Specific labeling strategies for new developments in liquid state protein NMR

Rachel Lenoir-Capello

▶ **To cite this version:**

Rachel Lenoir-Capello. Specific labeling strategies for new developments in liquid state protein NMR. Biochemistry, Molecular Biology. Sorbonne Université, 2020. English. NNT : 2020SORUS056 . tel-03202318

HAL Id: tel-03202318
https://theses.hal.science/tel-03202318

Submitted on 19 Apr 2021

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

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

Sorbonne Université

THESE de DOCTORAT

Domaine : *Chimie Physique et Chimie Analytique*

Présentée par

Rachel LENOIR-CAPELLO

Pour obtenir le grade de DOCTEUR EN SCIENCES de Sorbonne Université
Sous la direction du **Dr. Emeric Miclet**

Sujet : Specific labeling strategies for new developments in liquid state protein NMR

Soutenue le 5 novembre 2020 devant le jury composé de

Christina Sizun	Rapporteur
Benoît Odaert	Rapporteur
Emeric Miclet	Directeur de thèse
Teresa Carlomagno	Examinateur
Juliette Sirieix-Plénet	Examinateur
Jérôme Boisbouvier	Examinateur

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

Acknowledgments

Foremost, I would like to thank the revisers: Dr. Christina Sizun, and Dr. Benoît Odaert, and the jury members: Pr. Teresa Carlomagno, Dr. Juliette Sirieix-Plénet and Dr. Jérôme Boisbouvier for having accepted to evaluate the scientific relevance of my thesis, I am most grateful for their time and contributions.

This work was realized, in most part, at the Laboratoire des Biomolécules (LBM) of Sorbonne Université, directed by Dr. Sandrine Sagan, to whom I extend my thanks for having welcomed me to the research unit.

To my PhD supervisor, Dr. Emeric Miclet, I express my utmost gratitude, foremost, for having given me the unique opportunity to accomplish this project. It was an invaluable professional and personal experience during which I discovered so many aspects of laboratory research. I would also thank Dr. Emeric Miclet for his guidance, the knowledge in NMR and chemistry he continuously shared with me and the patience with which he did, it was an honor and a pleasure to be his student.

I would also like to express my gratitude to Lionel Imbert and Dr. Jérôme Boisbouvier from the Institut de Biologie Structurale for having welcomed me at the Cell-free platform and for the time and energy they spent transmitting me their knowledge, both in cell-free expression as in solution NMR spectroscopy. I would also thank them for having included me in their publication project.

I wish to acknowledge the help provided by Dr. Philippe Pelupessy in developing the three-dimensional spin state selective experiments.

I am particularly grateful for the assistance and expertise given by David Guarin and Dr. Daniel Abergel in starting the DNP project of my thesis.

I would also like to extend my thanks to Gilles Clodic, Dr. Gérard Bolbach and Dr. Emmanuelle Sachon of the mass spectrometry platform at Sorbonne Université for their help in running my samples and the time they dedicated to my project.

I would like to thank Pr. Olivier Lequin, Dr. Jean-Jacques Lacapère, Dr. Ludovic Carlier and Isabelle Correira, for their valuable and constructive recommendations during this research work, and for always finding the time to help me when I needed.

I thank Dr. Gregory Chaume and Dr. Chiara Zanatto for the help they provided synthesizing the ligands, and their individual contributions and suggestions throughout my study.

I wish to thank Soha Abou Ibrahim, Paula Milan Rodriguez and Edward Chalouhi for their continued support and help to this project, notably Soha for aiding me in protein expression and purification, Paula for creating and running scripts for data analysis.

I also thank the other members of the LBM whom I have not named here who have contributed by near or far to the realization of my project.

This work was made possible thanks to the CH2PROBE project funded by the Agence National de Recherche.

Specific labeling strategies for new developments in liquid state protein NMR

SUMMARY:

Abbreviations .9
Short Abstract .11
Court Abstract .13
Résumé en Français .15

Introduction .37

Chapter I: Concerns about sensitivity and resolution in solution NMR .43
1. Nuclear Magnetic Resonance phenomenon .43
1. Generalities .43
2. Two types of relaxation .45
3. Different sources of relaxation .47
2. Reducing T_1 and T_2 relaxation through deuteration labeling strategies .49
1. Deuteration for amide spin system analysis .51
2. Deuteration for glycine methylene spin system analysis .52
3. Deuteration for protein D-DNP of protons .52
3. Generalities concerning the TROSY experiments .53
1. Introduction .53
2. Theory of the CH_2-TROSY experiment .59
3. S3-Gly experiments .64
4. Dissolution-Dynamic Nuclear Polarization .66
1. Introduction .66
2. Electron Polarization Transfer .68
Bibliography .69

Chapter II: Structural properties of the two proteins studied .73
1. Pin1: a brief overview .73
1. Introduction .73
2. Activity, structure and mechanism of action .74
3. Partners and biological implications .78
4. Choice of Pin1 model .79
5. Expression of different constructs of Pin1 in *E.coli* and assignment .80
2. NMR study of H23 protein .82
1. Introduction and choice of H23 model .82
2. NMR assignment of H23 .84
3. Materials and Methods .90
1. *E.coli* expression and purification conditions for Pin1 .90
2. Recording conditions for Pin1 .92
3. H23 U-^{1}H-^{15}N-^{13}C labeled expression and purification for assignments .93
4. Assignment spectra recording conditions for H23 .96
Bibliography .98

Chapter III: Optimization and application of the cell free expression for the perdeuteration of proteins .101
1. Introduction .101
1. General principle of the cell free expression and set up .104
2. Configuration of the system .106
3. S30 cell extract from *E.coli* .107
4. Amino acids supply .110
5. Isotopic scrambling during the cell-free expression .112
6. Inhibition of PLP dependent enzymes .115
7. Plasmids for cell-free expression .117
2. Cell-free protocol for the perdeuteration of proteins in H_2O .118
1. Quantifying Cα protonation using the optimized protocol .118
2. Demonstrating the utility of the optimized cell-free protocol for the expression of the TET2 protein .122
3. Conclusions .122
3. Materials and Methods .123
1. *E.coli* extracts preparation (S30 extract) .124
2. Cell-free expression of proteins .126
3. 3D HNCA recording conditions for H23 in batch and CECF modes .129
Bibliography .131

Chapter IV: Optimization and application of the cell free expression for the specific labeling of Gly residues in proteins .133
1. Introduction .133
2. Challenges for the specific labeling of Gly residues in proteins. .135
1. Limitations *in vivo* .135
2. Limitations in cell free .136
3. Optimization of the Gly labeling protocol using the H23 protein .139
4. Application of the protocol to the Pin1 protein .153
3. Benefits of Gly specific labeling for NMR experiments .160
1. CH_2-TROSY experiments .160
2. S3-Gly experiments .174
3. Conclusions and perspectives .181
4. Materials and Methods .183
1. Cell free protocols for Gly specific H23 labeling .183
2. Recording conditions for ^{1}H-^{15}N HSQC spectra of H23 .185
3. Recording conditions for 1D spectra of H23 .185
4. Recording conditions for ^{1}H-^{13}C HSQC spectra of H23 .186
5. Recording conditions for ^{1}H-^{13}C HSQC spectra of feeding mixes .186
6. Recording conditions for MASS spectrometry .186
7. CH_2-TROSY and CH_2-HSQC recording conditions .187
8. Calculation of the dA proton apparent distance .187
9. PPIase titration recording conditions .188
10. s3-HACACO recording conditions .189
11. Docking simulations .189
Bibliography .191

Chapter V: Towards protein D-DNP NMR .193
1. Introduction .193
2. Importance of T1 relaxation for D-DNP .198
1. WW labeling strategy .200
2. Choice of the WW ligand .203
3. T_1 measurements of the different WW labeled samples .204
3. Protein D-DNP experimentations .209
1. D-DNP set up and "DNP juice" .209
2. Results .210
3. Conclusions and perspectives .212
4. Materials and Methods .213
1. ^{2}H-^{15}N labeled WW domain expression .213
2. Recording conditions for ^{1}H-^{15}N HSQC spectra of the WW domain .214
3. Recording conditions for ^{1}H 1D measurements of the WW domain .214
4. Recording conditions for ^{1}H T_1 measurements of the WW domain .215
5. MASS spectrometry of the WW samples .215
6. Sample conditions for WW domain polarization .216
Bibliography .217

Conclusions and perspectives .219

Annexes .223

Abbreviations

AMUPol: (15-{[(7-oxyl-3,11-dioxa-7-azadispiro[5.1.5.3]hexadec-15- yl)carbamoyl][2-(2,5,8,11-tetraoxatridecan-13-ylamino)}-[3,11-dioxa-7- azadispiro[5.1.5.3]hexadec-7-yl])oxidanyl
AOA: aminooxyacetic acid
BEST-TROSY: Band-selective Excitation Short-Transient TROSY
CECF: Continuous exchange Cell-free
CF: Cell-free
CRINEPT: cross relaxation-enhanced polarization transfer
CSA: Chemical Shift Anisotropy
CSP : Chemical Shift Perturbation
DD: Dipole Dipole
D-DNP: Dissolution-Dynamic Nuclear Polarization
DM: D-malic acid
DMF: Dimethylformamide
DMSO: Dimethyl sulfoxide
DNA: Desoxyribonucleic acid
DNP: Dynamic Nuclear Polarization
DTT: Dithiothreitol
DSS: Sodium trimethylsilylpropanesulfonate
EDTA: Ethylenediaminetetraacetic acid
FID: Free Induction Decay
FM: Feeding Mixture
HACAN: spin state selective 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: Ha proton of the backbone to the Ca, and then to the nitrogen of the amide of the residue.
HACACO: spin state selective 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: Ha proton of the backbone to the Ca, and then to the carbonyl of the same residue
HCCA: α-Cyano-4-hydroxycinnamic acid
HMQC: Heteronuclear Multiple Quantum Correlation
HNCA: 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid: amide proton to the amide nitrogen, to the alpha carbons of both the starting residue and the previous residue in the protein's amino acid sequence.
HNCACB: 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: amide proton to the amide nitrogen, and then to the alpha carbons of both the starting residue and the previous residue and then to the beta carbons of both the starting residue and the previous residue in the protein's amino acid sequence.
HNCO: 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: amide proton to the amide nitrogen, and then to the carbonyl of the previous residue in the protein's amino acid sequence.
HN(CO)CA: 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: amide proton to the amide nitrogen, then to the carbonyl, then to the alpha carbons of the previous residue in the protein's amino acid sequence.
HN(CO)CACB: 3D triple-resonance NMR experiment with magnetization transfer pathway for an amino acid residue: amide proton to the amide nitrogen, and then to the carbonyl of the previous residue in the protein's amino acid sequence, then to its alpha carbon and finally to its beta carbon.
HPLC: High-performance liquid chromatography

