was achieved using the implementation of Heteronuclear Single Quantum Coherence (HSQC)²⁶⁰ in 3D experiments such as ¹⁵N-¹H-HSQC-TOCSY and ¹⁵N-¹H-HSQC-NOESY, but also other schemes including Heteronuclear Multiple Quantum Coherence (HMQC) for instance. However, for α -helical proteins with molecular weight higher than 15 kDa, the efficiency of the ¹H TOCSY spin lock is often limited by the small values of the ³J_{HN-H α} coupling constants. To overcome such limitation, triple-resonance ¹H, ¹⁵N, ¹³C 3D experiments have been developed for backbone assignment of ¹⁵N/¹³C double-labeled proteins with molecular size from 15 to 30 kDa (nowadays, these efficient techniques are used even for small proteins). Most of these experiments use only coherence transfers via ¹J and ²J heteronuclear coupling constants¹⁵. Since the mid-1990s, assignment of proteins with a molecular weight higher than 50 kDa has been made possible thanks to new advances in the development of cryogenic probes and high field magnets, and the introduction of the TROSY scheme in 3D experiments²⁶¹ in combination with the use of ¹³C/¹⁵N/²H triple-labeled samples²⁶². The transverse relaxation-optimized spectroscopy (TROSY) experiment was introduced in 1997 in the group of K. Wüthrich to overcome the limitation of fast transverse relaxation in large proteins (> 30 kDa)²⁶¹. In addition, the new isotopic labeling strategies developed in the past few years allowed the assignment of very high-molecular weight proteins²⁶³, in particular with the expression of selectively-protonated perdeuterated proteins²⁶⁴. The idea of this approach is to express highly deuterated (> 98 %) proteins with targeted [¹³CH₃]-labeling at residue-specific methyl sites^{262,265,266}. Using these selectively labeled samples at methyl sites, new experiments such as optimized methyl

methods to speed up acquisition such as Non Uniform Sampling (NUS)²⁷⁴⁻²⁷⁶, and SO-FAST (Band-Selective Optimized-Flip Angle Short-Transient)²⁷⁷ or BEST (Band-selective Excitation Short Transient) experiments^{278,279}. The coherence transfer pathway of BEST experiments is similar to the standard experiments but uses a selective manipulation of the amide protons. Using selective pulses (E-BURP for instance) centered on the amide region, the polarizations of aliphatic and water protons show minimal perturbations. At the end of the sequence, due to the dipolar cross-relaxation between the detected amide and the unperturbed aliphatic protons, the polarization of amide protons is quickly recovered from the aliphatic and water polarization bath^{278,279}.

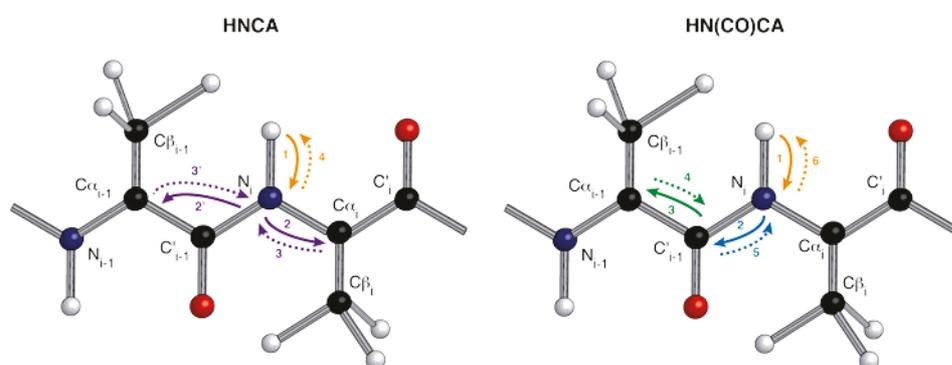


Figure 4.17: Magnetization transfers through scalar couplings used in the 3D experiments HNCA and HN(CO)CA.

In the case of disordered proteins, the assignment process may be more complex because of the lack of heterogeneity in chemical shifts values, resulting in the presence of overlapped regions in the 3D spectra. To avoid these problems, different strategies have been developed. For instance, by discriminating NMR signals onto a fourth (or more) dimension (as in the 5D-HACA(N)CONH or (HACA)CON(CA)CONH experiments)²⁸⁰, or by taking advantage of the larger chemical shift dispersion in carbon dimensions using ¹³C detected experiments²⁸¹. Experiments such as CON allow in parallel to identify and assign resonances from proline residues, usually abundant in IDPs. These ¹³C detected experiments were recently combined with BEST schemes (¹H^N-BEST¹³C^{ON})²⁸² and higher dimensionality experiments (5D (HACA)CON(CACO)NCO(CA)HA)²⁸³. In addition, the ¹⁵N chemical shift dispersion, relatively large in IDPs, can be used to establish complementary sequential correlations $N_{i+1}N_i$ between the sequential consecutive backbone amide ¹⁵N nuclei²⁸⁴⁻²⁸⁹. The

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

coherence transfer pathway ${}^1H_{(i+1)} \rightarrow {}^{15}N_{(i+1)} \rightarrow {}^{13}C'_{(i)} \rightarrow {}^{13}C^{\alpha}_{(i)} \rightarrow {}^{15}N_{(i)} \rightarrow {}^1H_{(i)}$ enables the recording of the 3D spectra that provide sequential correlations between the ${}^{15}N_{(i+1)}$, ${}^{15}N_{(i)}$ and ${}^1H_{(i)}$, backbone nuclei.²⁸⁷ In the present work, N, C α /C β and CO resonances of Art^{LigIV} were assigned using conventional heteronuclear triple resonance experiments.

3.2 Results of the assignment strategy

All the NMR data used for the backbone assignment were acquired on a Bruker Avance IIIHD 800 MHz spectrometer equipped with a room-temperature triple resonance 1H , ${}^{15}N$, ${}^{13}C$ probe with triple axis gradients. Spectra were processed with NMRPipe²⁰² and assigned with Sparky²⁹⁰. Backbone assignments 1H_N , ${}^{13}C_{\alpha}$, ${}^{13}C'$, and side-chains ${}^{13}C_{\beta}$ were achieved using a series of two-dimensional BEST-experiments ${}^{15}N$ - 1H -HSQC and triple resonance BEST-three-dimensional HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, and HN(CO)CACB spectra^{278,279} complemented with an (H)N(COCA)NH 3D spectrum²⁸⁷⁻²⁸⁹. Spectra were acquired on a sample containing 600 μM ${}^{15}N$ - ${}^{13}C$ Art^{LigIV} solubilized in 150 mM NaCl, 20 mM Tris-HCl (pH 6.5), and 1 mM EDTA.

The 1H - ${}^{15}N$ HSQC of Art^{LigIV} (Figure 4.18) shows a small dispersion of proton chemical shifts, from 7.5 to 8.5 ppm, which is typical of disordered proteins. Using the series of experiments detailed above we reached 96% of assignment for NH, C α , C β and CO resonances. Unassigned resonances include three prolines (487, 508, 519), two histidines (532, 542) and one glutamine (539). Proline resonances could not be assigned as we only recorded NH-base experiments. Interestingly, residues H532, Q539, and H542 are located in the same region and the lack of assignment for these residues may be due to internal dynamics or chemical exchange. Finally, several amino acids were not fully assigned mainly due to resonances overlap (Table 4.1).

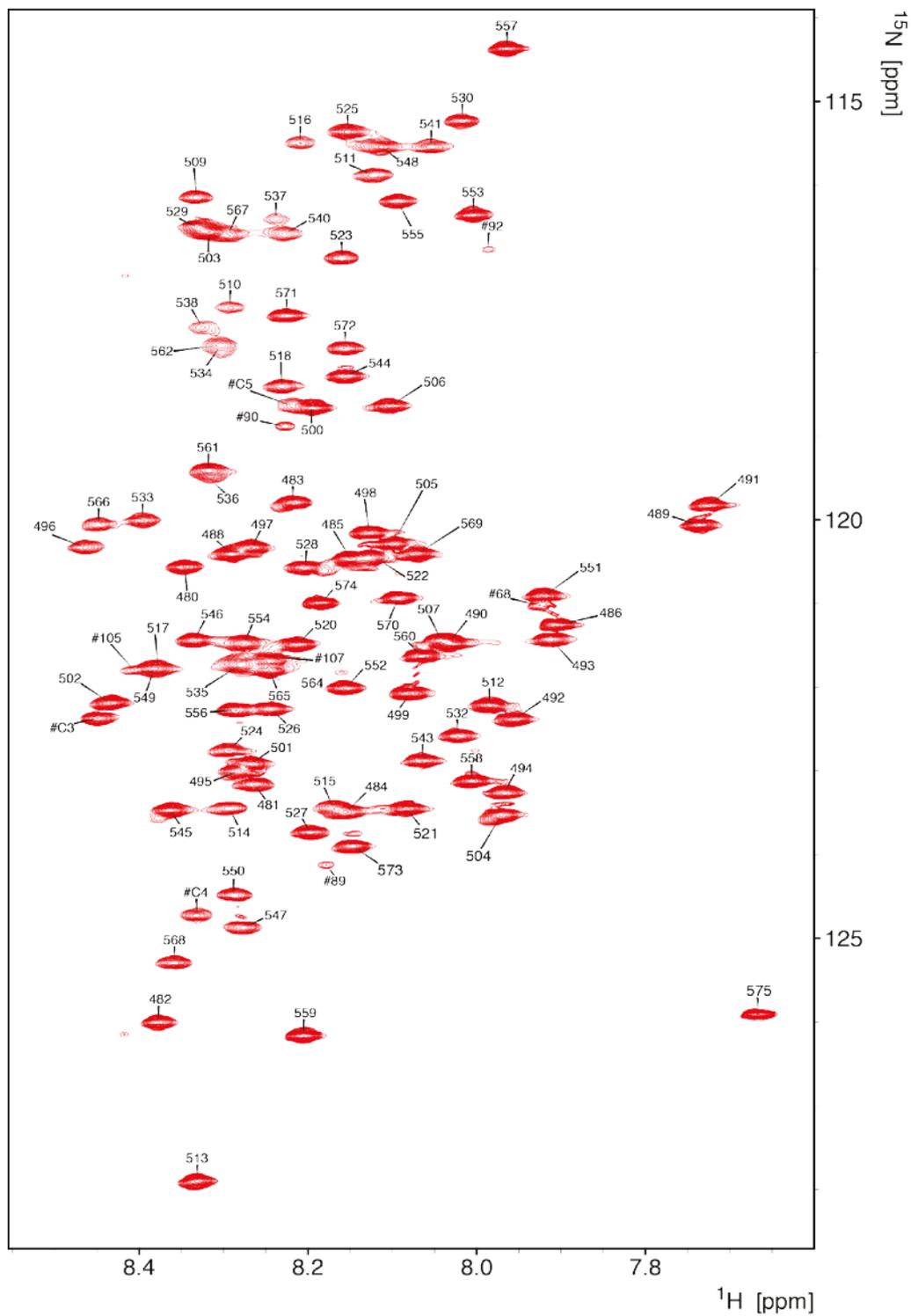


Figure 4.18: Two-dimensional ^1H - ^{15}H spectra of Art^{LigIV} in solution. Assignments of NH resonances are indicated on the spectrum.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Table 4.1: Backbone chemical shifts of Art^{LigIV}

Residue	Chemical Shift (ppm)					Residue	Chemical Shift (ppm)				
	¹⁵ N	¹ HN	¹³ C'	¹³ C α	¹³ C β		¹⁵ N	¹ HN	¹³ C'	¹³ C α	¹³ C β
Q480	120.564	8.346	175.89	55.768	29.492	E528	120.569	8.203	176.97	56.725	30.239
K481	123.158	8.264	176.178	56.23	33.148	S529	116.509	8.33	174.911	58.72	63.70
A482	126.006	8.377	177.386	52.381	19.346	T530	115.232	8.017	174.412	62.162	69.648
D483	119.797	8.218	176.702	54.547	41.173	H531					
G484	123.492	8.155	173.826	45.298		I532	122.586	8.023	176.291	61.26	38.656
D485	120.468	8.149	175.932	54.287	41.199	S533	120.009	8.396	<u>174.810</u>	58.438	63.873
V486	121.255	7.904	174.331	59.922	32.574	S534	117.914	8.303	<u>174.75</u>	58.697	63.718
P487						Q535	121.73	8.279	175.842	56.176	<u>29</u>
Q488	120.41	8.29	175.93	56.206	29.157	N536	119.532	8.314	175.433	53.361	38.923
W489	120.061	7.736	176.218	57.379	29.301	S537	116.405	8.238	175.40	58.717	63.846
E490	121.468	8.033	176.281	56.988	30.20	S538	117.702	8.325	174.855	58.77	63.71
V491	119.83	7.725	175.833	62.731	32.541	Q539					
F492	122.386	7.956	175.324	57.837	39.661	S540	116.583	8.228	176.092	58.506	63.786
F493	121.427	7.912	174.94	57.658	39.711	T541	115.545	8.054	174.25	62.029	69.696
K494	123.264	7.967	175.881	56.012	33.188	H542					
R495	123.001	8.282	176.15	56.134	31.057	I543	122.881	8.064	176.356	61.283	38.62
N496	120.329	8.464	174.971	53.455	38.844	T544	118.282	8.156	174.56	61.987	69.864
D497	120.332	8.267	176.023	54.451	41.00	E545	123.472	8.363	176.489	56.694	30.152
E498	120.16	8.127	176.309	56.614	30.293	Q546	121.439	8.335	176.595	56.332	29.244
I499	122.077	8.079	176.405	61.057	38.475	G547	124.868	8.276	174.311	45.346	
T500	118.666	8.195	174.162	61.563	70.071	S548	115.542	8.113	174.730	58.413	63.921
D501	122.919	8.269	176.501	54.435	41.157	Q549	121.777	8.382	176.418	56.187	29.216
E502	122.19	8.434	176.767	57.138	29.971	G550	124.489	8.288	174.083	45.324	
S503	116.561	8.317	174.865	58.994	63.679	W551	120.908	7.92	175.93	57.434	29.494
L504	123.53	7.971	177.53	55.379	42.21	D552	122.014	8.156	176.145	53.953	40.988
E505	120.288	8.1	175.936	56.724	30.131	S553	116.358	8.004	174.807	58.817	63.707
N506	118.642	8.103	174.293	53.053	39.095	Q554	121.48	8.276	176.186	56.094	29.174
F507	121.419	8.045	173.850	55.823	39.02	S555	116.192	8.094	174.341	58.638	63.9
P508						D556	122.275	8.286	176.4	54.616	41.048
S509	116.148	8.332	174.83	58.475	63.883	T557	114.371	7.964	174.491	62.237	69.813
S510	117.461	8.292		58.505	63.88	V558	123.121	8.006	175.949	62.617	32.516
T511	115.879	8.122	174.612	62.073	69.818	L559	126.167	8.205	177.283	55.144	42.277
V512	122.216	7.985	175.988	62.27	32.711	V560	121.628	8.065	176.259	62.422	32.745
A513	127.905	8.331	178.226	52.717	19.072	S561	119.41	8.318	174.79	58.379	63.876
G514	123.447	8.293	174.826	45.396		S562	<u>117.914</u>	<u>8.303</u>		<u>58.697</u>	<u>63.718</u>
G515	123.441	8.17	174.378	45.30		Q563	<u>121.638</u>	<u>8.247</u>	<u>176.096</u>		
S516	115.497	8.209	174.375	58.501	63.933	E564	<u>121.688</u>	<u>8.246</u>	<u>176.586</u>	<u>56.827</u>	<u>30.49</u>
Q517	121.777	8.382	175.74	55.59	29.41	R565	121.792	8.246	176.585	56.176	30.702
S518	118.405	8.231	175.734	56.578	63.309	N566	120.062	8.448	175.44	53.235	39.021
P519						S567	116.601	8.295	175.441	<u>59</u>	<u>63.7</u>
K520	121.491	8.213	176.243	56.171	32.871	G568	125.295	8.358	173.797	45.319	
L521	123.456	8.083	176.898	54.974	42.53	D569	120.41	8.071	173.796	54.25	41.174
F522	120.457	8.13	175.599	57.571	39.691	I570	120.937	8.093	176.723	61.506	38.574
S523	116.876	8.16	174.184	58.011	64.119	T571	117.563	8.225	174.692	62.26	69.837
D524	122.756	8.293	176.477	54.521	41.136	S572	117.952	8.156	174.415	58.295	63.787
S525	115.363	8.153	174.561	58.642	63.884	L573	123.9	8.147	177.032	55.232	42.351
D526	122.262	8.244	176.827	54.679	41.196	D574	120.993	8.185	175.009	54.457	41.008
G527	123.735	8.197	174.425	45.546		K575	125.91	7.667	175.644	57.551	33.746

3.3 Secondary structure from chemical shifts

It has been shown that it is possible to evaluate protein secondary structure from the chemical shifts of backbone (and C β) nuclei. Indeed, considering a given nucleus in a given residue, the deviation of the observed chemical shift from the random coil value ($\Delta\delta = \delta_{observed} - \delta_{coil}$) is a fair estimation of secondary structure propensity^{291,292}. To overcome the fact that the chemical shift of each individual backbone nucleus does not depend only on the local secondary structure²⁹³, the SSP program has been developed²⁹⁴. This program enables to take into account different nuclei ($^1\text{H}_N$, $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}'$, $^{13}\text{C}\beta$ and $^{15}\text{N}_H$) and has the advantage to weight the contribution of each residue depending on its contribution to the local secondary structure over five residues. This calculation in SSP program is done using the Equation (3.1) for a given residue i :

$$SSP_i = \frac{\sum_{j=i-2}^{i+2} \sum_X \begin{cases} \frac{|\Delta\delta X_j obs|}{\sigma_j \alpha} & \text{if } (\Delta\delta X_j obs)(\Delta\delta X_j \alpha) > 0 \\ -\frac{|\Delta\delta X_j obs|}{\sigma_j \beta} & \text{if } (\Delta\delta X_j obs)(\Delta\delta X_j \beta) > 0 \end{cases}}{\sum_{j=i-2}^{i+2} \sum_X \begin{cases} \frac{|\Delta\delta X_j obs|}{\sigma_j \alpha} & \text{if } (\Delta\delta X_j obs)(\Delta\delta X_j \alpha) > 0 \\ +\frac{|\Delta\delta X_j obs|}{\sigma_j \beta} & \text{if } (\Delta\delta X_j obs)(\Delta\delta X_j \beta) > 0 \end{cases}} \quad (4.1)$$

where $\Delta\delta X_j obs$ is the observed secondary chemical shift, $\Delta\delta X_j \alpha/\beta$ are the average secondary structure (α or β) and $\sigma_j \alpha/\beta$ are the standard deviations of $\delta X_j \alpha/\beta$. The summation over X corresponds to the nuclei types.