HSQC: Heteronuclear Single Quantum Correlation
IDP: Intrinsically Disordered Protein
INEPT: Insensitive Nuclei Enhanced Polarization Transfer
MALDI TOF: Matrix Assisted Laser Desorption Ionization - Time of Flight
NMR: Nuclear Magnetic Resonance
NOE: Nuclear Overhausser Effect
pIVEX: plasmid for *In Vitro* Expression
PLP: Pyridoxal Phosphate
RM: Reaction Mixture
RNA: Ribonucleic acid
SOFAST: Band-Selective Optimized Flip Angle Short Transient
TCA: Tricarboxylic acid cycle
TROSY: Transverse Relaxation Optimized SpectroscopY

Resumé

Solution Nuclear Magnetic Resonance (NMR) is a tool which provides valuable structural and dynamic information at the atomic scale. It has become an essential technique in many disciplines, particularly in biology and pharmacology, however, it suffers from two major disadvantages: low sensitivity and low resolution of signals, which rapidly preclude investigations of larger molecular objects. But, optimizing the biological sample can, in itself, enhance the quality of spectra. We present here three isotopic labeling strategies for different protein-solution NMR experiments and demonstrate, with the help of NMR developments, that they make it possible to push back the current limits for the structural study of biomolecules in solution.

Among the strategies we have considered, two are based on the optimization of *in vitro* protein expression (cell-free expression) to obtain selectively labeled proteins of a certain chemical group and/or amino acids in a perdeuterated environment. Obtaining such labeled samples can prove very difficult to achieve using *in vivo* expression. Perdeuteration is, however, essential for the optimal use of the TROSY (Transverse Relaxation Optimized Spectroscopy) pulse sequences, which are more efficient when the environment of the observed chemical group is deuterated. We demonstrated that the TROSY experiments allowed significant spectral gains when samples were specifically labeled on the amide groups or on the methylene of glycine residues while maintaining a very high rate of deuteration on all of the other chemical functions of the proteins. Optimization of labeling under cell-free conditions was undertaken on a subunit of the HMA8 protein, an ATPase from *Arabidopsis thaliana*. The protocol adopted for the selective labeling of amides was successfully applied to the TET2 protein and we were able to demonstrate its utility by recording different ^{1}H-^{15}N CRINEPT correlation spectra on this half-mega Dalton protein. To assess the benefit of specific labeling of glycine residues, we systematically quantified the gains in resolution and sensitivity obtained in CH_2-TROSY experiments recorded on different types of HMA8 samples. In comparison with a classical approach (uniformly labeled samples and ^{1}H-^{13}C HSQC experiment), the spectral improvements obtained for the resonances of the methylene groups allowed us to use them to precisely follow the interaction of the Pin1 protein with ligands developed in laboratory. Developments of 3D experiments, dedicated to glycine residues, were also undertaken. The many coupling constants that they allow to obtain could be useful for structural and dynamic studies of proteins.

Finally, the third isotopic protein labeling strategy that we employed is based on protocols already well described in the literature but used in innovative NMR applications. In recent years has emerged a technique of hyperpolarization of nuclei in solution which increases their sensitivity by several orders of magnitude. Dissolution-Dynamic Nuclear Polarization (D-DNP) transfers the polarization of electrons to atomic nuclei at cryogenic temperature, followed by dissolution of the sample to record hyperpolarized NMR signals. However, the lifetime of this hyperpolarization is governed by the longitudinal relaxation time of the nuclei, which are greatly reduced for proteins at room temperature. One way to lengthen relaxation times is to isolate the nuclei of interest in a perdeuterated environment. In this way, the dipolar interactions created by neighboring protons are eliminated and hyperpolarized nuclei relax much more slowly. The WW domain of the Pin1 protein was used for these purposes and deuterated in varying proportions. At around 80% deuteration, some measured T_1 relaxation times increased by a factor of 4 to 5 for some protons compared to a non-deuterated sample. Hyperpolarization of this domain has been successfully undertaken at 1K. The first tests that we have undertaken

show that this domain regains its fold very quickly after heating to 90°C. However, the dissolution conditions need to be improved in order to preserve a homogeneous aqueous phase.

Résumé

La Résonnance Magnétique Nucléaire (RMN) en solution est un outil qui permet d'obtenir des informations structurales et dynamiques précieuses à l'échelle atomique. Devenue indispensable dans de nombreuses disciplines, notamment en biologie et pharmacologie, elle est cependant limitée par la mauvaise résolution et la faible sensibilité des signaux détectés, limitant fortement la précision des mesures pour les systèmes de haut poids moléculaire. Toutefois, l'optimisation de l'échantillon permet, à lui seul, d'améliorer la qualité de ces signaux. Ce travail de thèse explore trois stratégies de marquage isotopique des protéines et démontre, à l'aide de développements RMN, qu'ils permettent de repousser les limites actuelles pour l'étude structurale des biomolécules en solution.

Parmi les stratégies que nous avons envisagées, deux reposent sur l'optimisation de l'expression des protéines *in vitro* (expression acellulaire) afin d'obtenir des marquages isotopiques spécifiques de certains groupes chimiques et/ou d'acide aminés dans un environnement perdeutéré. Ce type de marquage est très difficile à atteindre ou se réalise de manière imparfaite *in vivo*. La perdeutération est pourtant indispensable pour l'utilisation optimale des séquences d'impulsion de type TROSY (Transverse Relaxation Optimized Spectroscopy), qui sont d'autant plus efficaces que l'environnement du groupe chimique observé est deutéré. Nous avons démontré que les expériences TROSY permettaient des gains spectraux importants lorsque les échantillons étaient spécifiquement marqués sur les groupes amides ou au niveau des méthylènes des résidus glycines en conservant un très fort taux de deutération sur l'ensemble des autres fonctions chimique des protéines. L'optimisation des marquages en conditions acellulaires a été entreprise sur une sous unité de la protéine HMA8, une ATPase de *Arabidopsis thaliana*. Le protocole retenu pour le marquage sélectif des amides a été appliqué avec succès à la protéine TET2 et nous avons pu démontrer son intérêt en enregistrant différents spectres de corrélation ^{1}H-^{15}N CRINEPT sur cette protéine d'un demi méga Dalton. Pour évaluer l'intérêt des marquages spécifiques des résidus glycine, nous avons quantifié de manière systématique les gains en résolution et en sensibilité obtenus dans des expériences CH_2 TROSY enregistrées sur différents types d'échantillons de HMA8. En comparaison avec une approche classique (échantillons uniformément marqué et expérience ^{1}H-^{13}C HSQC), les améliorations de qualité spectrale obtenues pour les résonances des groupes méthylènes nous ont permis de nous en servir pour suivre avec précision l'interaction de la protéine Pin1 avec des ligands développés au laboratoire. Des développements d'expériences 3D, dédiées aux résidus glycine, ont également été entrepris. Les nombreuses constantes de couplages qu'elles permettent d'obtenir pourront être utiles pour les études structurales et dynamiques des protéines.

Enfin, la troisième voie de marquage isotopique de protéines que nous avons utilisée repose sur des protocoles déjà bien décrits dans la littérature mais pour une utilisation RMN originale. Ces dernières années ont vu apparaître une technique d'hyperpolarisation des noyaux en solution, qui permet d'accroître leur sensibilité par plusieurs ordres de grandeur. La Polarisation nucléaire dynamique par dissolution (D-DNP) permet de transférer la polarisation des électrons aux noyaux atomiques à température cryogénique, puis, par dissolution de l'échantillon, l'enregistrement de signaux RMN hyperpolarisés. Toutefois, la durée de vie de cette hyperpolarisation est régie par le temps de relaxation longitudinale des noyaux, et est fortement réduite pour les protéines à température ambiante. Un moyen de rallonger les temps de relaxation est d'isoler les noyaux d'intérêt dans un environnement perdeutéré. De cette manière, les interactions dipolaires crées par les protons voisins sont éliminés et les noyaux hyperpolarisés relaxent beaucoup plus lentement. Le domaine WW de la protéine Pin1 a été

utilisé à ces fins et, deutéré dans des proportions variables. Aux alentours de 80% de deutération, certains temps de relaxation T1 mesurés ont augmenté d'un facteur 4 à 5 pour certains protons par rapport à un échantillon non deutéré. L'hyperpolarisation de ce domaine a été entreprise avec succès à 1K. Les premiers essais que nous avons effectués montrent que ce domaine se replie très rapidement après chauffage à 90°C mais les conditions de dissolutions restent à améliorer pour notamment préserver une phase aqueuse homogène.

Résumé Français

Problématique :

La résonance magnétique nucléaire (RMN) est un phénomène physique observé lorsqu'un noyau, excité par une impulsion radiofréquence, libère l'énergie absorbée à une fréquence spécifique. Le signal généré offre des informations sur les interactions du noyau avec d'autres noyaux ou électrons de son environnement. Ainsi, la spectroscopie RMN est devenue un outil extrêmement polyvalent pour l'étude des molécules, avec de nombreuses applications notamment en biologie structurale, où la spectroscopie RMN en solution est désormais essentielle pour élucider la structure et la dynamique des biomolécules au niveau atomique.

Toutefois, la RMN liquide appliquée à la biologie comporte deux grandes limitations. Lors de l'étude des biomolécules de grande masse moléculaire (> 50 kDa), cette technique s'avère en effet peu sensible et la résolution des signaux acquis est mauvaise, même en appliquant les derniers développements méthodologiques décrits récemment [1–4]. Les expériences TROSY (Transverse Relaxation Optimized SpectroscopY), décrites à l'origine pour les fonctions amides[5] en 1997, ont cependant considérablement élargi les possibilités des études de RMN des protéines. Ces expériences de corrélation hétéronucléaire 2D permettent d'éditer uniquement la composante relaxant le plus lentement parmi toutes celles qui composent le multiplet résultant d'un système de spins couplés. Après la démonstration sur les paires d'amides ^{15}N-^{1}H, l'approche TROSY a été appliquée à de nombreux autres groupes chimiques, et notamment aux méthyles[6]. Ces expériences TROSY sont d'autant plus efficaces que le système de spin d'intérêt est isolé spatialement des autres protons. La perdeutération des échantillons est donc cruciale pour atteindre le plein potentiel des expériences TROSY. En combinant une expérience TROSY, le marquage isotopique sélectif de certains groupes méthyles et en enregistrant des données RMN à très haut champ, les chercheurs ont été en mesure d'obtenir des signaux bien résolus pour des protéines jusqu'à 1 méga Dalton[7]. La deutération complète des protéines (perdeutération) en introduisant que quelques noyaux ^{1}H sur des fonctions chimiques choisies reste cependant difficile. Pour le marquage spécifique des groupes amides, l'expression hétérologue dans l'eau s'accompagne d'une protonation élevé des carbones α, même en utilisant une source d'acides aminés deutérés. A l'inverse, l'expression des protéines dans l'eau lourde (D_2O) peut être une source de problèmes pour l'étude RMN des groupes amides de protéine de poids moléculaire important car l'échange ^{2}H/^{1}H est extrêmement ralenti au sein de leur cœur hydrophobe qui est très stable. Pour ce type de protéines, des étapes supplémentaires

de dénaturation et renaturation sont nécessaires, affectant considérablement les rendements finaux. Ainsi, le premier objectif de ce travail de thèse est d'optimiser un protocole d'expression pour la perdeutération des protéines en absence d'eau lourde, grâce à une approche « cell free ». Ce protocole a été appliqué à une protéine de près d'un demi-méga Dalton, Tet2. Le marquage sélectif ^{15}N-^{1}H des groupes amides a pu être obtenu, alors que les positions Cα présentent un taux de deutération supérieur à 95%, comme démontré par une analyse approfondie de la protonation résiduelle après expression.

Parmi toutes les expériences TROSY disponibles, l'une des moins exploitées est l'expérience CH_2-TROSY[8], dédiée aux groupes méthylènes. Les méthylènes représentent pourtant 46% des protons dans les protéines. Ceci peut être expliqué par le fait qu'il n'y ait pas encore de stratégie de marquage publiée permettant un marquage spécifique des méthylènes dans un environnement perdeutéré. Le deuxième objectif de cette thèse examine donc comment le protocole d'expression optimisé précédemment peut être adapté pour conserver sélectivement les deux protons Hα diastéréotopes des résidus glycine dans une protéine par ailleurs perdeutérée. Les bénéfices obtenus grâce à de tels échantillons dans l'expérience CH_2-TROSY ont été quantifiés. L'optimisation du protocole de marquage spécifique et des premières mesures RMN ont été entreprises sur la protéine HMA8[9] (H23). Les améliorations spectrales obtenues ont permis dans un second temps d'effectuer la caractérisation des interactions biomoléculaires impliquant la protéine Pin1[10]. Enfin, de nombreuses constantes de couplage ont pu être mesurées grâce aux expériences « spin state sélectives » 3D sur les résidus glycine de Pin1.

Dans une dernière partie, nous nous sommes intéressés aux opportunités offertes par la technique D-DNP pour l'étude des protéines. L'intensité du signal RMN ^{1}H est liée à la polarisation nucléaire qui est faible à température ambiante, même à des champs magnétiques élevés de plusieurs Teslas. En conséquence, une accumulation du signal est nécessaire pour compenser une faible sensibilité, et cela prolonge considérablement la durée des expériences RMN. Cependant, une technique appelée Polarisation Nucléaire Dynamique (DNP) permet d'améliorer considérablement la sensibilité des noyaux grâce à une hyperpolarisation des spins nucléaires. Le signal ainsi obtenu augmente de plusieurs ordres de grandeur, ouvrant une nouvelle voie pour les investigations biologiques RMN sur des échelles de temps courtes et avec des échantillons faiblement concentrés. La technique de DNP par dissolution[11] nécessite une instrumentation dédiée et couteuse, et reste donc limité à très peu de laboratoires dans le monde. Par conséquent, bien que certaines publications concernent des études de biomolécules

par D-DNP[12,13], il n'y a pas encore eu de travaux publiés utilisant la polarisation directe des protons de protéines. Nous disposons au laboratoire de l'équipement nécessaire pour le développement de la D-DNP à l'étude des protéines. La dernière partie de cette thèse se concentre donc sur une stratégie de marquage qui pourrait améliorer les investigations des protéines par D-DNP. Nous avons souhaité évaluer le potentiel d'échantillons faiblement protonés pour des applications en D-DNP. Une protonation dispersée et aléatoire dans un environnement perdeutéré permettrait en effet de limiter significativement la relaxation magnétique nucléaire pour tirer pleinement profit de l'hyperpolarisation ^{1}H des protéines. Pour ces études, nous avons utilisé la petite sous-unité de Pin1, WW, et évalué, en fonction du taux de protonation, les effets sur la relaxation longitudinale T_1. Des expériences préliminaires ont ensuite été menées pour mettre au point l'approche D-DNP sur un échantillon protéique.

L'objectif général de cette thèse est donc d'optimiser de nouvelles stratégies de marquages pour améliorer le potentiel des expériences NH-TROSY, CH_2 TROSY et D-DNP. Pour chacune de ces applications, le marquage sélectif et spécifique de certains protons a été nécessaire, tout en gardant une deutération stricte des autres hydrogènes des protéines. Les expériences RMN étudiées et développées bénéficient grandement de la suppression des interactions dipolaires, générées par la présence de protons. Les méthodes disponibles pour la deutération et/ou le marquage sélectif de protéines présentent plusieurs inconvénients. Lorsqu'il s'agit d'enregistrer des expériences de corrélation NH sur des protéines ou complexes de haut poids moléculaire, des échantillons perdeutérés doivent être produits, uniformément marqués en ^{15}N. En milieu cellulaire, une expression effectuée à partir d'une source d'acides aminés deutérés dans un tampon $^{1}H_2O$ se caractérise en définitive par un niveau de deutération relativement bas (moins de 10% des positions Hα)[14]. Ceci est lié à la présence dans les cellules vivantes d'enzymes qui métabolisent les acides aminés deutérés introduits, et/ou en synthétisent davantage mais non marqués en isotopes ^{2}H. Si un solvant deutéré est utilisé, on doit alors attendre l'échange spontané $^{2}H/^{1}H$ des hydrogènes amide lorsque le système biomoléculaire sera remis en solution dans de l'eau légère $^{1}H_2O$ pour l'enregistrement des données RMN. Cet échange peut être extrêmement lent pour les protéines de haut poids moléculaire et des étapes de dénaturation/renaturation sont alors nécessaires, ce qui n'est pas compatible avec certains échantillons sans perdre une quantité très importante de matériel. De la même manière, certaines enzymes utilisent la glycine comme substrat ou la produisent à partir d'autres précurseurs. Ainsi, le marquage sélectif des glycines ne peut pas être réalisé dans des cellules vivantes sans éviter une dilution isotopique des noyaux portés sélectivement par les glycines,

ou inversement portés uniquement les autres acides aminés. L'expression protéique *in vitro* constituera donc une stratégie de choix dans ces deux cas. En revanche, l'introduction aléatoire de quelques ^{1}H sur des protéines globalement deutérées pourra être obtenue par des techniques d'expression classique, en milieu cellulaire. Ces échantillons pourraient être utiles pour obtenir une polarisation efficace des protéines par D-DNP.

La synthèse de protéines en cell-free est une technique qui permet l'expression de protéines *in vitro* en exploitant la fraction de la machinerie cellulaire nécessaire à la synthèse de la protéine cible, tout en éliminant les éléments impliqués dans les mécanismes de viabilité et de croissance cellulaire. Cette approche permet ainsi de contrôler la synthèse de la protéine d'intérêt. Le milieu n'est donc pas contraint par des mécanismes de survie cellulaire, et sa nature ouverte permet l'accès et le contrôle de l'environnement de production. Les mécanismes de transcription et de traduction sont extraits de lysats cellulaires[15,16], l'ADN et l'ARN d'origine sont éliminés et l'extrait, nommé S30, est complété par l'ARN polymérase T7, des acides aminés portant le marquage isotopique choisi, des sources d'énergie et une matrice d'ADN (ou ARN). Des composés nécessaires à l'amélioration de l'expression et de la viabilité de la protéine d'intérêt, parmi lesquels divers stabilisants, peuvent être ajoutés à tout moment au cours de l'expression. Le cell-free est une alternative intéressante pour le marquage sélectif des acides aminés (ou groupes d'acides aminés) car il donne la possibilité de rajouter des inhibiteurs de transaminases[17,18]. Dans ce travail de thèse, son optimisation a permis la production de protéines perdeutérées sur tous les sites non-échangeables mais protonées sur tous les sites amides y compris au sein des cœurs hydrophobes. La culture « cell-free » a aussi permis le marquage isotopique des glycines avec une grande sélectivité, c'est-à-dire sans que les résidus glycines ne soient contaminés par les isotopes ^{2}H, ^{14}N, ^{12}C des autres acides aminés présents dans le milieu de culture, en assurant parallèlement que les isotopes ^{1}H, ^{15}N et ^{13}C restent « rivés » aux résidus glycine en fin de culture.

Protéines étudiées

Nous avons été amenés à étudier trois protéines dans le cadre de ce travail de thèse : la protéine Pin1, H23 et la protéine TET2. Cette dernière a été produite par nos collaborateurs de l'IBS à Grenoble et ne sera pas présentée dans le cadre de cette introduction. En revanche, nous avons concentré nous efforts au laboratoire sur les protéines Pin1 et H23 qui nous ont à la fois servi

de protéines modèles pour l'optimisation des marquages isotopiques en culture cell-free et à valider de nouvelles approches en spectroscopie RMN.