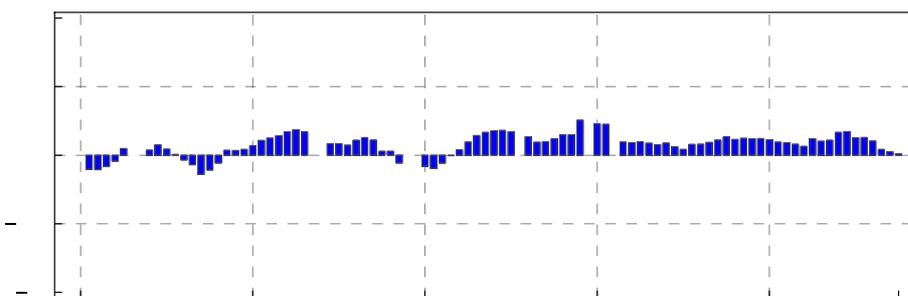


Figure 4.19: Secondary structure propensity (SSP) scores for Art^{LigIV} calculated from ^1HN , $^{13}\text{C}\alpha$, $^{13}\text{C}'$ and $^{13}\text{C}\beta$ chemical shifts. Positive and negative values represent α -helical and β -strand structure propensities, respectively²⁹⁴. An α -helix has an SSP score of 1 and a pure β -strand has an SSP score of -1.

In disordered proteins, the most informative nuclei for which chemical shift values are dominated by the secondary structure contribution are $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$. In our case, as we

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

did not assign $^1\text{H}\alpha$ [] only ^1HN , $^{13}\text{C}\alpha$, $^{13}\text{C}'$ and $^{13}\text{C}\beta$ chemical shifts were submitted to the SSP program. An SSP score of 1 usually reflects a fully formed α -helix whereas a SSP score of -1 reflects a β -strand. An intermediate SSP score of 0.7 at a particular position may indicate (i) an equilibrium between an α -helical structure populated at 70% and 30% of a random coil state or (ii) a somehow helical/turn-like structure populated at 100% or (iii) any more complex combination. As shown in Figure 4.19, all the SSP scores obtained for $\text{Art}^{\text{LigIV}}$ range from -0.15 to +0.3, showing the absence of stable secondary structure in the *apo*-form and a conformational ensemble close to the random coil state.

4. Mapping the interaction between $\text{Art}^{\text{LigIV}}$ and $\text{DBD}^{\text{LigIV}}$

4.1 Chemical shift mapping in binding studies by NMR

NMR is a powerful technique to investigate interactions of proteins due to its ability to provide site-specific information. In particular, when the interconversion between the protein-free and protein-bound conformational states is fast on the NMR timescale, one can perform a *chemical shift mapping* experiment. In this approach, a ^{15}N labeled protein is titrated with an unlabeled ligand and chemical shift changes of the protein resonances are monitored by 2D correlation spectroscopy. The chemical shift value strongly depends on the local magnetic environment of the observed nuclei. In the fast exchange regime, only one resonance will be observed for a given nucleus during the titration, and changes in the chemical environment of this nucleus are expected to modify its chemical shift. The common NMR experiments to follow chemical shift variations are the 2D ^1H - ^{15}N spectra such as HSQC, HMQC, or TROSY, which display a limited number of cross-peaks. Indeed, the number of NH correlations displayed in these 2D spectra roughly corresponds to the number of residues in the protein sequence (excluding NH correlations from side chains). One important point in a chemical shift mapping experiment is to keep the protein concentration almost constant through the addition of ligand, in particular when comparing cross-peak intensities. For this purpose it is more convenient to prepare a low-concentration protein sample and a highly concentrated ligand stock solution in order to add small volumes of the ligand solution to the protein sample, and neglect the dilution. Detailed protocols to perform chemical shift mapping can be found in the literature^{84,295,296}.

Chemical shift perturbations (CSP or Δ) can be calculated for each residue using weighted chemical shift variations $\Delta\delta_{\text{HN}}$ and $\Delta\delta_{\text{N}}$ for the proton and nitrogen, respectively²⁹⁷.

$$\Delta = \sqrt{\Delta\delta_{\text{HN}}^2 + (0.1\Delta\delta_{\text{N}})^2} \quad (4.2)$$

Cyril Charlier

The scaling factor of 0.1 was estimated from the difference in the gyromagnetic ratio between ^1H and ^{15}N ($\gamma_{\text{N}}/\gamma_{\text{H}} \sim 0.1$). It is used to scale ^{15}N CSP in the same order of magnitude as ^1H CSP.

Thus, the dissociation constant K_d is estimated by fitting the observed chemical shift evolution along the titration. For a typical two-state binding equilibrium:



The dissociation constant is given by:

$$K_D = \frac{[P][L]}{[PL]} \quad (4.4)$$

In the mixture the total protein ($[P]_t$) and ligand ($[L]_t$) concentration are defined as:

$$\begin{cases} [L]_t = [L] + [PL] \\ [P]_t = [P] + [PL] \end{cases} \Rightarrow \begin{cases} [L] = [L]_t - [PL] \\ [P] = [P]_t - [PL] \end{cases} \quad (4.5)$$

where $[P]$ and $[L]$ are respectively the protein and ligand concentrations in the free state, and $[PL]$ is the concentration of the complex, respectively. Combining Equations (4.4) and (4.5) leads to the determination of a second order Equation:

$$0 = [PL]^2 - (K_d + [L]_t + [P]_t)[PL] + [L]_t[P]_t \quad (4.6)$$

The solution of this Equation is given by the following Equation:

$$[PL] = \frac{(K_d + [L]_t + [P]_t) - \sqrt{(K_d + [L]_t + [P]_t)^2 - 4[L]_t[P]_t}}{2} \quad (4.7)$$

It is convenient to introduce the chemical shifts in Equation (4.7). In the case of a two-state equilibrium in a fast exchange regime, resonances shift continuously with the addition of ligand (Figure 4.20-a). In this exchange regime, peaks do not disappear from the spectra and the observed chemical shifts are directly related to the proportion of ligand-bound protein. Thus, we can introduce three different values of chemical shift where: δ_{free} is chemical shift in the free state, δ_{bound} is chemical shift in the bound state, and δ_{obs} is the weighted average of the free and bound states for each addition of ligand. The complex concentration can be written using the relationship:

$$[PL] = \left(\frac{\delta_{\text{obs}} - \delta_{\text{free}}}{\delta_{\text{bound}} - \delta_{\text{free}}} \right) [P]_t \Rightarrow [PL] = \frac{\Delta\delta_{\text{obs}}}{\Delta\delta_{\text{max}}} [P]_t \quad (4.8)$$

In addition, we define $\Delta\delta_{\text{obs}} = \delta_{\text{obs}} - \delta_{\text{free}}$ the chemical shift change, and $\Delta\delta_{\text{max}} = \delta_{\text{bound}} - \delta_{\text{free}}$ the chemical shift change at saturation. We obtain from Equation (4.7):

$$\Delta\delta_{\text{obs}} = \frac{\Delta\delta_{\text{max}}}{2} \left[\frac{(K_d + [L]_t + [P]_t) - \sqrt{(K_d + [L]_t + [P]_t)^2 - 4[L]_t[P]_t}}{[P]_t} \right] \quad (4.9)$$

The final step is to introduce the molar ratio defined as:

$$x = \frac{[L]_t}{[P]_t} \quad (4.10)$$

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Thus the dissociation constant K_d can be estimated by fitting the chemical shift changes ($\Delta\delta_{obs}$). To avoid the possible effect of dilution with the addition of ligand, the change in chemical shift at saturation ($\Delta\delta_{max}$) is also optimized.

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}}{2} \left[\frac{K_d}{[P]_t} + 1 + x - \sqrt{\left(\frac{K_d}{[P]_t} + 1 + x \right)^2 - 4x} \right] \quad (4.11)$$

More complete treatment and details about the determination of dissociation or association constants using NMR have been discussed^{298,299}.

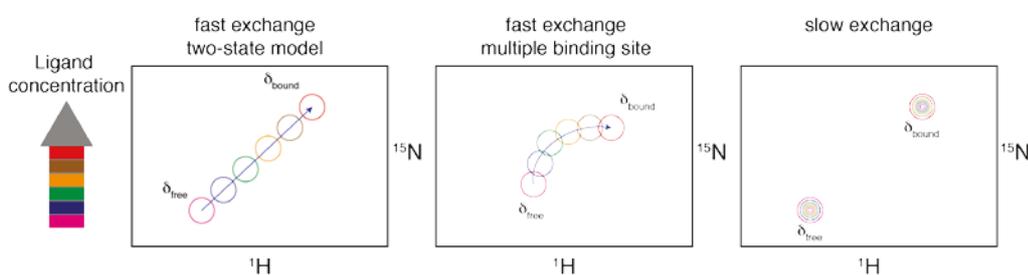


Figure 4.20: Dependence of chemical shift evolution on the mechanism and kinetic rates of the chemical exchange process. In fast exchange, peaks are moving along a straight line. When several sites are involved, the peaks follow a curved trajectory with addition a ligand. In the slow exchange regime, with the addition of ligand the intensity of the free signal decreases whereas the bound signal increases²⁹⁶.

In practice the chemical shift evolutions of the cross-peaks in the ^1H - ^{15}N spectra can be used as a “spy” to evaluate the equilibrium type between the free and ligand-bound type proteins. For a two-state equilibrium in a fast exchange regime, the correlation peaks move along a straight line throughout the titration whereas for a three-state equilibrium a curved path (or biphasic trajectory) is rather observed (Figure 4.20)^{296,300}. When the interconversion between the free and bound conformational states is slow on the NMR timescale, signals from both states are observed and the intensity of each peak can be used to estimate the fraction of bound protein²⁹⁶.

4.2 Results

^{15}N -Art^{LigIV} (150 μM) was titrated with a solution stock of unlabeled $^{\text{DBD}}$ LigIV at 775 μM up to 2 molar equivalents. Both proteins were dissolved in a buffer containing 150 mM NaCl, 20 mM Bis-Tris (pH 6.5), and 1 mM EDTA. A series of ^1H - ^{15}N HSQC experiments were acquired on a 600 MHz spectrometer using a triple-resonance probe. Spectra were processed with NMRPipe.

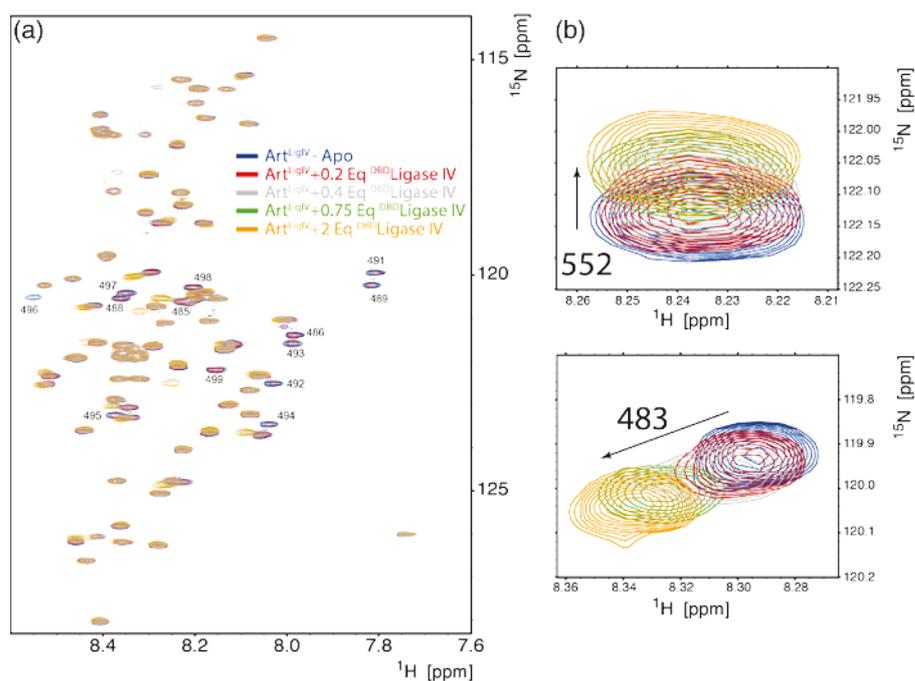


Figure 4.21: (a) ^{15}N - ^1H HSQC spectra recorded for the titration of $\text{Art}^{\text{LigIV}}$ by DBD-Ligase IV; *Blue:* $\text{Art}^{\text{LigIV}}$ Apo, *Red:* 0.2 molar equivalent DBD-Ligase IV, *Gray:* 0.4 molar equivalent DBD-Ligase IV, *Green:* 0.75 molar equivalent DBD-Ligase IV, *Yellow:* 2 molar equivalent DBD-Ligase IV. Labeled residues, discussed in the text, disappeared with addition of DBD-Ligase IV. (b) Selected regions of ^{15}N - ^1H HSQC with the signals of residues 552 and 483.

An overlay of the ^1H - ^{15}N HSQC spectra recorded on ^{15}N - $\text{Art}^{\text{LigIV}}$ in the presence of increasing amount of unlabeled $^{\text{DBD}}\text{LigIV}$ is shown in Figure 4.21-a. The binding to $^{\text{DBD}}\text{LigIV}$ leads to some significant changes in the HSQC spectra. The signals of the majority of residues are not altered but two types of chemical shift perturbations were observed for some residues. Residues at the heart of the interaction site (485-495) disappeared completely after adding ~ 0.3 molar equivalent. Interestingly, 4 other residues (496-499) at the C-terminal edge of the interaction site were also broadened beyond detection. The disappearance of NH correlations for residues 485-499 at an early stage of the titration suggests that this region undergoes slow to intermediate conformational exchange on the NMR timescale when bound to $^{\text{DBD}}\text{LigIV}$ with significant line broadening due to the exchange process. Only 2 new cross-peaks appear at the end of the titration, which may indicate unexpectedly fast relaxation for ^{15}N - ^1H spin systems of ^{15}N - $\text{Art}^{\text{LigIV}}$ in the bound form. The signals of many residues were slightly shifted upon addition of $^{\text{DBD}}\text{LigIV}$, as an indication of fast exchange (Figure 4.21-b). These observations are compatible with the presence of a single exchange process on a moderate timescale for which residues with large changes in ^{15}N chemical shifts are in the slow

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

exchange regime while residues with small changes in ^{15}N chemical shifts are in the fast exchange regime. Using Equation (4.2) chemical shifts perturbations (CSP) were calculated for all the residues of $\text{Art}^{\text{LigIV}}$ that are not in slow exchange (Figure 4.22). Overall, the low CSP values obtained for most residues in the region 510-575 suggest that this long region does not participate in the interaction with $^{\text{DBD}}\text{LigIV}$ although a few significant changes (residues 521, 523 or 552) might suggest a slight perturbation of the conformational ensemble. The highest CSP values are found for residues flanking the region 485-499. These CSP values, which are still quite small, probably reflect a change in the conformational ensemble due to proximity with $^{\text{DBD}}\text{LigIV}$ rather than a direct binding. The site-specific information obtained using chemical shift mapping indicates that residues 496-499, in slow exchange, are more affected than expected by $^{\text{DBD}}\text{LigIV}$ binding and are likely to play an important role in the interaction.

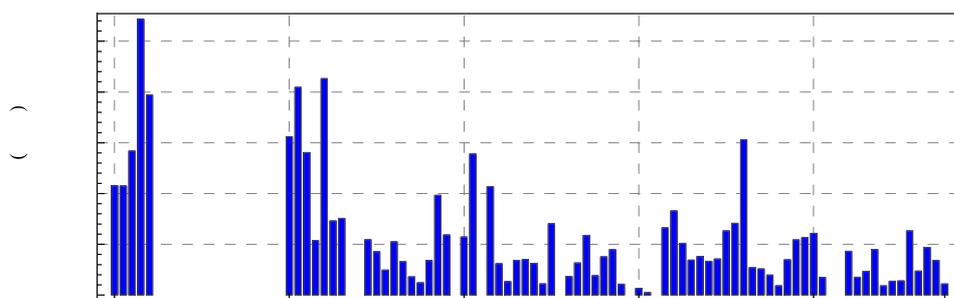


Figure 4.22: Plot of the weighted chemical shift perturbations calculated using Equation (3.2) versus residue number.

Figure 4.23 shows representative titration curves obtained from $\text{Art}^{\text{LigIV}}$ chemical shift values measured after each addition of $^{\text{DBD}}\text{Ligase IV}$. A global fitting procedure using Equation (4.11) applied on a selection of residues led to an estimated dissociation constant of $1.38 \pm 0.77 \mu\text{M}$. To take into account the effect of dilution, Δ_{max} was optimized. The fit was carried out under Mathematica with a Monte-Carlo procedure of 1000 steps. The value obtained is significantly lower than the dissociation constant reported by De Ioannes *et al.* using a peptide corresponding to the minimal interaction site 485-495, $4.8 \mu\text{M}$ by ITC²⁵². This result suggests that the $\text{Art}^{\text{LigIV}}$ fragment may have a higher binding affinity for $^{\text{DBD}}\text{LigIV}$ than the peptide 485-495, although two distinct techniques were used to evaluate the dissociation constants of two different constructs under slightly different experimental conditions.

A comparison under controlled conditions would be required to confirm the observed trend. To investigate the role of residues 496-499 in the binding to ^{DBD}LigIV, we further synthesized two Art^{LigIV} peptides: lPep-Art^{LigIV} (residues 485-501) and sPep-Art^{LigIV} (residues 485-495). The interaction of Art^{LigIV} and these two peptides with ^{DBD}LigIV was characterized by ITC, which is a more appropriate technique to determine accurate stoichiometry and thermodynamic parameters of an interaction.

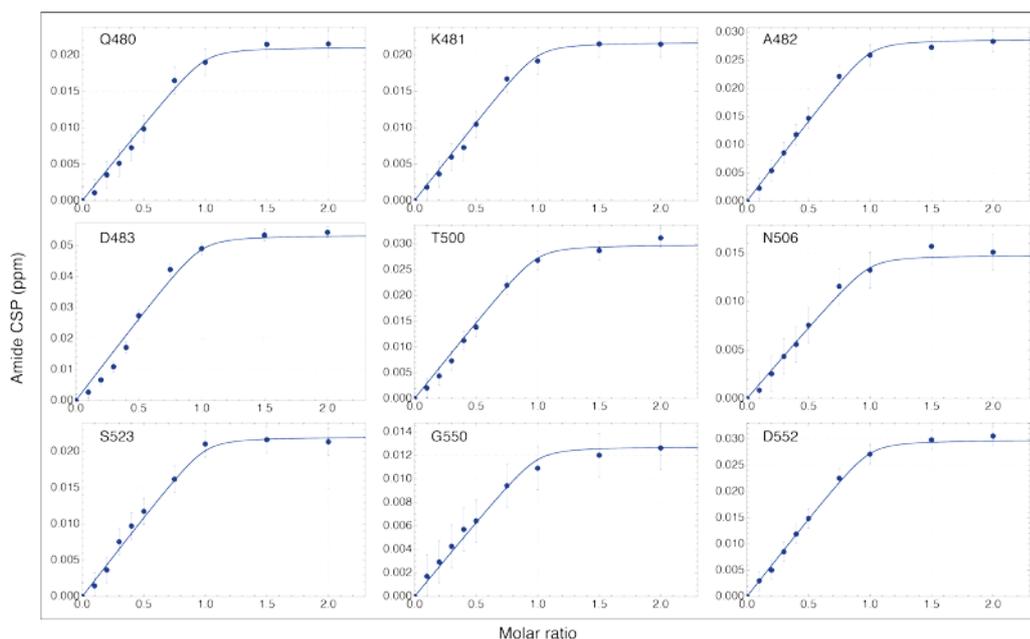


Figure 4.23: Amide CSP curves for selected residues used for the fit. Points correspond to the experimental value whereas the line is the fit obtained with a global K_d of $1.38 \pm 0.77 \mu\text{M}$.

4.3 Dissociation constant determination with isothermal titration calorimetry

4.3.1 Basic concept of isothermal titration calorimetry experiments

Isothermal titration calorimetry (ITC) is a technique used to determine the thermodynamic parameters of macromolecule-ligand interactions such as protein-protein, protein-DNA/RNA and protein-small molecule interactions. Small quantities of one of the binding partners are added to a solution of a macromolecule. Upon binding, heat is released (exothermic reaction) or absorbed (endothermic reaction) proportionally to the quantity of complex formed and the reaction enthalpy³⁰¹. Measurement of this heat transfer throughout the titration allows for a precise estimation of the dissociation constant (K_d), the reaction

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

stoichiometry (n) as well as the thermodynamic parameters, reaction enthalpy (ΔH), and reaction entropy (ΔS) of binding^{302,303}.

4.3.2 Experimental procedure

ITC experiments were performed by our colleague Astrid Walrant on a TA Instruments Nano ITC calorimeter at 23 °C in a buffer containing 20 mM BisTris pH 6.5, 150 mM NaCl, and 1 mM EDTA. Titrations were carried out by injecting 10 μ L aliquots of a Artemis solution (sPepArt: \sim 1.52-1.59 μ M, lPepArt: \sim 0.875 mM, Art^{LigIV} \sim 0.55 mM) into the calorimeter cell containing 1.2 mL of ^{DBD}LigIV (\sim 100-125 μ M for sPepArt and lPepArt, \sim 70 μ M for Art^{LigIV}), with 5 min delays between injections. Heats of dilution were measured by titrating Artemis constructs (sPepArt, lPepArt, Art^{LigIV}) into plain buffer under identical conditions. Thermodynamic parameters were determined by non-linear least-square fitting of the buffer-corrected data using the software NanoAnalyze (version 3.1.2) provided by TA Instruments.

4.3.3 Determination of the dissociation constant with ITC

ITC experiments were carried out to determine accurately the dissociation constant of the complex of Artemis and the DNA Binding Domain of Ligase IV. As shown in Figure 4.24 and Table 4.2, the enthalpy of binding (ΔH), the dissociation constant (K_d) and the stoichiometry (n) were determined for two peptides, sPepArt and lPepArt, (Figure 4.24-a-b) and for the fragment used in NMR experiments, Art^{LigIV} (Figure 4.24-c). The dissociation constant of sPepArt, $K_d = 11.7 \mu$ M, (Figure 4.24-d) slightly differs from the one previously reported by our collaborators ($K_d = 4.8 \mu$ M)²⁵². This variation could be explained by the different experimental conditions between the two measurements, in particular the pH in the present study was 6.5 instead of 7.5 in the study of *De Ioannes et al.* The dissociation constant measured for lPepArt, $K_d = 2.4 \mu$ M, and Art^{LigIV}, $K_d = 2.2 \mu$ M (in good agreement with NMR titration results), are similar and significantly smaller than the one measured for sPepArt (Figure 4.24-e-f). In both cases, the binding isotherms are characteristic of monophasic binding events and the best curve fittings were obtained using a one-site binding model. Overall, ITC results suggest that the five residues at the C-terminus of the originally described interaction site (485-495)²⁵² significantly contribute to the binding to ^{DBD}LigIV. In addition, it shows that the disordered region 502-575 does not affect directly the interaction.

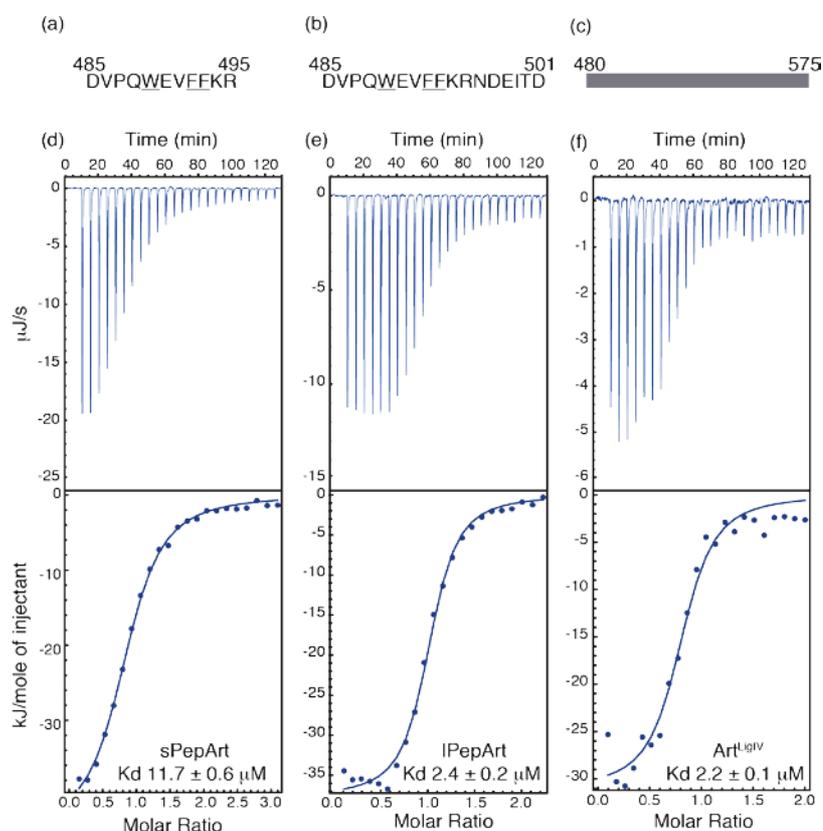


Figure 4.24: Isothermal titration of sPepArt (a) & (d), IPepArt (b) & (e), and Art^{LigIV} (c) & (f) with DBD-Ligase IV. The lower curves represent the heat of reaction (measured by peak integration) as a function of peptide/ligase molar ratio. The solid lines represent the best fits to experimental data. Thermodynamic parameters calculated from these experiments are shown in Table 4.2.

Table 4.2: Thermodynamic parameters for sPepArt, IPepArt and Art^{LigIV} binding to DBD-LigIV at 296.5 K. Data are given as average \pm SD, n=2 or 3.

	sPepArt	IPepArt	Art ^{LigIV}
Kd (μM)	11.7 \pm 0.6	2.4 \pm 0.2	2.2 \pm 0.1
ΔH for binding (kJ/mol)	-43.1 \pm 1.6	-38.7 \pm 1.4	-36 \pm 7
Binding stoichiometry (n)	0.92 \pm 0.08	1.01 \pm 0.01	0.92 \pm 0.17
Free energy ΔG (kJ/mol)	-28 \pm 0.13	-31.9 \pm 0.21	-32.1 \pm 0.12
Entropy ΔS (J/(mol.K))	-51 \pm 6.4	-23 \pm 2.9	-13.2 \pm 8.7

The free energy is more favorable for the interaction of the longer peptide ($\Delta G = -31.9$ kJ/mol) than the shorter peptide ($\Delta G = -28$ kJ/mol) with DBD-LigIV. The strong binding enthalpy observed for the two peptides contributes in a major way to the favorable free energy change that is measured for the peptides interaction with DBD-LigIV. Strong ion-pairs, hydrogen bonds, and van der Waals contacts dominate the unfavorable enthalpy associated with the desolvation of polar groups in the peptides and DBD-LigIV.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

The better free energy observed for the longer peptide is reflected in better affinity, compared to the shorter peptide, because for the longer peptide, the entropy change is less unfavorable. The additional residues in lPepArt are not expected to be pre-organized (see Figure 4.22) so that the binding of the longer peptide should be associated with a higher entropy cost. Entropy gains for the solvent due to the desolvation of charged residue might be at the origin of this less unfavorable entropy contribution to the binding of lPepArt.

Commentaire [FF4]: discuss with Ludovic?

4.4 Quantifying slow timescale motions with relaxation dispersion (RD)

In order to characterize the kinetics and mechanism of the binding reaction between Art^{LigIV} and DBD^{LigIV}, we performed a series of CPMG relaxation dispersion experiments on ¹⁵N-Art^{LigIV} with increasing quantities of DBD^{LigIV}, using the method described in chapter 2³⁰⁴.

4.4.1 Experimental section

Relaxation dispersion (RD) experiments were performed at 296.5 K on Bruker Avance 500 MHz using a triple-resonance indirect detection cryogenic probe equipped with a z-axis gradient coil. All the amide ¹⁵N RD experiments were recorded using the pulse scheme described in chapter 2. For each experiment, a series of 2D ¹⁵N-¹H correlation spectra with different CPMG frequencies $\nu_{CPMG} = 1/(4\tau_{cp})$ were recorded, where τ_{cp} is the spacing between successive 180° refocusing pulses. Typically, 25-30 points were recorded for each dispersion curve, corresponding to ν_{CPMG} between 20 and 1020 Hz. The total constant-time relaxation delay is set to $T_{relax} = 100$ ms. Error estimations were carried out by recording experiments with a single echo (typically 4 or 5 duplicates) and calculating the standard deviations between these experiments.

Commentaire [FF5]: This is not the most common definition of tauCP. You should make it appear in Figure 10 of the relaxation chapter so that it is fully clear.

Spectra were processed and analyzed using NMRPipe²⁰². Relaxation dispersion profiles ($R_{2,eff}(\nu_{CPMG})$) were calculated from peak heights according to $R_{2,eff}(\nu_{CPMG}) = -1/T_{relax} \ln(I(\nu_{CPMG})/I_0)$ with $I(\nu_{CPMG})$ being the peak height in the spectrum recorded with CPMG frequency ν_{CPMG} and I_0 the peak height in a reference spectrum recoded with $T_{relax} = 0$ ms.

Relaxation dispersion curves were fitted using an in-house Mathematica script employing the analytical expression of Carver-Richards Equation (see chapter 1)¹¹⁵.

4.4.2 Analysis of CPMG relaxation dispersion

NMR relaxation dispersion was performed on several mixtures of the Art^{LigIV}: DBD^{LigIV}. CPMG dispersion profiles were acquired on ¹⁵N-Art^{LigIV} with 0%, 3%, 6%, 10%

and 13% equivalent of $^{DBD}\text{LigIV}$ (Figure 4.26). Examination of the measured dispersion curves reveals the presence of an exchange contribution to relaxation for all residues in the region 485–499. No exchange is detectable with CPMG relaxation dispersion outside of this region.

The Carver-Richards¹¹⁵ Equation was fitted to the experimental data to extract exchange parameters. We assumed the population in the excited state to be equal to the molar ratio of $^{DBD}\text{LigIV}$ ($p_B = 0, 0.03, 0.06, 0.1$ and 0.13). The effective transverse relaxation rate, $R_{2,eff}$, in the free state was estimated using the average of the relaxation dispersion profile in the absence of $^{DBD}\text{LigIV}$. Experimental uncertainties on the experimental parameters reported in Table 4.3 were calculated using a *Monte Carlo* approach. Following the hypothesis that the binding event occurs on a single timescale, global fitting procedure for all residues at a given mixture (Table 4.3) and for all mixtures were performed (Table 4.4). At the time of writing the error estimation for the global fitting with all the mixture is not done yet (Table 4.4).