I) La peptidyl-prolyl isomérase Pin1

Dans les protéines, l'angle dièdre ω qui définit la géométrie de la liaison peptidique plane, peut prendre deux valeurs : 0 ° (*cis*) et 180 ° (*trans*). Cette restriction de l'espace conformationnel est liée au caractère de double liaison partielle de la liaison peptidique (40%). Les liaisons amide tertiaire des résidus prolines adoptent à 94% une conformation *trans*, et à 6% une conformation *cis*[19,20]. Cette population de conformères *cis* est très importante comparée à ce qui est rencontré pour les autres types d'acide aminés. Ainsi, deux géométries du squelette peptidique sont possibles en amont de chaque résidu proline des protéines. Ces deux conformations ne sont pas associées aux mêmes propriétés d'interaction, d'activité, de stabilité et sont en outre séparées par une barrière d'énergie élevée (environ 20 kcal/mol[21]) pour inter-convertir ces deux états. Par conséquent l'isomérisation *cis* ⇆ *trans* spontanée est lente (<0,002/s à 25 °C)[22]. Pin1 est une peptidyl prolyl *cis-trans* isomérase et agit en abaissant la barrière énergétique nécessaire pour réaliser cette isomérisation. Elle accélère la réaction d'un facteur 100 à 1000[10,23], agissant cependant uniquement sur les motifs phosphosérine/phospho-thréonine-proline (pS/pT-P). Depuis sa découverte, de nombreuses études ont été menées sur Pin1, mettant en évidence son implication dans la régulation de nombreuses voies métaboliques.

Pin1 humaine comporte 163 acides aminés, et correspond à un poids moléculaire de 18,243 kDa (Uniprot). Elle possède deux domaines : le domaine WW (nommé ainsi en raison des deux tryptophanes conservés W11 et W34) couvrant les résidus M1-G39, un linker des résidus N40 à Q49 et un domaine PPIase, possédant l'activité catalytique, compris entre les résidus G50 et E163. Les domaines WW et PPIase fixent les mêmes substrats mais le domaine WW possède une affinité multipliée par dix vis-à-vis de ces substrats en comparaison avec le domaine catalytique. Il reconnaît préférentiellement la conformation *trans*[24] des motifs peptidyl-prolyl mais ne possède lui-même aucune activité catalytique et n'est pas nécessaire à l'activité du domaine PPIase. Pour sa part, le domaine PPIase reconnaît préférentiellement les conformations *cis*. WW se compose de trois brins bêta reliés par deux boucles formant un feuillet bêta antiparallèle. Le domaine WW présente une stabilité élevée, avec une température de dénaturation Tm de 59,0°C[25,26]. Par ailleurs, la dénaturation thermique du domaine WW est

parfaitement réversible. La cinétique de repliement de la boucle 1 tant considérée comme très rapide : 59.4 μs[25] à 45°C. Ces caractéristiques thermodynamiques seront utilisées dans les développements D-DNP. Le domaine PPIase est constitué de 4 hélices alpha et d'un feuillet central de 4 brins β antiparallèles, face à l'hélice α3. Il n'y a pas de consensus général sur le mécanisme catalytique de Pin1, bien que plusieurs modèles aient été proposés[27].

Grace à la multitude des travaux menés sur la protéine Pin1, les données structurales disponibles dans la Protein Data Bank (pdb) sont nombreuses, dont la première structure RX déposée en 1998[28] (1PIN) et avec aujourd'hui 16 structures déterminées par RMN disponibles pour Pin1 et ses domaines séparés . Des études RMN et de modélisation moléculaire ont également permis de décrire la dynamique de cette protéine, notamment au niveau des mouvements entre ses deux domaines. Ce socle de connaissances sur la protéine Pin1 est un atout pour le développement de nouvelles méthode RMN car les nouvelles mesures effectuées pourront être comparées aux données de la littérature. Par ailleurs, la protéine Pin1 possède 15 glycines. 3 glycines se trouvent 3 dans le domaine WW, 3 au niveau du linker et 9 dans le domaine PPIase, elles sont donc bien dispersées à travers sa séquence et dans différents éléments de structure secondaire. Les 15 glycines représentent 9% des résidus de la séquence, soit sensiblement plus que leur abondance naturelle les protéines (7%)[29]. Les corrélations ^{1}H-^{15}N ou ^{1}H-$^{13}C\alpha$ étant par ailleurs bien éclatées sur les spectres HSQC, nous avons jugé que cette protéine était un bon modèle pour cette étude. En effet, les résidus glycine sont extrêmement pertinents pour les expériences RMN spécifiques du méthylène. Ils pourraient servir de sondes locales en permettant des mesures précises de nombreuses constantes de couplages (scalaires ou dipolaires). Le groupe méthylène des glycines fait partie intégrante du squelette peptidique. A ce titre, des changements de structure ou de dynamique induits par la présence de ligands pourront être mis en évidence sans ambiguïté par des mesures RMN et ces données pourront être comparées aux informations structurales déjà disponibles. Par ailleurs, Pin1 interagit *in vivo* avec de nombreux partenaires protéiques ce qui permet d'envisager la réalisation de complexes de hauts poids moléculaires *in vitro*. De tels échantillons de haut poids moléculaire seront pertinents pour évaluer les gains de résolution et de sensibilité obtenus dans l'expérience CH_2-TROSY appliquée aux résidus glycine. Dans le cadre de cette étude, nous avons quantifié ces gains spectraux en utilisant en premier lieu le domaine PPIase de la protéine Pin1. La comparaison des spectres ^{1}H-^{13}C CH_2-TROSY et HSQC révèle des augmentations de 2,5 et de 2,3 en résolution et en sensibilité dès les faibles poids moléculaires à 40°C. Ces améliorations spectrales ont permis de mener une étude d'interaction avec un ligand, basée sur les

perturbations de déplacement chimique Cα et Hα des glycines, qui donnent des informations complémentaires par rapport aux groupes NH qui sont habituellement utilisés dans les expériences de titrage.

Le domaine WW, en raison de ses remarquables propriétés de cinétiques de repliement, de sa petite taille (5 kDa), des très bons rendements d'expression et de sa purification aisée par HPLC, a été utilisé pour explorer une nouvelle approche D-DNP. En effet, les températures élevées atteintes lors de la phase de dilution nécessitent une protéine capable de se replier rapidement après dénaturation. Il sera en effet important de disposer de la forme native de la protéine après l'étape d'hyperpolarisation si ces expériences peuvent aboutir à un test de criblage de ligand.

II) Sous-unité ATPasique H23

La protéine H23 est une petite sous-unité de liaison à l'ATP d'une ATPase à métaux d'*Arabidopsis Thaliana,* HMA8, qui est présente dans la lumière des thylakoïdes des chloroplastes[9]. Dans ce travail de thèse, nous nous sommes intéressés à cette protéine car elle offre de bons rendements d'expression dans des milieux acellulaires et constitue donc un modèle de choix pour des mises en place de nouvelles stratégies de marquage spécifiques. Possédant 15 glycines réparties dans différents éléments de structure secondaire, la structure RX de H23 présente six brins β, quatre hélices α et des boucles dynamiques en solution (PDB : 5LBK)[9]. Cependant, cette protéine n'avait fait l'objet d'aucune étude RMN, et, dans une étude préliminaire, nous avons entrepris l'attribution de ses résonances ^{1}H, ^{15}N et ^{13}C au niveau de son squelette peptidique. L'attribution a été obtenue à partir d'un ensemble d'expériences triple résonances: HN(CO)CACB[30], HNCACB[31], HNCO[4] et HN(CA)CO[32], enregistrées dans leur version BEST-TROSY[33,34] sur un échantillon H23 marqué U-^{15}N-^{13}C. L'attribution des déplacements chimiques H^N, N, Cα, Cβ, CO a été déterminée pour 97% des résidus. Elle est indiquée sur le spectre ^{1}H-^{15}N TROSY présenté en Figure 1. Dans un second temps, le logiciel TALOS+[35] nous a permis de prédire les éléments de structure secondaire de H23 à partir des déplacements chimiques mesurés pour les atomes H^N, N, Cα, Cβ et CO. TALOS + s'appuie sur une banque de données de protéines de référence, dont la structure et les déplacements chimiques sont connus. Les angles φ et ϕ des acides aminés sont prédits à partir des déplacements chimiques mesurés pour chaque résidu i et les résidus i+1 et i-1 qui l'entourent. Il a été montré que des déplacements chimiques similaires pour ces fragments tripeptidiques

témoignent en effet de propriétés structurales communes. TALOS+ prédit non seulement la conformation mais aussi la dynamique présente au niveau de chaque résidu. Cette analyse permet d'obtenir rapidement des éléments de comparaison entre les données RMN obtenues en solution et les structures cristallines si elles sont disponibles. Les éléments de structures secondaires peuvent être directement comparés mais aussi les données de dynamique, qui sont traduites par les facteurs de température dans les structures cristallographiques. Dans le cas l'étude de la protéine H23, la comparaison des prédictions faites par TALOS+ avec la structure déterminée par cristallographie aux rayons X montrent de faibles divergences.

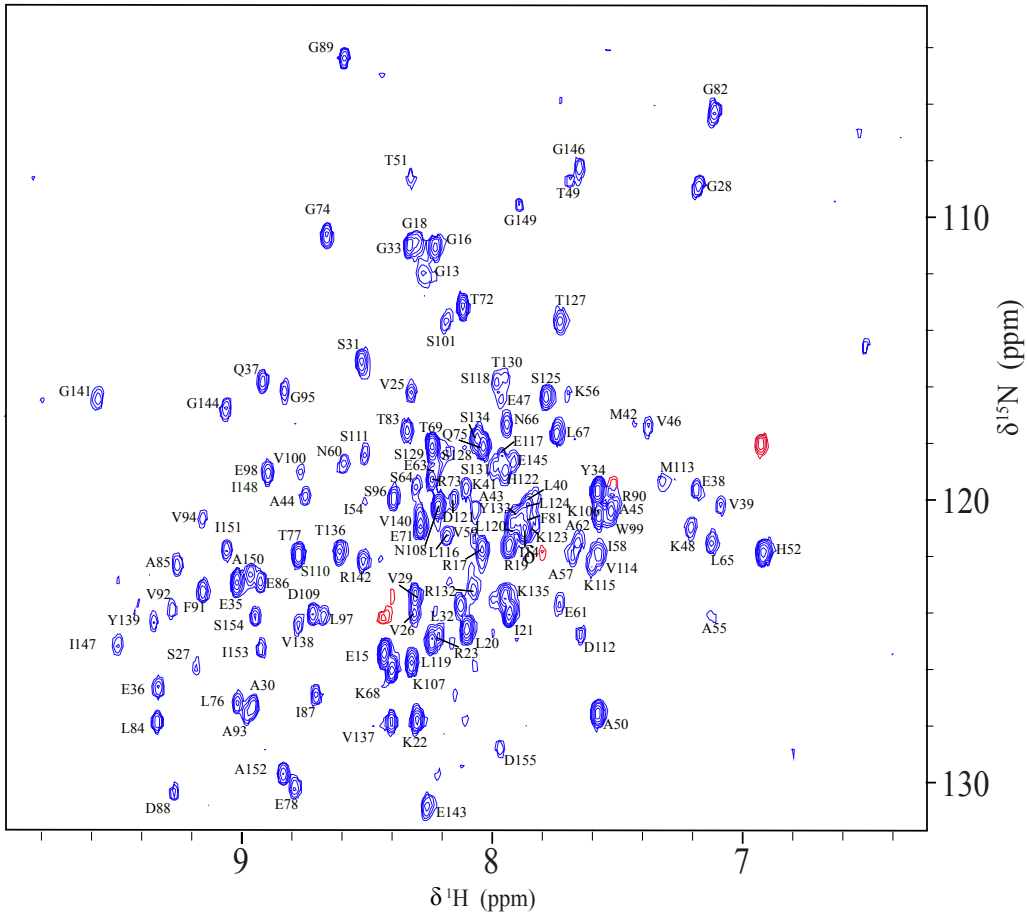


Figure 1: Spectre TROSY ^{15}N-^{1}H de la protéine H23 U ^{2}H-^{15}N-^{13}C à 850 MHz (300 µM, Phosphate 20 mM, pH 7, NaCl 100 mM, 25°C, 90% H_2O, 10% D_2O). Les corrélations non attribuées sont en rouge.

Résultats :

I) Optimisation de l'efficacité de la deutération lors des expressions de protéines dans l'eau légère et application à l'étude de protéines de haut poids moléculaire

Comme lors des expressions effectuées *in vivo*, l'extrait S30 utilisé pour l'élaboration des milieux cell-free contient des enzymes impliquées dans des voies métaboliques autres que la synthèse des protéines, qui concernent notamment le métabolisme des acides aminés. Ces enzymes sont sources de « scrambling » isotopique, notamment les transaminases qui sont directement problématiques pour la sélectivité de marquage ^{15}N des acides aminés mais aussi pour maintenir la deutération du carbone Cα lorsque l'expression a lieu dans du H_2O.

Les transaminases catalysent de manière réversible les réactions de transamination entre les acides aminés et les céto-acides, échangeant la fonction amine entre deux acides aminés, via le pyridoxal-phosphate (PLP)[36]. Elles sont responsables de la protonation Hα2 (*proR)* des positions Cα, à partir des protons présents en solution. Cette activité enzymatique limite considérablement la qualité des marquages dans l'eau lorsque la perdeutération est recherchée. Le milieu d'expression acellulaire permet cependant l'ajout de certains inhibiteurs de transaminases pour limiter le « scrambling » des acides aminés. Leur ajout évite cette incorporation de protons au niveau des positions Cα des acides aminés. Les inhibiteurs utilisés pour éviter l'activité transaminase ciblent directement le cofacteur PLP ou la base de Schiff qu'il établit au cours du mécanisme réactionnel avec l'enzyme ou l'acide aminé substrat. Dans notre protocole, nous avons utilisé trois inhibiteurs connus, l'acide aminooxyacétique (AOA), le borohydrure de sodium ($NaBH_4$) et l'acide D-malique (DM) . La source d'acides aminés marqués provient d'extraits d'algues, qui permettent d'obtenir un triple marquage ^{15}N-^{13}C-^{2}H à un coût raisonnable. L'utilisation de $NaBH_4$ pour inhiber les enzymes dépendantes du PLP réduit les rendements en protéines, parfois par plusieurs ordres de grandeurs, car ce réducteur non spécifique puissant peut également affecter l'intégrité d'autres protéines. L'optimisation du protocole cell-free a été entreprise sur la protéine H23, et a permis la quantification de la protonation résiduelle. Le protocole a ensuite été appliqué à d'autres protéines, Tet2 et la protéine Pin1 comme cela est présenté dans la suite de ce résumé.

La quantification de la protonation résiduelle a été effectuée à partir d'expériences HNCA 3D. Pour distinguer les deux groupes ^{1}H-Cα et ^{2}H-Cα au niveau de la résonance du carbone Cα, le deutérium a été découplé lors de l'édition ^{13}C, mais pas le ^{1}H. En raison d'un effet isotopique significatif sur le déplacement chimique du ^{13}C, le singulet Cα{^{2}H} est très proche de la composante « upfield » du doublet Cα-^{1}H . La composante « downfield » de ce doublet est pour sa part assez isolée et peut être intégrée séparément. Le pourcentage d'isotopes ^{1}H ou ^{2}H sur les

positions Cα a donc été déduit d'un rapport d'intégration entre les différentes corrélations extraites d'expériences HNCA 3D (Figure 2).

La présence de pics majeurs et mineur attestent de la présence de Cα protoné. La protonation résiduelle sur les sites Cα dans les protéines produites *in vitro* à l'aide des 3 inhibiteurs est réduite d'un ordre de grandeur. La protonation résiduelle maximale observée est de 10% pour quelques acides aminés, avec une moyenne globale inférieure à 5% de protonation résiduelle pour les deux modes de production, en batch ou en mode échange continu (CECF).

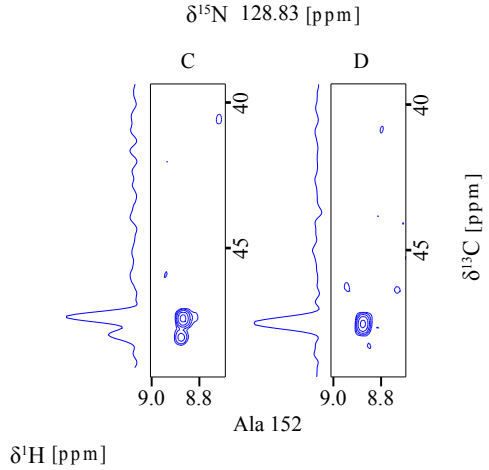


Figure 2: Bandes 2D H^N-C^α extraites du spectre 3D HNCA de H23 produite par cell free. Un marquage U- ^{15}N-^{13}C-^{2}H, dans H_2O a été effectué, sans (C) ou avec (D) 3 inhibiteurs de transaminases (AOA, DM et $NaBH_4$). 850 MHz, pH 7, tampon phosphate 30 mM, 100 mM NaCl, 25°C.

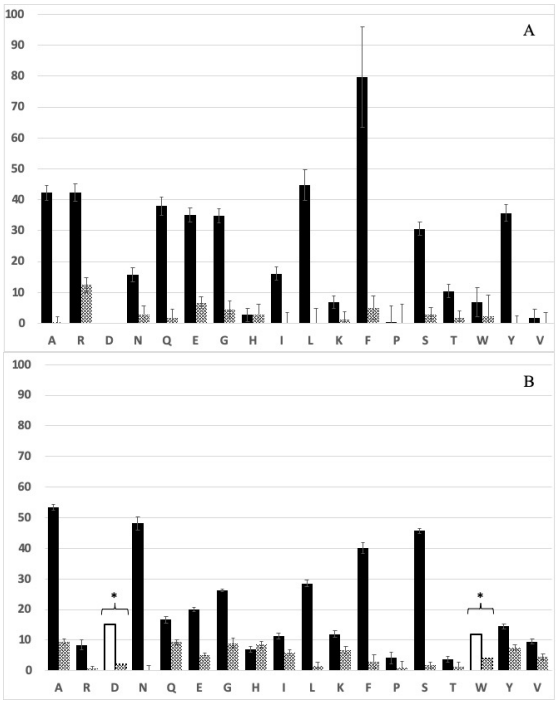


Figure 3: Niveaux de protonation résiduelle (%) détectées sur la position Cα de la protéine H23**.** Les expressions *in vitro* ont été faites dans $^{1}H_2O$ suivant des protocoles permettant d'obtenir un marquage U-[^{2}H, ^{13}C, ^{15}N]. Un mode batch (A) ou CECF (B) a été utilisé, en absence d'inhibiteurs (noir) ou en présence d'inhibiteurs (AOA, D-malate et $NaBH_4$) (gris). Pour chaque type de résidu, les pourcentages de protonation résiduelle affichés proviennent d'une moyenne calculées sur 2 à 7 acides aminés (sauf pour l'histidine dans les deux modes de production, la phénylalanine et le tryptophane en mode batch et la proline en mode CECF, pour lesquels un seul acide aminé était analysable). La quantification n'a pas été entreprise pour la méthionine et l'acide aspartique car les corrélations se chevauchaient. Les données pour les acides aminés signalés par un astérisque (*) ont été tirées de Otting et al, 2011, Journal of Biomolecular NMR. Il n'y a pas de quantification pour la cystéine car ce résidu est absent dans la séquence de H23.

Afin d'apprécier l'utilité de cette stratégie de marquage ^{2}H, l'application du protocole cell-free à la protéine TET2 (468 kDa) a été entrepris dans deux milieux différents : dans du H_2O ou dans du D_2O (travail réalisé à l'IBS par nos collaborateurs). Après purification, la protéine exprimée dans l'eau lourde a été dialysée contre un tampon 90% H_2O pour la réalisation du tube RMN. La protéine exprimée dans l'eau légère a été utilisée pour réaliser deux échantillons RMN, utilisant un tampon 100% D_2O ou 90% H_2O. Les spectre ^{15}N-^{1}H-CRINEPT-HMQC-TROSY[38] obtenus sont représentés en Figure 4.