Table 4.3: Global fitting parameters for each Art^{LigIV}:DBD-LigIV ratio

Residue	$ \Delta\omega $ (ppm)			$R_{2,eff}^{bound}$ (s^{-1})		
Global fit with 3 % DBD-LigIV : $k_{off} = 56.18 \pm 0.71 s^{-1}$						
486	1.57	±	0.06	9.6	±	0.2
488	3.30	±	0.10	45.9	±	2.6
491	4.52	±	0.40	71.3	±	8.6
492	2.75	±	0.46	58.6	±	5.6
493	1.98	±	0.19	38.8	±	2.0
494	3.11	±	0.39	31.4	±	2.4
495	1.63	±	0.04	12.5	±	0.9
496	1.50	±	0.16	12.7	±	3.5
497	1.76	±	0.19	15.0	±	1.5
498	1.05	±	0.05	13.5	±	3.5
499	0.95	±	0.06	18.0	±	0.7
Global fit with 6 % DBD-LigIV : $k_{off} = 51.44 \pm 0.18 s^{-1}$						
486	1.52	±	0.00	17.3	±	0.1
488	2.97	±	0.07	47.8	±	0.2
491	6.28	±	0.13	86.9	±	8.2
492	7.65	±	0.09	70.4	±	3.6
493	2.87	±	0.01	57.9	±	1.5
494	3.89	±	0.29	48.4	±	2.3
495	1.60	±	0.006	33.3	±	0.0
496	1.35	±	0.09	24.0	±	0.1
497	2.16	±	0.09	23.6	±	0.0
498	1.11	±	0.003	16.2	±	0.2
499	0.79	±	0.01	16.1	±	0.0

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Global fit with 10 % DBD-LigIV : $k_{\text{off}} = 60.44 \pm 0.06 \text{ s}^{-1}$						
486	1.46	±	0.01	16.4	±	0.2
488	3.38	±	0.05	51.0	±	0.2
491	7.23	±	0.37	57.5	±	4.7
492	6.07	±	0.56	67.4	±	8.3
493	3.06	±	0.21	52.0	±	5.6
494	4.13	±	0.18	50.3	±	3.2
495	1.57	±	0.13	22.0	±	0.1
496	1.53	±	0.15	25.6	±	0.1
497	2.10	±	0.11	25.5	±	1.5
498	1.08	±	0.008	16.0	±	0.4
499	0.89	±	0.002	19.8	±	0.1
Global fit with 13 % DBD-LigIV : $k_{\text{off}} = 62.41 \pm 0.61 \text{ s}^{-1}$						
486	1.66	±	0.02	17.5	±	0.1
488	3.27	±	0.08	47.3	±	0.5
491	6.70	±	0.17	73.9	±	0.9
492	6.57	±	0.23	73.9	±	4.3
493	2.78	±	0.12	51.3	±	2.3
494	3.87	±	0.10	47.9	±	0.0
495	1.85	±	0.008	24.1	±	0.1
496	1.44	±	0.06	23.1	±	0.0
497	2.12	±	0.009	22.8	±	0.3
498	1.23	±	0.01	14.6	±	0.3
499	0.85	±	0.00	15.2	±	0.0

The results of semi-global approach consisting in fitting of all the mixtures separately are very similar for each of them. In addition the global fitting with all mixtures together leads to a 60 s^{-1} exchange rate (Figure 4.26). However, the large transverse relaxation rates extracted in the bound state (Figure 4.25, Table 4.4) suggest the presence of an additional fast exchange process. For instance, this additional exchange could be between an aligned encounter complex^{82,305} and the well-defined structure solved by X-ray crystallography. To date, our data at a single field (500 MHz) and on a single nucleus (¹⁵N) are not sufficient to use a three state model in the fitting procedure. It seems necessary to obtain CPMG or rotating-frame relaxation dispersion at higher effective fields in order to characterize the faster exchange process. This may be accessible with ¹H relaxation dispersion using a deuterated sample, or ¹³C' relaxation dispersion^{82,304,306}. Finally the potential assignment of resonances in the interaction site (485-499) in the bound state will definitely help to interpret chemical shift differences extracted from the fit. This will likely require deuteration of Artemis as well.

Table 4.4: Global fitting parameters extracted for Art^{LigIV}

Residue	$ \Delta\omega $ (ppm)	$R_{2,eff}^{bound}$ (s ⁻¹)
	$k_{off} = 60 \text{ s}^{-1}$	
486	1.6042	17.4
488	3.2456	49.3
491	6.7407	76.6
492	6.4321	75.9
493	2.8327	55.6
494	3.9316	48.5
495	1.7219	27.5
496	1.4832	23.6
497	2.1574	23.8
498	1.1737	15.6
499	0.8443	16.3

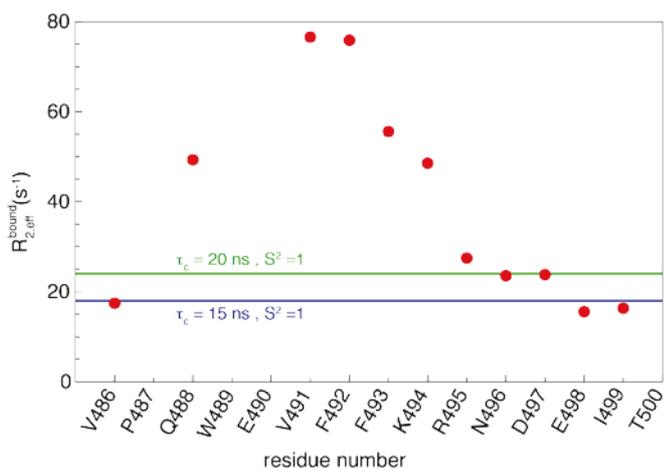


Figure 4.25: $R_{2,eff}^{bound}$ extracted from the global fit (Table 4.4). Expected R_2 calculated for a spectral density function with (green) $\tau_c = 20 \text{ ns}, S^2 = 1$ or (blue) $\tau_c = 15 \text{ ns}, S^2 = 1$ are shown as horizontal lines.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

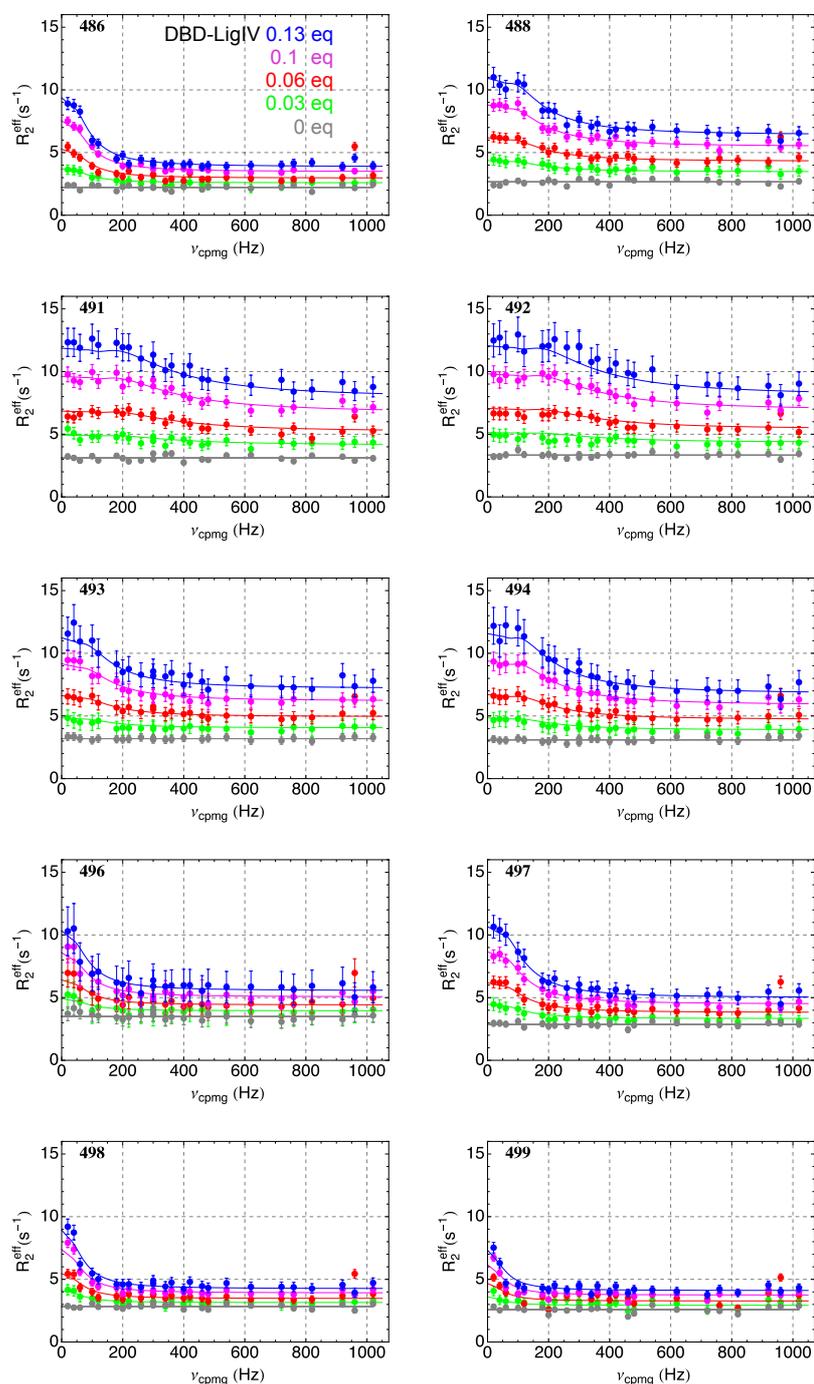


Figure 4.26: ¹⁵N CPMG relaxation dispersion curves for selected residues measured at 11.4 T (500 MHz ¹H Larmor frequency) for 0% (gray), 3% (green), 6% (red), 10% (magenta) and 13% (blue) molar ratio admixtures of Art^{LigIV}:DBD-LigIV. Solid lines correspond to a global fit of k_{off} constant and residue specific $\Delta\omega$ and $R_{2,\text{eff}}^{\text{bound}}$.

5. Quantifying pico- to nanosecond motions from ^{15}N relaxation

While CPMG relaxation dispersion provides information on processes that occur typically on millisecond timescales, ^{15}N relaxation measurements lead to the characterization of pico-nanosecond motions. We measured ^{15}N relaxation rates at various static high-fields in the presence or absence of Ligase IV. In addition, we complete our study with the measurements of a series of longitudinal relaxation rates at low field using high-resolution relaxometry^{83,141}.

5.1 Experimental sections

Nitrogen-15 relaxation rates have been recorded at four static magnetic fields, corresponding to proton Larmor frequencies of 950 MHz, 800 MHz, 600 MHz and 500 MHz. Experiments were acquired at 950 MHz using a 5-mm TCI cryoprobe, at 800 MHz and 600 MHz using room-temperature triple resonance probes and on a Bruker DMX 500 MHz spectrometer equipped with a 5-mm TCI cryoprobe. Readers are referred to chapter 1 for extensive details about the pulse sequences use in this work.

Relaxometry experiments were carried out on 700 MHz spectrometer equipped with a 5-mm TCI cryoprobe. The pulse program used here is a slightly modified version of the pulse sequence used in the chapter on ubiquitin, in order to avoid radiation dumping during the t_1 evolution in the same way as in the sequence used to measure steady-state ^{15}N - $\{^1\text{H}\}$ nuclear Overhauser effects.

Spectra were processed with NMR-Pipe²⁰² and analyzed with an in-house Mathematica program. Experiments were carried out using samples of 600 μM ^{15}N Art^{LigIV} (pH 6.5, 150 mM NaCl, 20 mM Bis-Tris and 1mM EDTA with 8% D₂O). Experiments in the presence of ^{DBD}LigIV were performed by mixing with a stock of unlabeled ^{DBD}LigIV at 775 μM in the same buffer conditions. Extensive details about the sample preparation are available above.

5.2 ^{15}N relaxation rate measurements on Art^{LigIV} apo at high fields

For four static high-fields, we measured a set of relaxation experiments: longitudinal relaxation rates (R_1), transverse relaxation rates (R_2) using train of ^{15}N echoes (CPMG pulse train) and using a single echo (only at 800 MHz and 950 MHz), Nuclear Overhauser Effect (NOE) and the longitudinal and transverse cross-relaxation rates due to correlated fluctuations of the nitrogen-15 CSA with the dipolar coupling between the ^{15}N nucleus and the amide proton (Figure 4.27).

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

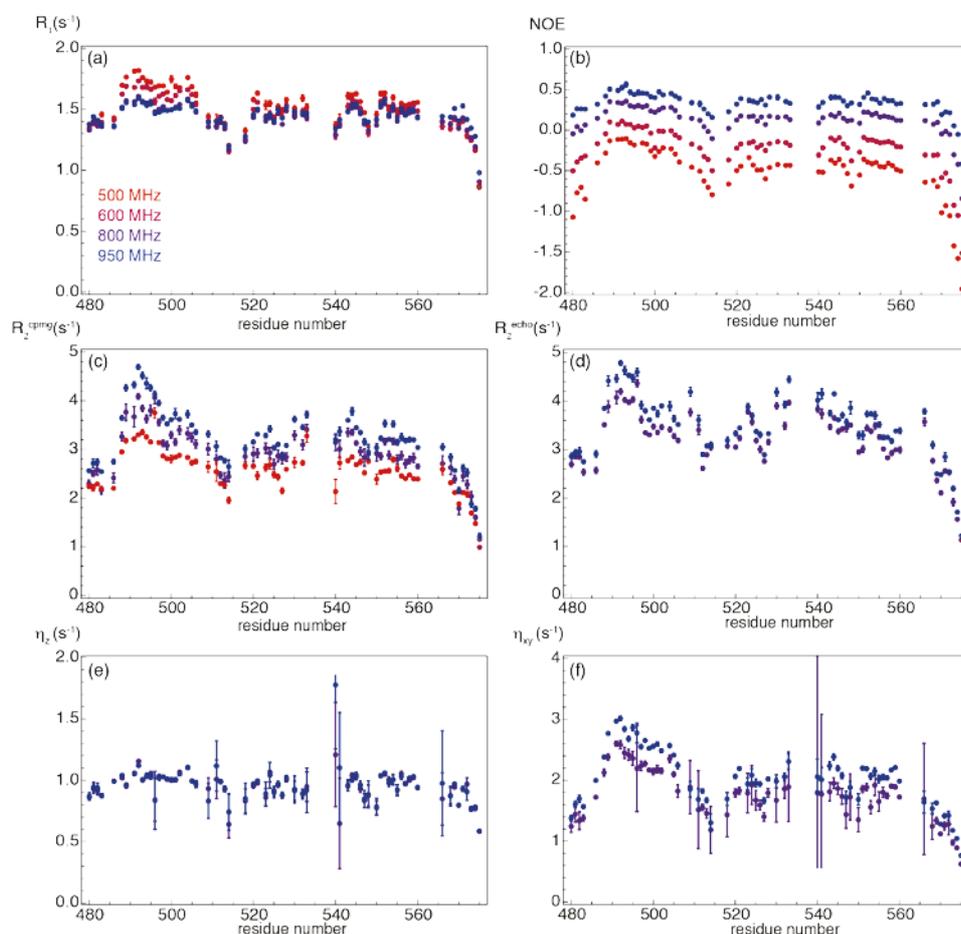


Figure 4.27: Backbone ^{15}N relaxation measurements on $\text{Art}^{\text{LigIV}}$ at 500 MHz (red), 600 MHz (burgundy), 800 MHz (purple) and 950 MHz (blue) (a) Longitudinal relaxation rates R_1 ; (b) $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE ratios; (c) Transverse relaxation under a CPMG train (R_2^{cpmg}); (d) Transverse relaxation under a simple echo (R_2^{echo}); (e) CSA-DD longitudinal cross-relaxation rates (η_z); (f) CSA-DD transverse cross-relaxation rates (η_{xy});

The measured rates are characteristic of a disordered protein. Indeed, the R_1 rates are very similar depending of the field (Figure 4.27-a) and the NOE are fully negative at moderate fields (500 MHz and 600 MHz) and as high as 0.5 at the highest field (950 MHz) for some of the residues (Figure 4.27-b). The transverse relaxation rates measured with CPMG (Figure 4.27-c & Figure 4.28) and single echo schemes (Figure 4.27 & Figure 4.28) do not show evidence of significant ($R_{\text{ex}} > 1 \text{ s}^{-1}$) chemical exchange contributions in *apo* $\text{Art}^{\text{LigIV}}$.