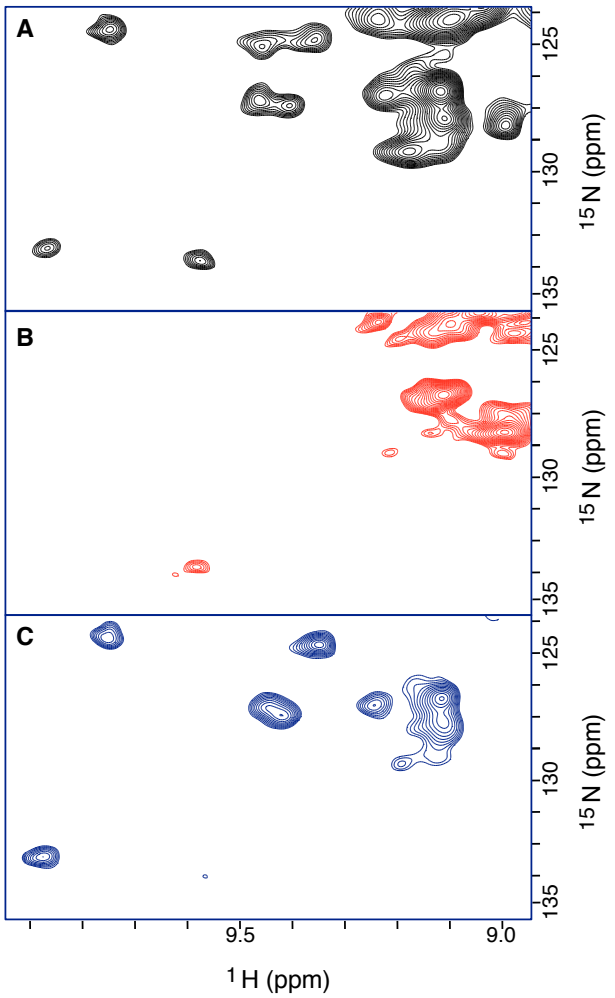


Figure 4: extraits 2D-(^{15}N,^{1}H) des spectres 1H-15N de la protéine TET2 utilisant l'expérience CRINEPT-HMQC-TROSY (Riek, PNAS 1999). Les données ont été enregistrées à 950 MHz sur un spectromètre Bruker Avance III HD, équipé avec une sonde cryogénique 5 mm triple résonances. Le marquage uniforme U-[^{2}H,^{15}N] de la protéine Tet2 a été obtenu *in vitro* (CECF mode), Des concentrations de 10 µM (200 µL, tubes Shigemi 4 mm) ont été utilisées pour les échantillons RMN, dans du tampon TRIS 20 mM (pH 7.4), 20 mM NaCl. Les données ont été enregistrées à 50°C. (A) échantillon produit dans $^{1}H_2O$ avec du Celtone® U- [^{15}N, ^{2}H] hydrolysé, complété par des acides aminés W et C non marqués comme sources d'acides aminés. L'échantillon (B) a été produit en utilisant la même source d'acides aminés mais dans de l'eau lourde D_2O. La protéine purifiée a ensuite été dialysée dans un tampon H_2O et stockée à température ambiante pendant 10 jours pour permettre les éventuelles substitutions ^{2}H->^{1}H. L'échantillon (C) a été produit selon le protocole utilisé pour l'échantillon (A), mais la protéine purifiée a été dialysée contre un tampon D_2O et stockée à température ambiante pendant 10 jours pour permettre les échanges ^{1}H-> ^{2}H. Seules les positions amides protégées du solvant sont visibles sur ce spectre C.

Ces données spectrales nous permettent d'illustrer clairement l'utilité de ce protocole de culture cell-free optimisé pour la deutération. En premier lieu, l'utilisation d'inhibiteurs de transaminases permet de réduire à moins de 5% en moyenne la protonation résiduelle. Ce taux de deutération très élevé est essentiel pour limiter au maximum la relaxation induite par les protons et obtenir des spectres ^{1}H-^{15}N de qualité pour une protéine de masse moléculaire aussi importante que TET2. Le spectre A permet de visualiser distinctement des corrélations ^{1}H-^{15}N dans la zone 9-10 ppm dans la dimension ^{1}H, qui rassemble habituellement des groupes amides de régions très structurées. Le spectre B correspond à la même région pour un échantillon

produit dans du D_2O. Une dialyse dans H_2O permet l'échange D/H pour les groupes amides accessibles au solvant, certaines corrélations ne sont cependant pas visibles sur ce spectre ce qui indique que les amides correspondant sont ne sont pas en contact avec le solvant, même après un long temps d'incubation de 10 jours. Un tel ralentissement de l'échange H/D est fréquent au sein des cœurs hydrophobes de protéines de haut poids moléculaire. Pour visualiser à nouveau ces résonances, une étape de dénaturation dans H_2O serait nécessaire, et impliquerait bien sûr une étape de renaturation, à l'issue incertaine. Enfin, L'image C correspond à la protéine TET2 exprimée dans de l'eau légère H_2O mais dialysée contre un solvant D_2O. Il s'agit en définitive d'un spectre « négatif » par rapport au précédent où toutes les corrélations qui étaient non visibles sont à présent les seules apparentes. Cela s'explique car les acides aminés accessibles au solvant deutéré disparaissent alors du spectre ^{1}H-^{15}N. A l'inverse, la protonation est conservée au sein des cœurs hydrophobes malgré l'utilisation d'eau lourde. Une superposition des images B et C nous permettrait de retrouver le spectre A. On comprend donc bien l'intérêt du marquage cell free pour la deutération de protéines de haut poids moléculaire. Une expression dans *E. coli* utilisant H_2O comme solvant n'aurait pas permis d'obtenir un taux de deutération satisfaisant, même en ajoutant une source d'acides aminés deutérés, en raison des conversions enzymatiques nombreuses au sein des cellules, en particulier des activités transaminases. Une culture dans *E. coli* utilisant D_2O comme solvant aurait assuré une deutération complète mais n'aurait pas permis d'obtenir un échange H/D de la totalité des groupes amides. Le protocole cell-free décrit dans ce travail de thèse permet en revanche de mener l'expression de protéines hautement deutérées dans l'eau. Cette approche garantit de pouvoir observer tous les signaux RMN des groupes amides, en préservant un taux de deutération très important au niveau des chaînes carbonées ce qui est nécessaire pour les objets de haut poids moléculaire.

II) Optimisation du marquage spécifique ^{1}H,^{15}N, ^{13}C Glycine, ^{2}H,^{14}N,^{12}C non-Glycine et application à des expériences RMN dédiées aux résidus glycines

L'activité transaminase décrite précédemment doit être également contrôlée pour le marquage spécifique des résidus glycines. Son action conduirait à échanger l'azote 15 porté uniquement par les glycines avec les azotes 14 des autres résidus, et inversement. D'autre part, le proton $^{1}H\alpha 2$ de la glycine pourrait être remplacé par un deutérium du solvant dans le cas d'une

expression dans D_2O, ou, dans le cas d'une expression dans l'eau légère, ce serait les deutériums $^{2}H\alpha 2$ des autres acides qui seraient échangés par des protons, faisant ainsi baisser considérablement le taux de deutération final recherché. Parmi les enzymes métabolisant la glycine, La sérine hydroxyméthyl-transférase (glyA) est également présente de manière significative dans le milieu S30. Son action conduit à un échange réversible de la chaîne latérale de la sérine avec un hydrogène du solvant. La production de glycine à partir de sérine ^{2}H, ^{14}N, ^{12}C conduirait donc un marquage non souhaité au niveau des hétéroatomes mais aussi à une paire $^{2}H\alpha 2$-$^{1}H\alpha 3$ non exploitable pour les expériences CH_2-TROSY que nous souhaitons développer. Par ailleurs, l'action de glyA conduirait également à la production de sérines ^{1}H, ^{15}N, ^{13}C à partir des glycines introduites[39]. Son activité peut être éliminée ou limitée très fortement par l'utilisation des 3 inhibiteurs AOA, DM et $NaBH_4$, car, à l'instar des transaminases, glyA est PLP dépendante. Des spectres ^{15}N-^{1}H HSQC ont été enregistrés sur des échantillons de H23 exprimés dans un milieu acellulaire avec ou sans ces trois inhibiteurs. Le marquage spécifique des résidus glycine a bien été mis en évidence, ainsi que l'inhibition de l'enzyme glyA (Figure 5). L'absence d'inhibiteurs a conduit à un marquage non spécifique d'autres acides aminés (principalement des sérines). Les faibles corrélations correspondantes ne sont plus visibles dans l'échantillon produit en présence d'inhibiteurs. Les inhibiteurs ont été également jugés importants pour ne pas observer le transfert partiel des isotopes ^{12}C, ^{14}N et ^{2}H des autres acides aminés sur les résidus glycine (« scrambling » isotopique).

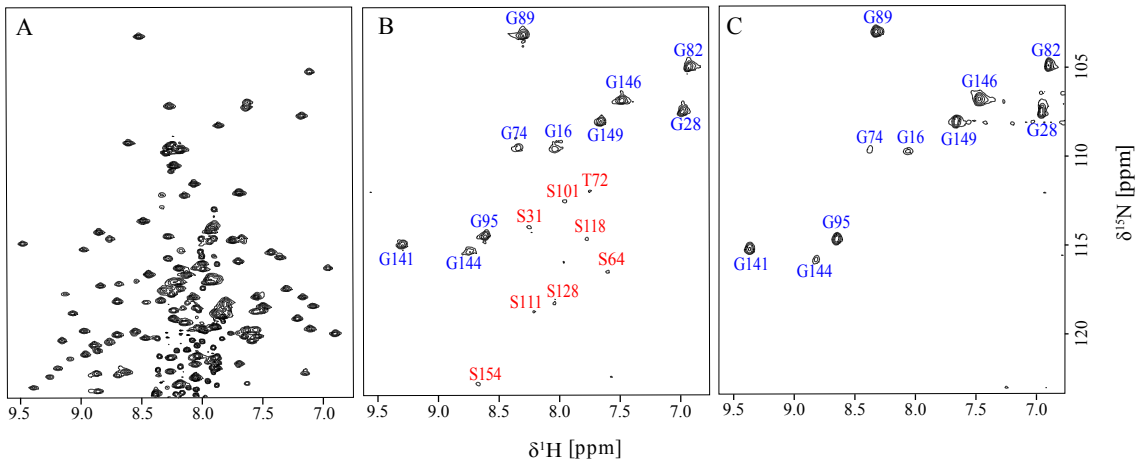


Figure 5 : Spectres ^{1}H-^{15}N HSQC de la protéine H23 exprimée par cell-free. A) U-^{15}N, 200 µM, tampon phosphate 20 mM, pH 4,5, NaCl 100 mM, 25°C, 700 MHz. B) Gly-^{1}H-^{15}N-^{13}C, autres acides aminés U-^{2}H-^{14}N-^{12}C, sans inhibiteurs. C) Gly-^{1}H-^{15}N-^{13}C, autres acides aminés U-^{2}H-^{14}N-^{12}C, avec inhibiteurs 20 mM DM, 20 mM AOA et $NaBH_4$. Concentration H23 300 µM pour B et 330 µM pour C, tampon phosphate 20 mM, pH 6,5, NaCl 100 mM, 0,11 mM DSS, 10% D_2O, 500 MHz, 40°C.