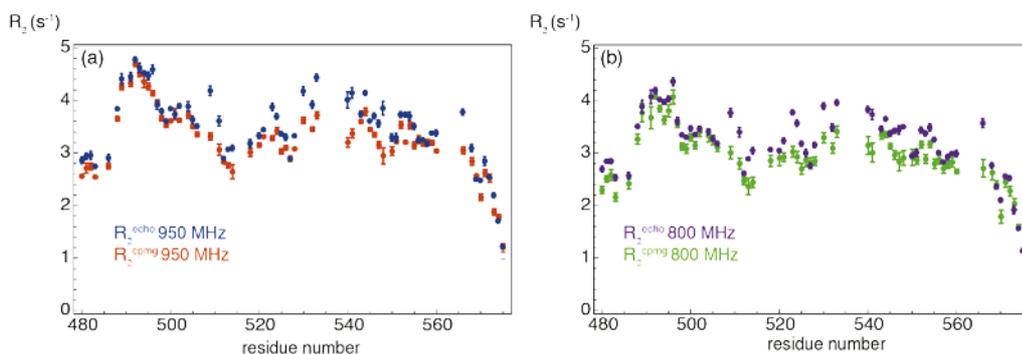


Figure 4.28: Backbone ^{15}N transverse relaxation measurements on $\text{Art}^{\text{LigIV}}$ at (a) 950 MHz and (b) 800 MHz. (a) R_2^{cpmg} (orange) and R_2^{echo} (blue); (b) R_2^{cpmg} (green) and R_2^{echo} (purple);

5.3 Interpretation of pico- to nanosecond motions in $\text{Art}^{\text{LigIV}}$ using IMPACT

The recent method proposed to interpret relaxation data from intrinsically disordered proteins, IMPACT, (see chapter 3) has to be changed from the original approach. Indeed in the particular case of Engrailed, for which we have only high-field data, it is possible to perform spectral density mapping at all the fields. In contrast, the measurements at low field are, to date, restricted to the longitudinal relaxation rates. Thus any fit of the spectral density function has to be done through the theoretical expression of the R_1 and not directly on the spectral density function.

Here, we show preliminary results on the interpretation of high-field relaxation data obtained on $\text{Art}^{\text{LigIV}}$. The data used for the following analysis were carried out with only 500 MHz, 800 MHz and 950 MHz datasets and the program is identical to the one used in Engrailed. In particular it uses spectral density mapping to perform the fit (data not showed here). The only difference with the work on Engrailed is the selection criteria, which is simply a reduced χ^2 . The optimization of the number and the range of correlation times leads to $n = 5$ and $[\tau_{\min}, \tau_{\max}] = [35 \text{ ps}, 35 \text{ ns}]$ (Figure 4.29).

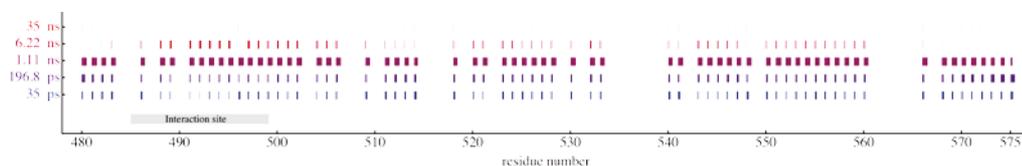


Figure 4.29: Graphical representations of a preliminary IMPACT analysis of the spectral density function in $\text{Art}^{\text{LigIV}}$.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

Note the very low coefficient for the Lorentzian function with the longest correlation time (35 ns) while the dominant timescale is around 1 ns (similarly to what was found in Engrailed). This led us to the conclusion that the ratio $\tau_{\max}/\tau_{\min} = 10^3$ used in our study of Engrailed is not the optimum for Art^{LigIV} probably because of more prevalent disorder. This means that for the first “IMPACT step” of the combined IMPACT-ICARUS approach the ratio τ_{\max}/τ_{\min} has to be optimized. We may notice that the interaction site with ^{DBD}LigIV is characterized by the presence of more supra-ns motions than the rest of our construct, which may suggest the presence of some residual structure. Interestingly, residues at the N-terminal and C-terminal edges of the protein exhibit more prominent motions on sub-ns timescales, as was seen in Engrailed.

5.4 ¹⁵N relaxation rate measurements on 90% Art^{LigIV} apo and 10% holo at high fields

Thus we were interested to compare the relaxation occurring in the absence or in the presence of its partner, ^{DBD}LigIV. Because all the resonances of the interaction site disappeared by adding ^{DBD}LigIV, we first decided to measure relaxation in presence of 0.1 molar equivalent of ^{DBD}LigIV (Figure 4.30). We measured a series of relaxation rates at 800 MHz including the same rates in the apo form: R_1 , R_2^{cpmg} , R_2^{echo} , NOE, η_{xy} and η_z (Figure 4.27).

As expected, most of the rates are similar in the absence and in the presence of ^{DBD}LigIV outside of the heart interaction site (485-500) and its C-terminal edge (500-510). This confirms the results from chemical shift mapping and isothermal titration calorimetry (see above) suggesting that the interaction site is limited to residues 485-500.

One has to be careful about the measured rates that correspond to an average of 90% of Art^{LigIV} and 10% of a complex Art^{LigIV}: ^{DBD}LigIV in the 485-510 region. In particular, measured rates are not necessarily averages between the free and bound form but effective rates that result from the spin dynamics during each experiment. The measured longitudinal relaxation rates (R_1) from residue 485 to 510, are lower than in the apo form alone. The size of the complex (~ 40 kDa) makes that R_1 rates in the bound form are expected to be very low (Figure 4.30-a). However, the measured rates drop by more than 10% for some residues (492 - 493) at the heart of the interaction region (residues 485-500), which challenges a naïve treatment of the results. Interestingly, NOE values do not vary upon the addition of ^{DBD}LigIV (Figure 4.30-b). This observation is somehow unexpected if we consider that NOE ratios in the bound form should be much higher than in the apo form for tightly bound residues (residues 485-500). This observation can be rationalized in the following manner: the typical timescale of exchange $1/k_{ex} \sim 17$ ms is longer than the duration of one element of the

Cyril Charlier

saturation period in the NOE experiment $\tau_{\text{NOE}} = 11$ ms. We can approximate that the evolution of the nitrogen-15 longitudinal polarization during the saturation experiment is a series of evolutions during saturations τ_{NOE} with no averaging of relaxation rates during τ_{NOE} . Under these conditions, nitrogen-15 longitudinal polarization will evolve towards the saturated steady state of the *apo* state for 90% of the saturation elements, with the longitudinal relaxation rate of the *apo* state, which is large and, for 10% of the time, towards the saturated steady state of the *holo* state, with the longitudinal relaxation rate of the *holo* state, which is small. The difference of the rates of evolution is expected to skew the averaging between the two steady states towards the one of the *apo* state, which is what we observe. We will investigate this effect in more detail, with simulations and experiments.

Transverse relaxation rates are shown in Figure 4.30-e. Many residues of the interaction site are characterized by high transverse relaxation rates. On the contrary, transverse relaxation rates for most of the protein are similar to those of the *apo* state. Both R_2^{cpmg} and R_2^{echo} , are enhanced in the presence of $^{\text{DBD}}\text{LigIV}$. As expected from the relaxation dispersion study, R_2 rates measured under a single echo are significantly higher than those measured under CPMG train. Residues with high R_2^{cpmg} relaxation rates are those for which high relaxation rates in the *holo* state were found in the RD analysis, where we suspect the existence of a second, fast, exchange process. Residues outside of the direct interaction site, in particular 500-523 and 552 show small, but significant enhancements of their R_2^{echo} rates as compared to the R_2 rates in the *apo* state. These changes correlate reasonably well with the presence of chemical shift perturbations under binding (see Figure 4.22). These changes, which are too small to be observable in dispersion profiles (data not shown), suggest some degree of reorganization of the conformational space upon binding that reach far from the core of the binding site.

The longitudinal (η_z) and transverse (η_{xy}) cross-correlated relaxation rates due to the fluctuations of ^{15}N CSA and dipole-dipole interaction of [didier](#) N-H pair are slightly similar in presence and absence of $^{\text{DBD}}\text{LigIV}$ (Figure 4.30-c-d). Fast transverse relaxation in the bound form likely prevents the build-up of polarization under cross-relaxation at the η_{xy} rate of the *holo* state. In addition, these rates could be affected by chemical exchange with the solvent and so have significant error bars.

Chapter 4: Interaction and dynamics of the C-terminal region of Artemis

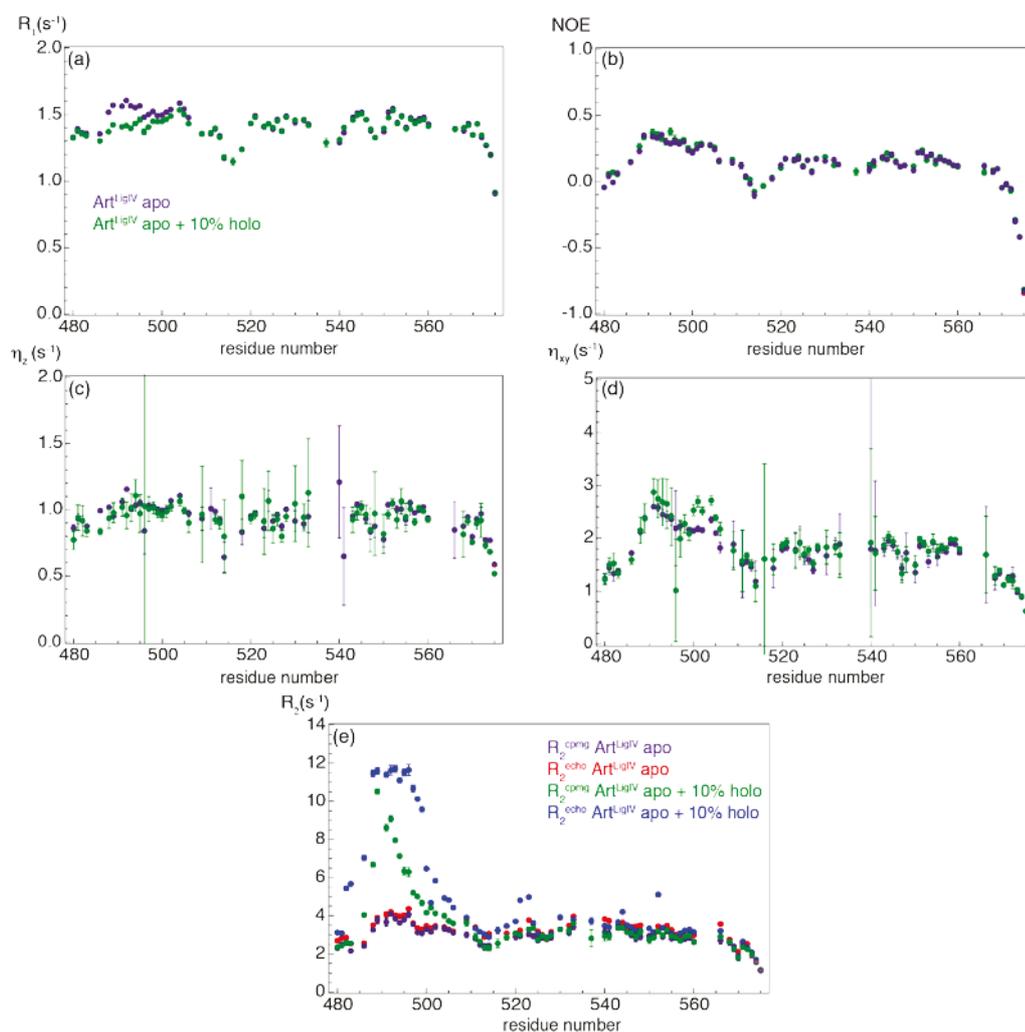


Figure 4.30: Backbone ^{15}N relaxation measurements on $\text{Art}^{\text{LigIV}}$ in absence (purple & red) and in presence (green & blue) of 0.1 equivalent of $^{\text{c}}\text{DBD}^{\text{LigIV}}$ at 800 MHz; (a) Longitudinal relaxation rates R_1 ; (b) steady-state $^{15}\text{N}\{-^1\text{H}\}$ NOE ratios; (c) CSA-DD longitudinal cross-relaxation rates (η_z); (d) CSA-DD transverse cross-relaxation rates (η_{xy}); (e) Transverse relaxation under a CPMG train (R_2^{cpmg}) and under a simple echo (R_2^{echo}).

5.5 ^{15}N relaxation rate measurements fully saturated $\text{Art}^{\text{LigIV}}$ high fields

In addition we tried to measure relaxation rates in $^{15}\text{N}\text{-ArtLigIV}$ with 1.3 molar equivalent of unlabeled $^{\text{DBD}}\text{LigIV}$ at 500 MHz. Unfortunately due to the short life time of the sample we could only measured longitudinal relaxation rates (R_1) and Nuclear Overhauser Effect (NOE). As expected, signals of the interaction site (485-499) are missing, due to the chemical exchange broadening. (Figure 4.31) Longitudinal relaxation rates do not reveal major difference outside of the interaction site (Figure 4.31-a). In contrast with what we

observed with only 10% of $^{DBD}LigIV$, NOE ratios are significantly higher in the complex at the edges of the interaction site (Figure 4.31-b). Overall, these results show significant changes in spectral density functions (Figure 4.32) for residues at the N-terminus (residues 480–484) and at the C-terminus (500–520) of the core of the binding site. At the C-terminus (residues 500–507) the spectral density function at the Larmor frequency of nitrogen-15 decreases as reported by the longitudinal relaxation rates. The opposite can be observed at the N-terminus, where sub-ns motions are likely depleted in the complex. The drop in the spectral density at high frequency in the complex is more significant, both at the N- and C-terminal parts of the interaction region. In particular, the decrease of fast motions at the C-terminus of the interaction site reaches at least residue 515, which is at a distance about twice the persistence length (ca. 7 residues in disordered proteins) of the interaction site.

Commentaire [FF6]: approx

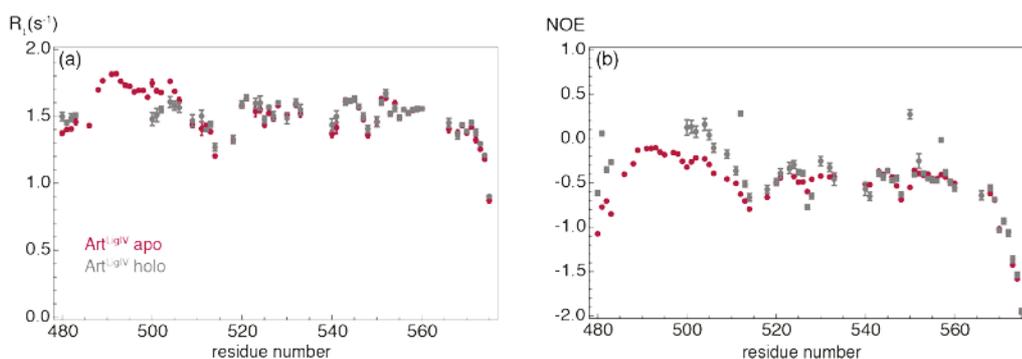


Figure 4.31: Backbone ^{15}N relaxation measurements on Art^{LigIV} with 0% (burgundy) or 100% (gray) $^{DBD}LigIV$ at 500 MHz; (a) Longitudinal relaxation rates R_1 ; (b) Steady-state ^{15}N - $\{^1H\}$ NOE ratios;

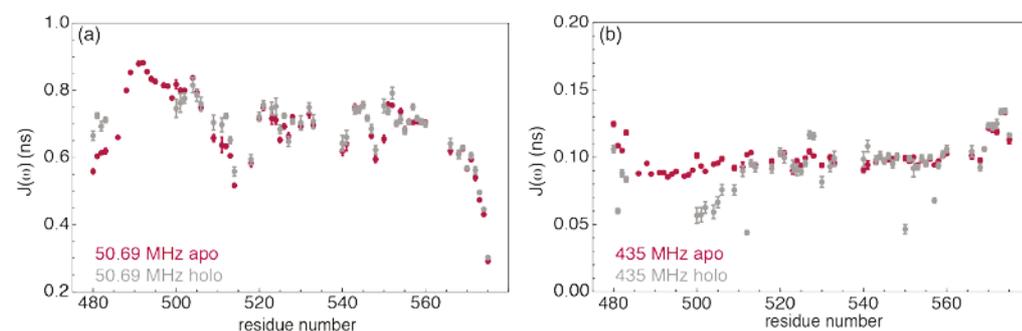


Figure 4.32: Spectral density functions for backbone NH vectors for Art^{LigIV} with 0% (burgundy) or 100% (gray) $^{DBD}LigIV$ at 500 MHz. (a) Effective spectral density near the proton Larmor frequency $J(0.87\omega_H)$ (ns). (b) Spectral density at the Larmor frequency of nitrogen-15, $J(\omega_N)$ (ns).

5.6 Low-field longitudinal ^{15}N relaxation measurements on $\text{Art}^{\text{LigIV}}$ with high-resolution relaxometry

As we earlier describe with the example of ubiquitin, high-resolution relaxometry provide complementary information on the spectral density function⁸³. We applied this idea to $\text{Art}^{\text{LigIV}}$ by measuring a series of longitudinal relaxation rates at low field using high-resolution relaxometry^{83,141}. The shuttle used for R_1 measurements at low fields, was designed by C.-Y. Chou and coworkers¹⁴¹ and constructed by C.-Y. Chou in collaboration with the teams of T.-H. Huang (Academia Sinica, Taiwan) and D. Sakellariou (CEA, Saclay). This device has a much higher sensitivity than the Bruker shuttle (Chapter 2). As a conventional 5-mm tube can be mounted on this shuttle, it can be adapted on any liquid-state 5 mm probe, including a cryoprobe. The expected sensitivity on a 700 MHz spectrometer equipped with a cryoprobe is at least an order of magnitude better than the sensitivity of the Bruker shuttle. We used such a system for measuring longitudinal relaxation rates on a sample $^{15}\text{N}\text{-Art}^{\text{LigIV}}$ at a concentration of 600 μM (Figure 4.33).

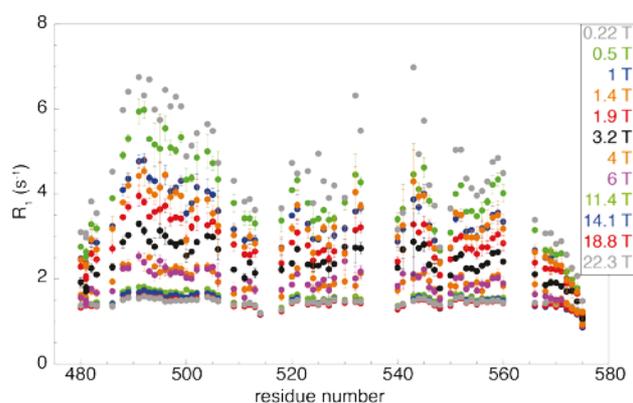


Figure 4.33: Experimental longitudinal relaxation rates $R_1 = 1/T_1$ of ^{15}N in backbone amide groups of $^{15}\text{N}\text{-Art}^{\text{LigIV}}$ as a function of magnetic fields $0.22 \leq B_0^{\text{low}} \leq 6$ T and $11.7 \leq B_0^{\text{high}} \leq 22.3$ T. From bottom to top: $B_0^{\text{high}} = 22.3, 18.8, 14.1,$ and 11.4 T; $B_0^{\text{low}} = 6.0, 4.0, 3.2, 1.9, 1.4, 1.0, 0.5,$ and 0.22 T. Note that all rates increase with decreasing field. The lower B_0^{low} , the greater the variations of the rates along the backbone

As mentioned in the case of ubiquitin, the variations in R_1 along the sequence are more pronounced at low fields for $\text{Art}^{\text{LigIV}}$. The interpretation these data in terms of microdynamics parameters is not as straightforward as in ubiquitin. Indeed, as indicated in the case of Engrailed (see chapter 3) relaxation data cannot be analyzed using the basic model-free approach with a single overall diffusion tensor. Here, we need to develop an approach combining the ICARUS and IMPACT approaches as followed:

Cyril Charlier

- IMPACT approach using only high-field data to approximate properly the general evolution of spectral densities with magnetic field.
- ICARUS approach to simulate the evolution of the spin system during a relaxometry experiment in order to correct the data.
- IMPACT approach using both high-field and corrected low field relaxation rates with an optimized array of correlation times.

(As in the original approach of ICARUS, the last two steps should work as a loop)

At the time of writing only preliminary simulations have been carried out and will have to be completed to determine accurately the site-specific distribution of correlation times in Art^{LigIV}.

6. Conclusion

Artemis, a critical factor for the maintenance of the genome and adaptive immunity, contains a long disordered C-terminal region of ~ 300 amino acids. This C-terminal region interacts with the DNA Binding Domain of Ligase IV. The aim of this work was to study this interaction with NMR and to use Artemis to develop methodological tools based on nuclear spin relaxation measurements in IDPs.

Overall this study demonstrates the presence of five additional residues in the C-terminal of the interaction site. The characterization of the kinetics and the thermodynamics of the binding sheds light on the mechanism of the interaction. *In vitro* and *in vivo* assays would be useful to study the effect of these additional residues on the efficiency of Artemis to repair DNA. Preliminary studies on the dynamics of Artemis in presence or in absence of the DNA-Binding Domain of Ligase IV indicate variations in pico- and nanosecond timescale motions in the vicinity of the interaction site. In addition, the conformational space is clearly modified in the C-terminal region of the interaction site. Finally, the measurements of extensive longitudinal relaxation rates using high-resolution relaxometry should provide access to a quantitative characterization of nanosecond timescale motions in Artemis.

Bibliography

- (1) Fischer, E. *Berichte Dtsch. Chem. Ges.* **1894**, 27 (3), 2985.
- (2) Anfinsen, C. B. *Science* **1973**, 181 (4096), 223.
- (3) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, 181 (4610), 662.
- (4) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature* **1960**, 185 (4711), 422.
- (5) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, 185 (4711), 416.
- (6) Ban, N. *Science* **2000**, 289 (5481), 905.
- (7) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. *Nature* **2013**, 497 (7451), 643.
- (8) Greber, B. J.; Boehringer, D.; Leibundgut, M.; Bieri, P.; Leitner, A.; Schmitz, N.; Aebersold, R.; Ban, N. *Nature* **2014**.
- (9) Zhao, M.; Wu, S.; Zhou, Q.; Vivona, S.; Cipriano, D. J.; Cheng, Y.; Brunger, A. T. *Nature* **2015**, 518 (7537), 61.
- (10) Wüthrich, K. *NMR of proteins and nucleic acids*; The George Fisher Baker non-resident lectureship in chemistry at Cornell University; Wiley: New York, 1986.
- (11) Shahid, S. A.; Bardiaux, B.; Franks, W. T.; Krabben, L.; Habeck, M.; van Rossum, B.-J.; Linke, D. *Nat. Methods* **2012**, 9 (12), 1212.
- (12) Wang, S.; Munro, R. A.; Shi, L.; Kawamura, I.; Okitsu, T.; Wada, A.; Kim, S.-Y.; Jung, K.-H.; Brown, L. S.; Ladizhansky, V. *Nat. Methods* **2013**, 10 (10), 1007.
- (13) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, 124 (34), 10025.
- (14) Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J. *J. Biomol. NMR* **2013**, 57 (3), 251.
- (15) Sattler, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, 34 (2), 93.
- (16) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry (Mosc.)* **1989**, 28 (23), 8972.
- (17) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, 27, 357.
- (18) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, 125 (45), 13868.
- (19) Crublet, E.; Kerfah, R.; Mas, G.; Noirclerc-Savoie, M.; Lantéz, V.; Vernet, T.; Boisbouvier, J. In *Structural Genomics*; Chen, Y. W., Ed.; Humana Press: Totowa, NJ, 2014; Vol. 1091, pp 229–244.
- (20) Kerfah, R.; Plevin, M. J.; Pessey, O.; Hamelin, O.; Gans, P.; Boisbouvier, J. *J. Biomol. NMR* **2015**, 61 (1), 73.
- (21) Kwan, A. H.; Mobli, M.; Gooley, P. R.; King, G. F.; Mackay, J. P. *FEBS J.* **2011**, 278 (5), 687.
- (22) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, 293 (2), 321.
- (23) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **1998**, 437.
- (24) Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2010**, 1804 (4), 996.
- (25) Xue, B.; Dunker, A. K.; Uversky, V. N. *J. Biomol. Struct. Dyn.* **2012**, 30 (2), 137.
- (26) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, 37 (1), 215.
- (27) Uversky, V. N. *Int. J. Biochem. Cell Biol.* **2011**, 43 (8), 1090.
- (28) Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biochem.* **2014**, 83 (1), 553.

- (29) Yegambaram, K.; Bulloch, E. M. M.; Kingston, R. L. *Protein Sci. Publ. Protein Soc.* **2013**, *22* (11), 1502.
- (30) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120* (27), 6836.
- (31) Song, J.; Lee, M. S.; Carlberg, I.; Vener, A. V.; Markley, J. L. *Biochemistry (Mosc.)* **2006**, *45* (51), 15633.
- (32) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. *Biophys. J.* **2007**, *92* (5), 1439.
- (33) He, B.; Wang, K.; Liu, Y.; Xue, B.; Uversky, V. N.; Dunker, A. K. *Cell Res.* **2009**, *19* (8), 929.
- (34) Brucale, M.; Schuler, B.; Samori, B. *Chem. Rev.* **2014**, *114* (6), 3281.
- (35) Baker, D. *Nature* **2000**, *405* (6782), 39.
- (36) Baker, D. *Biochem. Soc. Trans.* **2014**, *42* (2), 225.
- (37) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2010**, *1804* (6), 1231.
- (38) Romero, null; Obradovic, null; Dunker, null. *Genome Inform. Workshop Genome Inform.* **1997**, *8*, 110.
- (39) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337* (3), 635.
- (40) Dunker, A. K.; Garner, E.; Guillot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **1998**, 473.
- (41) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* **2001**, *19* (1), 26.
- (42) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42* (1), 38.
- (43) Uversky, V. N. *Protein Sci. Publ. Protein Soc.* **2002**, *11* (4), 739.
- (44) Garner, null; Cannon, null; Romero, null; Obradovic, null; Dunker, null. *Genome Inform. Workshop Genome Inform.* **1998**, *9*, 201.
- (45) Dyson, H. J.; Wright, P. E. *Adv. Protein Chem.* **2002**, *62*, 311.
- (46) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. *Biochemistry (Mosc.)* **2002**, *41* (21), 6573.
- (47) Wright, P. E.; Dyson, H. J. *Nat. Rev. Mol. Cell Biol.* **2014**, *16* (1), 18.
- (48) Tompa, P. *FEBS Lett.* **2005**, *579* (15), 3346.
- (49) Tompa, P. *Structure and function of intrinsically disordered proteins*; Chapman & Hall/CRC Press: Boca Raton, 2010.
- (50) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15* (1), 35.
- (51) Peti, W.; Nairn, A. C.; Page, R. *Curr. Phys. Chem.* **2012**, *2* (1), 107.
- (52) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323* (3), 573.
- (53) Xu, W.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. *Mol. Cell* **1999**, *3* (5), 629.
- (54) Tompa, P.; Csermely, P. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2004**, *18* (11), 1169.
- (55) Tompa, P.; Kovacs, D. *Biochem. Cell Biol. Biochim. Biol. Cell.* **2010**, *88* (2), 167.
- (56) Bernadó, P.; Svergun, D. I. In *Intrinsically Disordered Protein Analysis*; Uversky, V. N., Dunker, A. K., Eds.; Springer New York: New York, NY, 2012; pp 107–122.
- (57) Receveur-Brechot, V.; Durand, D. *Curr. Protein Pept. Sci.* **2012**, *13* (1), 55.
- (58) Ferreon, A. C. M.; Moran, C. R.; Gambin, Y.; Deniz, A. A. In *Methods in Enzymology*; Elsevier, 2010; Vol. 472, pp 179–204.

- (59) Neri, D.; Wider, G.; Wüthrich, K. *FEBS Lett.* **1992**, *303* (2-3), 129.
- (60) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257* (5076), 1559.
- (61) Neri, D.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (10), 4397.
- (62) Arcus, V. L.; Vuilleumier, S.; Freund, S. M.; Bycroft, M.; Fersht, A. R. *Proc. Natl. Acad. Sci.* **1994**, *91* (20), 9412.
- (63) Arcus, V. L.; Vuilleumier, S.; Freund, S. M.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254* (2), 305.
- (64) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263* (2), 369.
- (65) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry (Mosc.)* **1997**, *36* (29), 8977.
- (66) Shortle, D. R. *Curr. Opin. Struct. Biol.* **1996**, *6* (1), 24.
- (67) Bermel, W.; Bertini, I.; Felli, I.; Piccioli, M.; Pierattelli, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48* (1), 25.
- (68) Bermel, W.; Bertini, I.; Csizmok, V.; Felli, I. C.; Pierattelli, R.; Tompa, P. *J. Magn. Reson.* **2009**, *198* (2), 275.
- (69) Felli, I. C.; Pierattelli, R. *J. Magn. Reson.* **2014**, *241*, 115.
- (70) Nováček, J.; Židek, L.; Sklenář, V. *J. Magn. Reson.* **2014**, *241*, 41.
- (71) Zawadzka-Kazimierczuk, A.; Koźmiński, W.; Šanderová, H.; Krásný, L. *J. Biomol. NMR* **2012**, *52* (4), 329.
- (72) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104* (8), 3607.
- (73) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. *Molecules* **2013**, *18* (9), 10802.
- (74) Konrat, R. *J. Magn. Reson.* **2014**, *241*, 74.
- (75) *Protein NMR spectroscopy: practical techniques and applications*; Lian, L.-Y., Roberts, G. C. K., Eds.; Wiley: Chichester, West Sussex, 2011.
- (76) Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. *Curr. Opin. Struct. Biol.* **2013**, *23* (3), 426.
- (77) Rezaei-Ghaleh, N.; Blackledge, M.; Zweckstetter, M. *ChemBioChem* **2012**, *13* (7), 930.
- (78) Kjaergaard, M.; Poulsen, F. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *60*, 42.
- (79) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132* (24), 8407.
- (80) Salmon, L.; Bouvignies, G.; Markwick, P.; Blackledge, M. *Biochemistry (Mosc.)* **2011**, *50* (14), 2735.
- (81) Lawrence, C. W.; Showalter, S. A. *J. Phys. Chem. Lett.* **2012**, *3* (10), 1409.
- (82) Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D. F.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2015**, *137* (3), 1220.
- (83) Charlier, C.; Khan, S. N.; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F. *J. Am. Chem. Soc.* **2013**, *135* (49), 18665.
- (84) *Protein NMR spectroscopy: principles and practice*, 2nd ed.; Cavanagh, J., Ed.; Academic Press: Amsterdam ; Boston, 2007.
- (85) Korzhnev, D. M.; Billeter, M.; Arseniev, A. S.; Orekhov, V. Y. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38* (3), 197.
- (86) Bloch, F. *Phys. Rev.* **1946**, *70* (7-8), 460.
- (87) Solomon, I. *Phys. Rev.* **1955**, *99* (2), 559.
- (88) Keeler, J. *Understanding NMR spectroscopy*, 2nd ed.; John Wiley and Sons: Chichester, U.K., 2010.
- (89) Wangsness, R. K.; Bloch, F. *Phys. Rev.* **1953**, *89* (4), 728.
- (90) Redfield, A. G. *IBM J. Res. Dev.* **1957**, *1* (1), 19.
- (91) Redfield, A. G. *Advances In Magnetic Resonance*. 1965, pp 1–32.

- (92) Goldman, M. *J. Magn. Reson.* **2001**, *149* (2), 160.
- (93) Abragam, A. *The principles of nuclear magnetism*; Clarendon Press ; Oxford University Press: Oxford [Oxfordshire]; New York, 1983.
- (94) Hubbard, P. S. *Phys. Rev.* **1969**, *180* (1), 319.
- (95) Halle, B.; Andersson, T.; Forsen, S.; Lindman, B.; Lindman, B. *J. Am. Chem. Soc.* **1981**, *103* (3), 500.
- (96) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104* (17), 4546.
- (97) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104* (17), 4559.
- (98) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112* (12), 4989.
- (99) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120* (38), 9870.
- (100) Meirovitch, E.; Shapiro, Y. E.; Polimeno, A.; Freed, J. H. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56* (4), 360.
- (101) Peng, J. W.; Wagner, G. *Biochemistry (Mosc.)* **1992**, *31* (36), 8571.
- (102) Farrow, N.; Zhang, O.; Szabo, A.; Torchia, D.; Kay, L. *J. Biomol. NMR* **1995**, *6* (2).
- (103) Saitô, H.; Ando, I.; Ramamoorthy, A. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57* (2), 181.
- (104) Palmer, A. G.; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204.
- (105) Carr, H. Y.; Purcell, E. M. *Phys. Rev.* **1954**, *94* (3), 630.
- (106) Meiboom, S.; Gill, D. *Rev. Sci. Instrum.* **1958**, *29* (8), 688.
- (107) Palmer, A. G. *Chem. Rev.* **2004**, *104* (8), 3623.
- (108) Loria, J. P.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1999**, *121* (10), 2331.
- (109) Hill, R. B.; Bracken, C.; DeGrado, W. F.; Palmer, A. G. *J. Am. Chem. Soc.* **2000**, *122* (47), 11610.
- (110) Neudecker, P.; Lundström, P.; Kay, L. E. *Biophys. J.* **2009**, *96* (6), 2045.
- (111) Wang, C.; Palmer, A. G. *Magn. Reson. Chem.* **2003**, *41* (10), 866.
- (112) Orekhov, V. Y.; Korzhnev, D. M.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126* (6), 1886.
- (113) Korzhnev, D. M.; Neudecker, P.; Mittermaier, A.; Orekhov, V. Y.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127* (44), 15602.
- (114) Korzhnev, D. M.; Kay, L. E. *Acc. Chem. Res.* **2008**, *41* (3), 442.
- (115) Carver, J. ; Richards, R. . *J. Magn. Reson.* **1969** *1972*, *6* (1), 89.
- (116) Ishima, R.; Torchia, D. A. *J. Biomol. NMR* **1999**, *14* (4), 369.
- (117) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. . *J. Magn. Reson.* **1969** *1992*, *97* (2), 359.
- (118) Ferrage, F. *Methods Mol. Biol. Clifton NJ* **2012**, *831*, 141.
- (119) Cavanagh, J.; Palmer, A. G.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1969** *1991*, *91* (2), 429.
- (120) Palmer, A. G.; Skelton, N. J.; Chazin, W. J.; Wright, P. E.; Rance, M. *Mol. Phys.* **1992**, *75* (3), 699.
- (121) Skelton, N. J.; Palmer, A. G.; Akke, M.; Kordel, J.; Rance, M.; Chazin, W. J. *J. Magn. Reson. B* **1993**, *102* (3), 253.
- (122) Palmer, A. G. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129.
- (123) Burum, D. ; Ernst, R. . *J. Magn. Reson.* **1969** *1980*, *39* (1), 163.
- (124) Morris, G. A.; Freeman, R. *J. Am. Chem. Soc.* **1979**, *101* (3), 760.
- (125) Pelupessy, P.; Ferrage, F.; Bodenhausen, G. *J. Chem. Phys.* **2007**, *126* (13), 134508.
- (126) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1969** *1989*, *85* (2), 393.
- (127) Ferrage, F.; Cowburn, D.; Ghose, R. *J. Am. Chem. Soc.* **2009**, *131* (17), 6048.
- (128) Ferrage, F.; Reichel, A.; Battacharya, S.; Cowburn, D.; Ghose, R. *J. Magn. Reson. San Diego Calif 1997* **2010**, *207* (2), 294.