Un protocole cell-free a pu être établi pour obtenir un marquage spécifique de très bonne qualité pour les résidus Glycine, tout en assurant un rendement d'expression correct. L'expression dans

H_2O a été préférée à une expression dans D_2O pour des raisons de rendements. Les échantillons spécifiquement marqués sur les résidus glycine ont été utilisés pour des quantifications en RMN. En particulier, les gains qu'ils apportent dans l'expérience CH_2-TROSY en termes de sensibilité et de résolution ont été déterminés. Les mesures ont été faites de manière comparative avec un échantillon uniformément marqué U-^{1}H,^{15}N,^{13}C et avec l'expérience CH_2-HSQC, dédiée au groupes méthylène mais non-optimisée du point de vue de la relaxation. Les spectres CH_2-HSQC et CH_2-TROSY ont été enregistrés dans deux solvants différents : H_2O ou, après lyophilisation, dans 100% D_2O et à trois températures différentes (40°C, 25°C et 10°C) pour chacun des échantillons, uniformément marqué sur les carbones et azotes U-^{15}N,^{13}C, ou spécifiquement marqué ^{1}H,^{15}N,^{13}C-Gly et ^{2}H,^{14}N,^{12}C -pour les autres acides aminés.

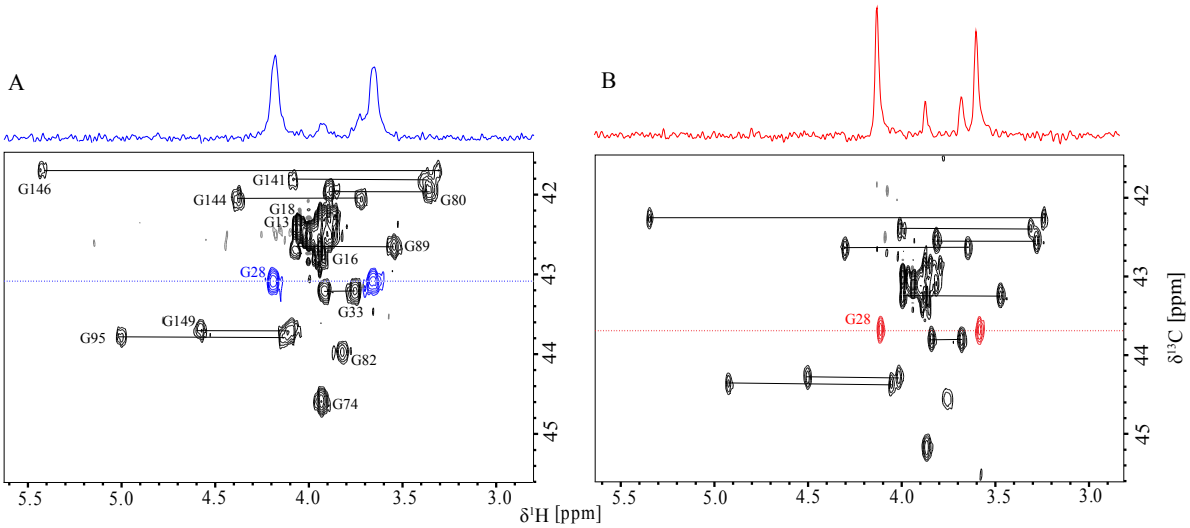


Figure 6 : spectres CH_2-HSQC (A) et CH_2-TROSY (B) de H23. Les traces 1D proton correspondant aux lignes pointillées apparaissent au-dessus des spectres 2D. L'échantillon est marqué Gly-^{1}H-^{15}N-^{13}C- et U-^{2}H-^{15}N pour les autres acides aminés. Ce marquage spécifique a été obtenu par expression cell-free en utilisant le protocole que nous avons établi. 400µM, phosphate 20 mM, pH 7,0, 100 mM NaCl, 1 mM DTT, 100% D_2O. 950 MHz, 25°C.

La Figure 6 présente un spectre CH_2-HSQC et un spectre CH_2-TROSY, enregistrés à 25° C pour l'échantillon spécifiquement marqué Gly dans 100% D_2O. La qualité spectrale obtenue grâce à l'utilisation du CH_2-TROSY permet de limiter les zones de recouvrement nombreuses qui sont observées sur une spectre de corrélation ^{1}H-^{13}C classique. L'ensemble des couples de protons diastéréotopes est visible, avec de surcroît, une sensibilité nettement accrue.

La quantification du rapport signal sur bruit (S/N) et la largeur de raies à mi-hauteur des corrélations ont été entreprises sur 9 paires de protons diastéréotopes des résidus glycine. Nous avons calculé les « améliorations spectrales » observées entre les spectres TROSY et les spectres HSQC pour plusieurs types de marquage à plusieurs températures. Pour cela, nous avons effectué les rapports des largeurs de raies et des sensibilités S/N sur chaque résonance

des résidus glycines enregistrée en version TROSY ou HSQC. Les gains en sensibilité et en résolution sont rapportés en fonction du paramètre d_A, défini comme la distance moyenne apparente des protons voisins(voir figure 7). Comme attendu, il apparaît clairement que ces améliorations spectrales augmentent en fonction de d_A, c'est-à-dire lorsque le taux de deutération est plus élevé. Que ce soit au niveau de la sensibilité ou de la résolution, des gains d'un facteur 2 à 2,5 sont observés en moyenne lorsque les échantillons sont spécifiquement marqués sur les glycines. Alors que les améliorations les plus grandes ont été obtenues pour l'échantillon spécifiquement marqué glycine dans D_2O (correspondant à un d_A moyen de 3.75 Å), l'échange H/D des protons HN à l'aide d'un tampon deutéré permet, à lui seul, une amélioration significative des paramètres spectraux dans des échantillons uniformément marqués. Dans les structures des protéines, le proton HN est en effet très souvent celui qui est le plus proche des protons Hα de la glycine. A ce titre, il représente une source importante de relaxation des groupes méthylènes. Lorsque le temps de corrélation τ_C diminue (augmentation de la température), les améliorations spectrales sont plus marquées passant d'une moyenne de 2,2 et 1,8 respectivement pour la résolution et le S/N à 10°C à 2,5 et 2,3 à 40°C pour un échantillon spécifiquement marqué dans du D_2O. Ces résultats attestent des gains offerts par l'expérience CH_2-TROSY à la fois au niveau de la sensibilité et de la résolution dans la dimension proton pour des échantillons hautement deutérés. Toutefois, ces gains apparaissent plus modestes à mesure que la température diminue (que τ_C augmente) ce qui semble indiquer la limite de cette stratégie pour l'étude de complexes de haut poids moléculaire. Les bénéfices spectraux observés pour des protéines de poids moléculaire intermédiaires peuvent cependant être exploités. Nous avons ainsi proposé de caractériser une interaction protéine ligand à partir d'expériences de titrage suivies par spectroscopie CH_2-TROSY. En mesurant précisément les perturbations de déplacement chimiques des résidus glycine de la protéine Pin1, nous avons pu obtenir des mesures de constantes d'affinité comparables à celles calculées plus classiquement à l'aide de spectres ^{1}H-^{15}N HSQC. Une telle titration, réalisée à partir des résonances des protons méthylène, n'aurait pas été précise en utilisant une expériences ^{1}H-^{13}C HSQC classique en raison de nombreux recouvrement de résonances d'une part, et du manque de résolution d'autre part car les perturbations de déplacements chimiques mesurées sur les méthylènes des glycines sont beaucoup plus faibles en intensité que sur les groupes amides. Une très grande précision de mesure est donc nécessaire.

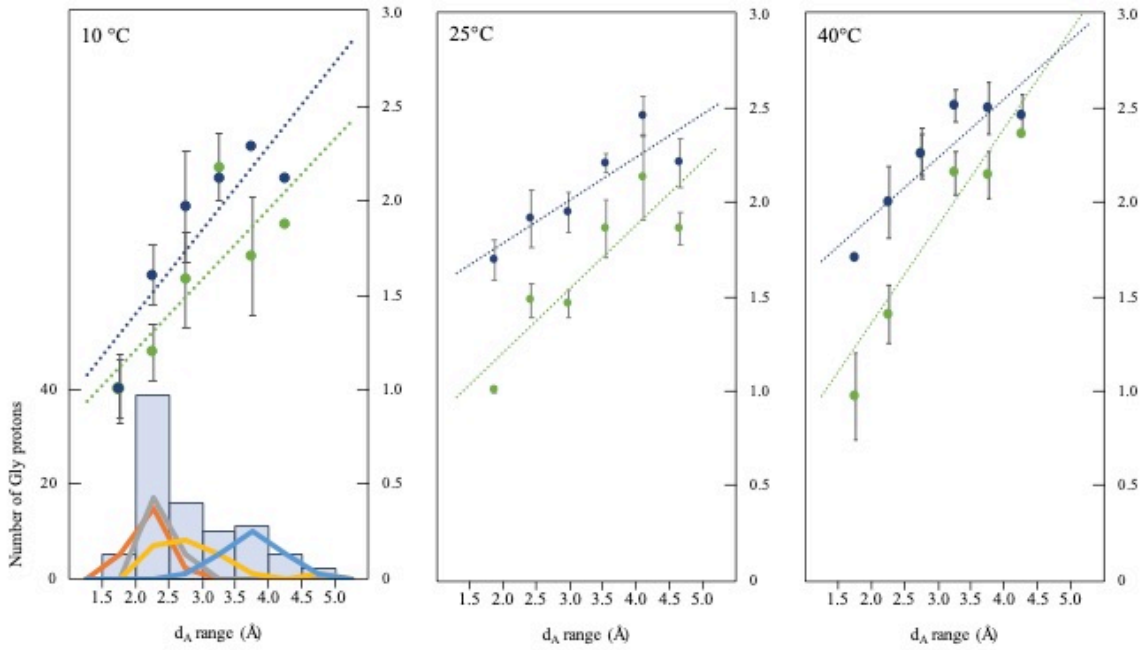


Figure 7 : Améliorations des caractéristiques spectrales (S/N et résolution) des protons diastéréotopes des Gly en fonction du d_A. Pour chaque gamme de d_A, les points verts correspondent aux améliorations S/N, les points bleu foncé aux améliorations de résolution. Les mesures ont été faites à trois températures, 10°C (panneau de gauche), 25°C (panneau central) et 40°C (panneau de droite). Sur le panneau de gauche, l'histogramme représente la distribution des distances d_A sur l'ensemble des mesures effectuées : pour près de 40 Hα Gly, le d_A est compris entre 2 et 2,5 Å. Les courbes en couleurs détaillent ces distributions pour chaque type de marquage : en orange, pour un marquage uniforme dans H_2O, en gris pour un échantillon Gly spécifiquement marqué dans H_2O, en jaune pour un marquage uniforme U-^{1}H-^{13}C-^{15}N dans D_2O et en bleu clair pour un marquage spécifique Gly dans D_2O.