- (129) Tjandra, N.; Szabo, A.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118* (29), 6986.
- (130) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. *J. Magn. Reson. San Diego Calif* **1997** **2003**, *161* (2), 258.
- (131) Kroenke, C. D.; Loria, J. P.; Lee, L. K.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1998**, *120* (31), 7905.
- (132) Redfield, A. G. *Magn. Reson. Chem.* **2003**, *41* (10), 753.
- (133) Noack, F. *Prog. Nucl. Magn. Reson. Spectrosc.* **1986**, *18* (3), 171.
- (134) Modig, K.; Halle, B. *J. Am. Chem. Soc.* **2002**, *124* (40), 12031.
- (135) Kimmich, R.; Anorado, E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *44* (3-4), 257.
- (136) Luchinat, C.; Parigi, G. *J. Am. Chem. Soc.* **2007**, *129* (5), 1055.
- (137) Wagner, S.; Dinesen, T. R. J.; Rayner, T.; Bryant, R. G. *J. Magn. Reson.* **1999**, *140* (1), 172.
- (138) Victor, K.; Kavolius, V.; Bryant, R. G. *J. Magn. Reson.* **2004**, *171* (2), 253.
- (139) Kerwood, D. J.; Bolton, P. H. *J. Magn. Reson.* **1969** **1987**, *75* (1), 142.
- (140) Redfield, A. G. *J. Biomol. NMR* **2012**, *52* (2), 159.
- (141) Chou, C.-Y.; Chu, M.; Chang, C.-F.; Huang, T. *J. Magn. Reson.* **2012**, *214*, 302.
- (142) Roberts, M. F.; Redfield, A. G. *Proc. Natl. Acad. Sci.* **2004**, *101* (49), 17066.
- (143) Sivanandam, V. N.; Cai, J.; Redfield, A. G.; Roberts, M. F. *J. Am. Chem. Soc.* **2009**, *131* (10), 3420.
- (144) Clarkson, M. W.; Lei, M.; Eisenmesser, E. Z.; Labeikovsky, W.; Redfield, A.; Kern, D. *J. Biomol. NMR* **2009**, *45* (1-2), 217.
- (145) Gradziel, C. S.; Wang, Y.; Stec, B.; Redfield, A. G.; Roberts, M. F. *Biochemistry (Mosc.)* **2014**, *53* (3), 462.
- (146) Pickart, C. M.; Eddins, M. J. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2004**, *1695* (1-3), 55.
- (147) Dikic, I.; Wakatsuki, S.; Walters, K. J. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (10), 659.
- (148) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120* (27), 6836.
- (149) Vijay-Kumar, S.; Bugg, C. E.; Wilkinson, K. D.; Cook, W. J. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (11), 3582.
- (150) Vijay-kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194* (3), 531.
- (151) Di Stefano, D. L.; Wand, A. J. *Biochemistry (Mosc.)* **1987**, *26* (23), 7272.
- (152) Weber, P. L.; Brown, S. C.; Mueller, L. *Biochemistry (Mosc.)* **1987**, *26* (23), 7282.
- (153) Wang, A.; Grzesiek, S.; Tschudin, R.; Lodi, P.; Bax, A. *J. Biomol. NMR* **1995**, *5* (4).
- (154) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117* (50), 12562.
- (155) Salmon, L.; Bouvignies, G.; Markwick, P.; Lakomek, N.; Showalter, S.; Li, D.-W.; Walter, K.; Griesinger, C.; Brüschweiler, R.; Blackledge, M. *Angew. Chem. Int. Ed.* **2009**, *48* (23), 4154.
- (156) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129* (31), 9799.
- (157) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131* (46), 16968.
- (158) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3* (3), 961.
- (159) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmüller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angew. Chem. Int. Ed.* **2011**, *50* (48), 11437.
- (160) Farès, C.; Lakomek, N.-A.; Walter, K. F. A.; Frank, B. T. C.; Meiler, J.; Becker, S.; Griesinger, C. *J. Biomol. NMR* **2009**, *45* (1-2), 23.
- (161) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker,

- S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320* (5882), 1471.
- (162) Massi, F. *Protein Sci.* **2005**, *14* (3), 735.
- (163) Zhang, Y.; Zhou, L.; Rouge, L.; Phillips, A. H.; Lam, C.; Liu, P.; Sandoval, W.; Helgason, E.; Murray, J. M.; Wertz, I. E.; Corn, J. E. *Nat. Chem. Biol.* **2013**, *9* (1), 51.
- (164) Reese, M.; Türke, M.-T.; Tkach, I.; Parigi, G.; Luchinat, C.; Marquardsen, T.; Tavernier, A.; Höfer, P.; Engelke, F.; Griesinger, C.; Bennati, M. *J. Am. Chem. Soc.* **2009**, *131* (42), 15086.
- (165) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. *J. Magn. Reson.* **1992**, *97* (2), 359.
- (166) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3), 277.
- (167) Salvi, N.; Ulzega, S.; Ferrage, F.; Bodenhausen, G. *J. Am. Chem. Soc.* **2012**, *134* (5), 2481.
- (168) Walker, O.; Varadan, R.; Fushman, D. *J. Magn. Reson.* **2004**, *168* (2), 336.
- (169) Fushman, D.; Cahill, S.; Cowburn, D. *J. Mol. Biol.* **1997**, *266* (1), 173.
- (170) Briggman, K. B.; Tolman, J. R. *J. Am. Chem. Soc.* **2003**, *125* (34), 10164.
- (171) Lakomek, N.-A.; Walter, K. F. A.; Farès, C.; Lange, O. F.; de Groot, B. L.; Grubmüller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41* (3), 139.
- (172) Salmon, L.; Bouvignies, G.; Markwick, P.; Lakomek, N.; Showalter, S.; Li, D.-W.; Walter, K.; Griesinger, C.; Brüschweiler, R.; Blackledge, M. *Angew. Chem. Int. Ed.* **2009**, *48* (23), 4154.
- (173) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ø.; Xu, H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112* (19), 6155.
- (174) Clore, G. M.; Bax, A.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry (Mosc.)* **1990**, *29* (35), 8172.
- (175) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133* (27), 10336.
- (176) Long, D.; Brüschweiler, R. *PLoS Comput. Biol.* **2011**, *7* (4), e1002035.
- (177) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci.* **2005**, *102* (39), 13885.
- (178) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450* (7171), 913.
- (179) Ban, D.; Gossert, A. D.; Giller, K.; Becker, S.; Griesinger, C.; Lee, D. *J. Magn. Reson.* **2012**, *221*, 1.
- (180) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121* (37), 8671.
- (181) Tugarinov, V.; Liang, Z.; Shapiro, Y. E.; Freed, J. H.; Meirovitch, E. *J. Am. Chem. Soc.* **2001**, *123* (13), 3055.
- (182) Wong, V.; Case, D. A.; Szabo, A. *Proc. Natl. Acad. Sci.* **2009**, *106* (27), 11016.
- (183) Ryabov, Y.; Clore, G. M.; Schwieters, C. D. *J. Chem. Phys.* **2012**, *136* (3), 034108.
- (184) Peng, J. W.; Wagner, G. *Biochemistry (Mosc.)* **1992**, *31* (36), 8571.
- (185) Bussell, R.; Eliezer, D. *J. Biol. Chem.* **2001**, *276* (49), 45996.
- (186) Ochsenbein, F.; Guerois, R.; Neumann, J. M.; Sanson, A.; Guittet, E.; van Heijenoort, C. *J. Biomol. NMR* **2001**, *19* (1), 3.
- (187) Ishima, R.; Nagayama, K. *J. Magn. Reson. B* **1995**, *108* (1), 73.
- (188) Ishima, R.; Nagayama, K. *Biochemistry (Mosc.)* **1995**, *34* (10), 3162.
- (189) Alexandrescu, A. T.; Shortle, D. *J. Mol. Biol.* **1994**, *242* (4), 527.
- (190) Cole, K. S.; Cole, R. H. *J. Chem. Phys.* **1941**, *9* (4), 341.

- (191) Buevich, A. V.; Shinde, U. P.; Inouye, M.; Baum, J. J. *Biomol. NMR* **2001**, *20* (3), 233.
- (192) Ochsenbein, F. *Protein Sci.* **2002**, *11* (4), 957.
- (193) Modig, K.; Poulsen, F. M. *J. Biomol. NMR* **2008**, *42* (3), 163.
- (194) Banerjee-Basu, S.; Baxevanis, A. D. *Nucleic Acids Res.* **2001**, *29* (15), 3258.
- (195) Morgan, R. *FEBS Lett.* **2006**, *580* (11), 2531.
- (196) Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O. *Protein Sci. Publ. Protein Soc.* **1994**, *3* (10), 1779.
- (197) Carlier, L.; Balayssac, S.; Cantrelle, F.-X.; Khemtémourian, L.; Chassaing, G.; Joliot, A.; Lequin, O. *Biophys. J.* **2013**, *105* (3), 667.
- (198) Khan, Shahid. The Conformational Space and Dynamics of the Partially Disordered Transcription Factor Engrailed-2 Explored by Magnetic Resonance, Pierre er Marie Curie (Paris).
- (199) Augustyniak, R.; Balayssac, S.; Ferrage, F.; Bodenhausen, G.; Lequin, O. *Biomol. NMR Assign.* **2011**, *5* (2), 229.
- (200) Augustyniak, Rafal, (Prénom). Etude de la protéine partiellement désordonnée Engrailed 2 par RMN et nouvelles méthodes pour la mesure de coefficients de diffusion et l'attribution des signaux des chaines latérales de protéine, Pierre er Marie Curie (Paris).
- (201) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391* (2), 359.
- (202) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6* (3).
- (203) Prasch, S.; Schwarz, S.; Eisenmann, A.; Wöhrl, B. M.; Schweimer, K.; Rösch, P. *Biochemistry (Mosc.)* **2006**, *45* (14), 4542.
- (204) Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry (Mosc.)* **1995**, *34* (3), 868.
- (205) Ropars, V.; Bouguet-Bonnet, S.; Auguin, D.; Barthe, P.; Canet, D.; Roumestand, C. *J. Biomol. NMR* **2007**, *37* (3), 159.
- (206) Lemaster, D. M. *J. Biomol. NMR* **1995**, *6* (4), 366.
- (207) Sugiura, N. *Commun. Stat. - Theory Methods* **1978**, *7* (1), 13.
- (208) Hurvich, C. M.; Tsai, C.-L. *Biometrika* **1989**, *76* (2), 297.
- (209) Burnham, K. P.; Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*, 2. ed.; Springer: New York, NY, 2010.
- (210) Akaike, H. In *Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer New York: New York, NY, 1998; pp 199–213.
- (211) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124* (16), 4522.
- (212) Calandrini, V.; Abergel, D.; Kneller, G. R. *J. Chem. Phys.* **2010**, *133* (14), 145101.
- (213) Xue, Y.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2011**, *133* (37), 14614.
- (214) Parigi, G.; Rezaei-Ghaleh, N.; Giachetti, A.; Becker, S.; Fernandez, C.; Blackledge, M.; Griesinger, C.; Zweckstetter, M.; Luchinat, C. *J. Am. Chem. Soc.* **2014**, *136* (46), 16201.
- (215) Hall, J. B.; Fushman, D. *J. Am. Chem. Soc.* **2006**, *128* (24), 7855.
- (216) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-R.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinforma. Oxf. Engl.* **2012**, *28* (11), 1463.
- (217) Market, E.; Papavasiliou, F. N. *PLoS Biol.* **2003**, *1* (1), e6.
- (218) Harris, L. J.; Larson, S. B.; Hasel, K. W.; McPherson, A. *Biochemistry (Mosc.)* **1997**, *36* (7), 1581.
- (219) Schatz, D. G.; Ji, Y. *Nat. Rev. Immunol.* **2011**, *11* (4), 251.
- (220) Mansilla-Soto, J.; Cortes, P. *J. Exp. Med.* **2003**, *197* (5), 543.
- (221) Shrivastav, M.; De Haro, L. P.; Nickoloff, J. A. *Cell Res.* **2008**, *18* (1), 134.
- (222) Sancar, A.; Lindsey-Boltz, L. A.; Unsal-Kaçmaz, K.; Linn, S. *Annu. Rev. Biochem.* **2004**, *73*, 39.

- (223) Brandsma, I.; Gent, D. C. *Genome Integr.* **2012**, *3* (1), 9.
- (224) Ciccia, A.; Elledge, S. J. *Mol. Cell* **2010**, *40* (2), 179.
- (225) Lieber, M. R. *Annu. Rev. Biochem.* **2010**, *79* (1), 181.
- (226) Lieber, M. R. *J. Biol. Chem.* **2008**, *283* (1), 1.
- (227) Noordzij, J. G.; Verkaik, N. S.; van der Burg, M.; van Veelen, L. R.; de Bruin-Versteeg, S.; Wiegant, W.; Vossen, J. M. J. J.; Weemaes, C. M. R.; de Groot, R.; Zdzienicka, M. Z.; van Gent, D. C.; van Dongen, J. J. M. *Blood* **2003**, *101* (4), 1446.
- (228) Moshous, D.; Callebaut, I.; de Chasseval, R.; Corneo, B.; Cavazzana-Calvo, M.; Le Deist, F.; Tezcan, I.; Sanal, O.; Bertrand, Y.; Philippe, N.; Fischer, A.; de Villartay, J. P. *Cell* **2001**, *105* (2), 177.
- (229) Nicolas, N.; Moshous, D.; Cavazzana-Calvo, M.; Papadopoulo, D.; de Chasseval, R.; Le Deist, F.; Fischer, A.; de Villartay, J. P. *J. Exp. Med.* **1998**, *188* (4), 627.
- (230) Moshous, D.; Li, L.; Chasseval, R.; Philippe, N.; Jabado, N.; Cowan, M. J.; Fischer, A.; de Villartay, J. P. *Hum. Mol. Genet.* **2000**, *9* (4), 583.
- (231) Callebaut, I.; Moshous, D.; Mornon, J.-P.; de Villartay, J.-P. *Nucleic Acids Res.* **2002**, *30* (16), 3592.
- (232) Poinsignon, C. *J. Exp. Med.* **2004**, *199* (3), 315.
- (233) Wang, Z.; Fast, W.; Valentine, A. M.; Benkovic, S. J. *Curr. Opin. Chem. Biol.* **1999**, *3* (5), 614.
- (234) Ma, Y.; Pannicke, U.; Schwarz, K.; Lieber, M. R. *Cell* **2002**, *108* (6), 781.
- (235) Rooney, S.; Sekiguchi, J.; Zhu, C.; Cheng, H. L.; Manis, J.; Whitlow, S.; DeVido, J.; Foy, D.; Chaudhuri, J.; Lombard, D.; Alt, F. W. *Mol. Cell* **2002**, *10* (6), 1379.
- (236) Schlissel, M. S. *Cell* **2002**, *109* (1), 1.
- (237) Ma, Y.; Pannicke, U.; Lu, H.; Niewolik, D.; Schwarz, K.; Lieber, M. R. *J. Biol. Chem.* **2005**, *280* (40), 33839.
- (238) Geng, L.; Zhang, X.; Zheng, S.; Legerski, R. J. *Mol. Cell. Biol.* **2007**, *27* (7), 2625.
- (239) Soubeyrand, S.; Pope, L.; De Chasseval, R.; Gosselin, D.; Dong, F.; de Villartay, J.-P.; Haché, R. J. G. *J. Mol. Biol.* **2006**, *358* (5), 1200.
- (240) Zhang, X.; Succi, J.; Feng, Z.; Prithivirajsingh, S.; Story, M. D.; Legerski, R. J. *Mol. Cell. Biol.* **2004**, *24* (20), 9207.
- (241) Wang, J.; Aroumougame, A.; Loblrich, M.; Li, Y.; Chen, D.; Chen, J.; Gong, Z. *Genes Dev.* **2014**, *28* (24), 2693.
- (242) Yan, Y.; Zhang, X.; Legerski, R. J. *Cell Cycle* **2011**, *10* (23), 4098.
- (243) Wang, H.; Zhang, X.; Geng, L.; Teng, L.; Legerski, R. J. *J. Biol. Chem.* **2009**, *284* (27), 18236.
- (244) Zhang, X.; Zhu, Y.; Geng, L.; Wang, H.; Legerski, R. J. *Oncogene* **2009**, *28* (22), 2196.
- (245) Huang, Y.; Giblin, W.; Kubec, M.; Westfield, G.; St Charles, J.; Chadde, L.; Kraftson, S.; Sekiguchi, J. *J. Exp. Med.* **2009**, *206* (4), 893.
- (246) Deng, X.; Eickholt, J.; Cheng, J. *BMC Bioinformatics* **2009**, *10* (1), 436.
- (247) Goodarzi, A. A.; Yu, Y.; Riballo, E.; Douglas, P.; Walker, S. A.; Ye, R.; Härer, C.; Marchetti, C.; Morrice, N.; Jeggo, P. A.; Lees-Miller, S. P. *EMBO J.* **2006**, *25* (16), 3880.
- (248) 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.; Vezzoni, P.; Villa, A.; Aggarwal, A. K.; Cortes, P. *J. Exp. Med.* **2012**, *209* (5), 955.
- (249) Wei, Y. F.; Robins, P.; Carter, K.; Caldecott, K.; Pappin, D. J.; Yu, G. L.; Wang, R. P.; Shell, B. K.; Nash, R. A.; Schär, P. *Mol. Cell. Biol.* **1995**, *15* (6), 3206.
- (250) Martin, I. V.; MacNeill, S. A. *Genome Biol.* **2002**, *3* (4), REVIEWS3005.
- (251) Ellenberger, T.; Tomkinson, A. E. *Annu. Rev. Biochem.* **2008**, *77* (1), 313.
- (252) De Ioannes, P.; Malu, S.; Cortes, P.; Aggarwal, A. K. *Cell Rep.* **2012**, *2* (6), 1505.

- (253) Ochi, T.; Gu, X.; Blundell, T. L. *Structure* **2013**, *21* (4), 672.
- (254) Marley, J.; Lu, M.; Bracken, C. *J. Biomol. NMR* **2001**, *20* (1), 71.
- (255) Rath, A.; Glibowicka, M.; Nadeau, V. G.; Chen, G.; Deber, C. M. *Proc. Natl. Acad. Sci.* **2009**, *106* (6), 1760.
- (256) Wüthrich, K. *NMR of proteins and nucleic acids*; The George Fisher Baker non-resident lectureship in chemistry at Cornell University; Wiley: New York, 1986.
- (257) Driscoll, P. C.; Clore, G. M.; Marion, D.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry (Mosc.)* **1990**, *29* (14), 3542.
- (258) Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M. *Biochemistry (Mosc.)* **1989**, *28* (15), 6150.
- (259) Driscoll, P. C.; Gronenborn, A. M.; Wingfield, P. T.; Clore, G. M. *Biochemistry (Mosc.)* **1990**, *29* (19), 4668.
- (260) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69* (1), 185.
- (261) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (23), 12366.
- (262) Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. *Curr. Opin. Struct. Biol.* **2015**, *32*, 113.
- (263) *Isotope labeling in Biomolecular NMR*; Atreya, H. S., Ed.; Advances in Experimental Medicine and Biology; Springer Netherlands: Dordrecht, 2012; Vol. 992.
- (264) Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119* (32), 7599.
- (265) Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J. *J. Biomol. NMR* **2013**, *57* (3), 251.
- (266) Frueh, D. P.; Goodrich, A. C.; Mishra, S. H.; Nichols, S. R. *Curr. Opin. Struct. Biol.* **2013**, *23* (5), 734.
- (267) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125* (34), 10420.
- (268) Amero, C.; Schanda, P.; Durá, M. A.; Ayala, I.; Marion, D.; Franzetti, B.; Brutscher, B.; Boisbouvier, J. *J. Am. Chem. Soc.* **2009**, *131* (10), 3448.
- (269) Tugarinov, V.; Kanelis, V.; Kay, L. E. *Nat. Protoc.* **2006**, *1* (2), 749.
- (270) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445* (7128), 618.
- (271) Ikura, M.; Kay, L. E.; Bax, A. *Biochemistry (Mosc.)* **1990**, *29* (19), 4659.
- (272) Kanelis, V.; Forman-Kay, J. D.; Kay, L. E. *IUBMB Life* **2001**, *52* (6), 291.
- (273) Frueh, D. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *78*, 47.
- (274) Kazimierczuk, K.; Zawadzka, A.; Koźmiński, W. *J. Magn. Reson.* **2009**, *197* (2), 219.
- (275) Kazimierczuk, K.; Stanek, J.; Zawadzka-Kazimierczuk, A.; Koźmiński, W. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57* (4), 420.
- (276) Mobli, M.; Hoch, J. C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *83*, 21.
- (277) Schanda, P.; Brutscher, B. *J. Am. Chem. Soc.* **2005**, *127* (22), 8014.
- (278) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128* (28), 9042.
- (279) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187* (1), 163.
- (280) Zawadzka-Kazimierczuk, A.; Koźmiński, W.; Šanderová, H.; Krásný, L. *J. Biomol. NMR* **2012**, *52* (4), 329.
- (281) Felli, I. C.; Pierattelli, R. *J. Magn. Reson.* **2014**, *241*, 115.
- (282) Gil, S.; Hošek, T.; Solyom, Z.; Kümmerle, R.; Brutscher, B.; Pierattelli, R.; Felli, I. C. *Angew. Chem. Int. Ed Engl.* **2013**, *52* (45), 11808.
- (283) Piai, A.; Hošek, T.; Gonnelli, L.; Zawadzka-Kazimierczuk, A.; Koźmiński, W.; Brutscher, B.; Bermel, W.; Pierattelli, R.; Felli, I. C. *J. Biomol. NMR* **2014**, *60* (4), 209.
- (284) Grzesiek, S.; Anglister, J.; Ren, H.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115* (10), 4369.
- (285) Weisemann, R.; Rüterjans, H.; Bermel, W. *J. Biomol. NMR* **1993**, *3* (1), 113.

- (286) Matsuo, H.; Kupce, E.; Li, H.; Wagner, G. *J. Magn. Reson. B* **1996**, *111* (2), 194.
- (287) Bracken, C.; Palmer, A. G.; Cavanagh, J. *J. Biomol. NMR* **1997**, *9* (1), 94.
- (288) Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V. *J. Biomol. NMR* **2001**, *20* (2), 135.
- (289) Sun, Z.-Y. J.; Frueh, D. P.; Selenko, P.; Hoch, J. C.; Wagner, G. *J. Biomol. NMR* **2005**, *33* (1), 43.
- (290) Goddard, T. D.; Kneller, D. G. *University of California, San Francisco*.
- (291) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363.
- (292) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4* (2), 171.
- (293) Wang, Y.; Jardetzky, O. *Protein Sci. Publ. Protein Soc.* **2002**, *11* (4), 852.
- (294) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15* (12), 2795.
- (295) Ziarek, J. J.; Peterson, F. C.; Lytle, B. L.; Volkman, B. F. In *Methods in Enzymology*; Elsevier, 2011; Vol. 493, pp 241–275.
- (296) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1.
- (297) Farmer, B. T.; Constantine, K. L.; Goldfarb, V.; Friedrichs, M. S.; Wittekind, M.; Yanchunas, J.; Robertson, J. G.; Mueller, L. *Nat. Struct. Biol.* **1996**, *3* (12), 995.
- (298) Fielding, L. *Tetrahedron* **2000**, *56* (34), 6151.
- (299) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51* (4), 219.
- (300) Ababou, A.; Pfuhl, M.; Ladbury, J. E. *J. Mol. Biol.* **2009**, *387* (5), 1120.
- (301) Liang, Y. *Acta Biochim. Biophys. Sin.* **2008**, *40* (7), 565.
- (302) Duff, Jr., M. R.; Grubbs, J.; Howell, E. E. *J. Vis. Exp.* **2011**, No. 55.
- (303) Leavitt, S.; Freire, E. *Curr. Opin. Struct. Biol.* **2001**, *11* (5), 560.
- (304) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Phys. Chem. B* **2008**, *112* (19), 5898.
- (305) Scanu, S.; Foerster, J. M.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2013**, *135* (20), 7681.
- (306) Korzhnev, D. M.; Kay, L. E. *Acc. Chem. Res.* **2008**, *41* (3), 442.

Charlier Cyril

Date of Birth: 14th April 1989, Lesquin (France)

Nationality: French

Married

Tel: + 33 6 68 21 51 54

E-mail: cyril.charlier14@gmail.com



Professional Information

Département de Chimie; Ecole Normale Supérieure

24 Rue Lhomond

75005 Paris (France)

Tel: + 33 1 44 32 24 11

E-mail: cyril.charlier@ens.fr

Education

2012 – ... **PhD. student in Physical Chemistry**

*Département de Chimie; Ecole Normale Supérieure &
Université Pierre et Marie Curie*

Topic: Protein Dynamics from Nuclear Spin Relaxation: High-Resolution Relaxometry, Disordered Proteins and Applications to the C-Terminal Domain of the Protein Artemis

Advisors: Dr. Fabien Ferrage, Dr. Ludovic Carlier

2010 – 2012 **M. Sc. in Fundamental Physics**

Paris Sud University (Orsay, France)

2009 – 2010 **B. Sc. in Fundamental Physics**

Paris Sud University (Orsay, France)

Research experiences

2012 (6 months) **Paris Sud University**

Topic: Characterization of flow reaction by Nuclear Magnetic Resonance and Magnetic Resonance Imaging

Advisor: Prof. Denis Merlet

2011 (3 months) **Lille 1 University – Science and Technologies**

Topic: Study of the protein Cyclophilin A by in cell-NMR

Advisor: Dr. Guy Lippens

2010 (2 months) **Lille 1 University – Science and Technologies**

Topic: Study of the protein Cyclophilin A by in cell-NMR

Advisor: Dr. Guy Lippens

Publications

6. Cousin S., **Charlier C.**, Kadeřávek P., Marquardsen T., Tyburn J.M., Bovier P.A., Speck T., Ulzega S., Chou C.Y., Bodenhausen G., Sakellariou D., Pelupessy P., Maas W., Wilhelm D., Engelke F., Ferrage F.; High-resolution two-fields NMR spectroscopy. *In Preparation*

5. **Charlier C.**, Chou C. Y., Pelupessy P., Bodenhausen G., Cortes P., Aggarwal A. K., Carlier L., Sakellariou D., Ferrage F.; Nanosecond motions of the disordered C-terminal domain of Artemis by a combination of high-resolution relaxometry and high-field relaxation. *In Preparation*

4. **Charlier C.**, Pelupessy P., Walrant A., Marquant R., Sagan S., Bodenhausen G., Cortes P., Aggarwal A. K., Carlier L., Ferrage F.; Interactions and dynamics of the disordered C-terminal domain of Artemis. *In Preparation*

3. Billaut I., **Charlier C.**, Gillet B., Sebré C., Scherrmann M.C., Merlet D.; New meso fluidic reactors for flow synthesis studied by MRI. *In Preparation*

2. Khan SN[†], **Charlier C[†]**, Augustyniak R., Salvi N., Déjean V., Bodenhausen G., Lequin O., Pelupessy P., Ferrage F.; Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation: a simple array of correlation times. *Under review*

1. **Charlier C[†]**, Khan SN[†], Marquardsen T., Pelupessy P., Reiss V., Sakellariou D., Bodenhausen G., Engelke F., Ferrage F.; Nanosecond timescale motions in proteins by high-resolution NMR relaxometry. *J. Am. Chem. Soc.* 135, 18665-18672 (2013)

Presentations & Awards

56th Experimental Nuclear Magnetic Resonance Conference – Asilomar – 2015

« *Interactions and dynamics of the C-terminal domain of Artemis* » (**Poster**)

55th Experimental Nuclear Magnetic Resonance Conference – Boston – 2014

« *Nanosecond Timescale Motion in Proteins Revealed by High-Resolution NMR Relaxometry* » (**Oral Presentation, Poster and Student travel stipend**)

Réunion RMN grand bassin Parisien – Lille – 2013

« *Exploration of protein dynamics by high-resolution relaxometry* » (**Oral Presentation**)

Journée des jeunes chercheurs de l'ENS – Paris – 2013

« *Exploration of protein dynamics by high-resolution relaxometry* » (**Oral Presentation**)

EUROMAR 2013 – Hersonissos (**Poster and Student travel stipend**)

Languages

French: mother tongue

English: fluent

Computer Skills

Operating systems: Windows, Linus and Mac Os X

Programming languages: Mathematica, MATLAB

Softwares: LATEX, *Office Skills*, Adobe Illustrator

Title: Protein Dynamics from Nuclear Spin Relaxation: High-Resolution Relaxometry, Disordered Proteins and Applications to the C-Terminal Region of the Protein Artemis

The intimate relation between the dynamics and function of biomolecules is nowadays obvious. Indeed, proteins are very dynamic over a broad range of timescales. NMR is a unique technique to extract information on both structure and dynamics of biomolecules. Recent advances of NMR spectroscopy have substantially extended the description of protein dynamics and their relation to biological functions. The work presented here, focuses on the study of pico- to nanosecond timescale motions in proteins using nuclear spin relaxation and in particular studying the reorientation of backbone amide bond vectors in proteins. Conventional high-field relaxation measurements provide information on dynamics with a limited sampling of the density of motions.

We have used a new shuttle device designed for high-resolution relaxometry to measure relaxation rates over nearly two orders of magnitude of magnetic field. The application of high-resolution relaxometry provides unprecedented insights on internal dynamics in ubiquitin. A protocol of correction, called ICARUS, (Iterative Correction and Analysis of Relaxation Under Shuttling) has been developed in order to analyze relaxation rates. In particular it is essential to take into account relaxation pathways when the sample is in the stray field of the magnet. Defying expectations, we found that the nanosecond motions of the $\beta 1$ - $\beta 2$ turn can be revealed by relaxation measurements, in agreement with studies of residual dipolar couplings and accelerated molecular dynamics.

Over the last decades a new class of protein challenging the “structure-function paradigm” has emerged: Intrinsically Disordered Proteins (IDPs). These proteins exhibit a broad conformational ensemble due to the lack of a stable tri-dimensional conformation. However, the interpretation of relaxation data measured on IDPs requires the use of adapted models of protein dynamics. We used a fragment of the transcription factor Engrailed 2 to develop a new approach for the interpretation of relaxation data measured on IDPs called IMPACT (Interpretation of Motions by a Projection onto an Array of Correlation Times). IMPACT consists in the reconstruction of the distribution of correlation times without relying on a particular model of motions. Our results reveal the distribution of pico- to nanosecond timescale motions in the disordered region of Engrailed 2.

In parallel, we have studied a disordered domain of the DNA-repair factor Artemis. This protein, crucial for the immune system, interacts with the DNA Binding Domain of Ligase IV (DBD-LigIV). We have optimized the sample preparation and assigned backbone resonances in Artemis. We have then conducted a study of the interaction between Artemis and the DBD-LigIV. We have shown the role of 5 additional amino acids past the C-terminus of the previously identified interaction site and characterized the kinetics and thermodynamics of the interaction by NMR relaxation dispersion and isothermal titration calorimetry. Independently, we performed a series of relaxation measurements using high-resolution relaxometry (from 0.2 T to 6 T) and high-field relaxation measurements (from 11.4 T to 22.3 T). This combination of high-resolution relaxometry and high-field relaxation measurements will expand our understanding of ps-ns dynamics in intrinsically disordered proteins.

Overall, this work contributes to the development of new concepts for the interpretation of extensive nuclear spin relaxation data in proteins.

Keywords: [Biomolecular NMR, Protein Dynamics, Intrinsically Disordered Proteins (IDPs), Relaxation, High-Resolution Relaxometry, ubiquitin, Engrailed, Artemis]