Des expériences sélectives d'états de spins 3D ont par ailleurs été imaginées à partir de cette expériences CH_2-TROSY. Ces expériences sélectives d'état de spin permettent l'expression de nombreux couplages scalaires entre les différents noyaux du système. La mesure de ces constantes de couplage donne accès à des informations précises sur la structure et la dynamique du résidu, agissant ainsi comme une sonde locale dans la protéine.

III) Optimisation du T_1 des protons des protéines pour la DNP

La relaxation des spins est influencée par leur environnement moléculaire. Le marquage isotopique au deutérium d'un échantillon protéique peut considérablement allonger les temps de relaxation T_1 et T_2, notamment par suppression des interactions dipolaires, augmentant les gains de sensibilité que la DNP peut offrir. Les gains obtenus pourraient améliorer la polarisation D-DNP ^{1}H pour des études dynamiques et d'interactions protéine/ligand et peut-être permettre d'explorer de nouvelles techniques de criblage. Ainsi, l'objectif de cette partie de ma thèse est d'utiliser le domaine WW, exprimé avec des densités de protons différentes placées

de manière éparpillée et aléatoire dans un environnement perdeutéré, pour obtenir des temps des relaxation T_1 plus longs des protons afin de pouvoir maintenir leur état hyperpolarisé. En effet, pour que l'hyperpolarisation subsiste dans le processus de dissolution et de transfert du polariseur au spectromètre, le T_1 des protons individuels doit être supérieur à quelques secondes, ce qui n'est pas le cas pour les protéines entièrement protonées.

Pour mesurer le T_1, une série de 14 expériences avec un délais TI (temps d'inversion) croissant ont été enregistrées sur des échantillons avec des densités croissantes de deutération (0, 80.1%, 90.6% et 92.6%). L'ajustement des courbes à été effectué pour différents protons ou groupes de protons du domaine WW. Le graphique ci-dessous reprend les valeurs de T_1 trouvées pour différents protons suivis lors de marquages à densité variable en protons.

T1 sec	100% ^{1}H	ERROR	80.1% ^{2}H	ERROR	90.6% ^{2}H	ERROR	92.6% ^{2}H	ERROR
DSS	3,00	0,07	3,59	0,10	3,40	0,09	3,50	6,37
METHYL 1	0,54	0,01	0,93	0,03	1,12	2,22	1,34	2,69
METHYL 2	0,79	0,03	1,86	0,20	2,00	0,30		
METHYL 3	0,54	0,01	1,70	0,20	3,22	6,75	5,32	10,65
ALI 1	0,72	0,02	1,81	0,20	1,86	0,57	5,85	11,69
ALI 2	0,68	0,02	1,77	0,24	1,08	2,24	3,98	8,03
ALI 3	0,51	0,01	0,77	0,02	0,99	2,03	5,99	12,41
AROM 1	1,45	0,16	3,31	0,60	10,29	20,95		
AROM 2	1,48	0,05	3,92	0,14	8,26	0,43		
AROM 3	1,87	3,49	4,26	0,35	11,40	0,89		
AROM 4	1,64	0,34	3,18	0,52	10,25	19,31		

Table 1 : Table the mesures de T_1 pour différentes résonances proton des 4 échantillon de WW marqués avec des concentrations différentes en deutérium

Pour l'échantillon WW non marqué, où seuls les groupes amide sont deutérés, les temps de relaxation varient entre 0,5 s pour les protons relaxants les plus rapidement (ALI 3) à 1,87 s pour les plus lents (AROM 3). Les protons aliphatiques ont les temps de relaxation les plus courts, car ils sont entourés par de nombreux protons, étant au cœur des chaînes latérales, donc très sujets aux interactions dipolaires. Il est à noter que ces temps de relaxation T_1 sont déjà allongés car les amides sont deutérés, mais, dans un tampon utilisant de l'eau légère, les temps de relaxation seraient encore plus faibles. Au-delà d'une deutération de 80%, les valeurs de T_1 font plus que doubler en moyenne (bien qu'avec une certaine disparité, presque triplant pour le MÉTHYL 3). Une augmentation du facteur 8 est visible pour l'échantillon deutéré à 92.6%. Cependant, à 92.6% de deutération, le signal devient faible pour certaines résonances et le T_1 n'a pas pu être déterminé de manière adéquate. Les temps T_1 obtenus dans ces conditions sont

compatibles avec des études par D-DNP, car ils permettent de conserver l'hyperpolarisation pour des études de dynamique ou interactions protéine/ligand.

Conclusion

Au-delà du développement de l'appareillage (intensité du champ magnétique, miniaturisation et/ou facteurs de qualités des sondes) pour augmenter la qualité des signaux RMN, il y a un enjeu à travailler sur l'échantillon lui-même. Des progrès importants ont été effectués grâce à de nouvelles stratégies de marquages (deutération, marquages spécifiques, marquages segmentaux) et aux travaux réalisés pour faciliter leur mise en œuvre. Les protéines deutérées présentant uniquement quelques protons sur des groupes méthyles permettent par exemple d'enregistrer des spectres méthyle-TROSY de grande qualité, y compris pour des protéines de très haut poids moléculaire. Trois nouvelles stratégies de marquage ont été explorées ici pour des utilisations RMN différentes. Dans un premier temps, l'expression cell-free a été employée pour assurer la perdeutération des protéines tout en conservant la protonation des groupements amides. Des échantillons ainsi produits permettent l'enregistrement d'expériences NH-TROSY sans observer la disparition des corrélations correspondant à certains résidus hydrophobes. En effet, dans le cas des protéines de poids moléculaire important, l'échange H/D nécessaire après production en milieu deutéré dans la bactérie est extrêmement ralenti au sein des cœurs hydrophobes.

L'obtention d'un échantillon marqué sélectivement sur un acide aminé dans un environnement perdeutéré permet le développement d'expériences dédiées. Dans ce cadre, nous avons entrepris le marquage spécifique des glycines. L'apport de ces échantillons pour les expériences de corrélation ^{1}H-^{13}C CH_2-TROSY a été étudié. Comparé aux spectres HSQC classique, ces échantillons ont permis d'observer des gains en résolution et sensibilité de 2.5. Les améliorations spectrales ont été avantageusement utilisées pour réaliser des études d'interactions, de manière complémentaire aux études ^{15}N-^{1}H habituelles. De plus, cette stratégie de marquage a permis de réaliser des mesures de couplage inédites au niveau des résidus glycine de la protéine Pin1. Ces mesures, basées sur une expérience 3D dérivée de l'expérience CH_2-TROSY, a permis d'obtenir des informations structurales au niveau du squelette peptidique de la protéine Pin1. Contrairement aux expériences CH_3-TROSY, les gains en résolution et sensibilité de l'expérience CH_2-TROSY deviennent plus modestes à mesure

que la masse moléculaire augmente ce qui n'a pas permis d'obtenir des données exploitables sur des échantillons de haut poids moléculaire. La dynamique restreinte liée à la localisation du méthylène des glycines directement sur le squelette et l'autorelaxation importante liée à la proximité des deux protons expliquent les limites de cette approche, même pour des échantillons spécifiquement marqués. Toutefois, le groupe méthylène des résidus glycine représente une sonde locale intéressante, abondante dans les protéines et placée souvent au niveau de boucles ou en limite des éléments de structure secondaires. Il est possible d'extraire un nombre important de constantes de couplages à partir des expériences sélectives d'état de spin que nous avons décrites pour obtenir des informations structurales et dynamiques.

Enfin, dans l'optique d'appliquer la technique d'hyperpolarisation D-DNP aux protéines, nous avons utilisé le domaine WW de la protéine Pin1 pour réaliser des expériences préliminaires. La deutération partielle de ce domaine permet d'obtenir des augmentations importantes des temps de relaxations longitudinaux T_1 des protons résiduels. Nous avons mis au point des conditions de « DNP-juice » permettant de créer des états hyperpolarisés de ce domaine WW. La technique de dissolution reste pour le moment difficile à mettre en œuvre pour un échantillon protéique mais les efforts seront poursuivis au laboratoire pour permettre l'observation en solution de ce domaine hyperpolarisé.

1. Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **101,** 760–762 (1979).
2. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69,** 185–189 (1980).
3. Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29,** 4659–4667 (1990).
4. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89,** 496–514 (1990).
5. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94,** 12366–12371 (1997).
6. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced ^{1}H–^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society* **125,** 10420–10428 (2003).
7. Amero, C., Schanda, P., Durá, M. A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B. & Boisbouvier, J. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *Journal of the American Chemical Society* **131,** 3448–3449 (2009).
8. Miclet, E., Williams, D. C., Clore, G. M., Bryce, D. L., Boisbouvier, J. & Bax, A. Relaxation-Optimized NMR Spectroscopy of Methylene Groups in Proteins and Nucleic Acids. *J. Am. Chem. Soc.* **126,** 10560–10570 (2004).
9. Mayerhofer, H., Sautron, E., Rolland, N., Catty, P., Seigneurin-Berny, D., Pebay-Peyroula, E. & Ravaud, S. Structural Insights into the Nucleotide-Binding Domains of the P1B-type ATPases HMA6 and HMA8 from Arabidopsis thaliana. *PLOS ONE* **11,** e0165666 (2016).
10. Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380,** 544–547 (1996).
11. Ardenkjær-Larsen, J. H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M. H., Servin, R., Thaning, M. & Golman, K. Increase in Signal-to-Noise Ratio of >10,000 Times in Liquid-State NMR. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 10158–10163 (2003).
12. Ragavan, M., Chen, H.-Y., Sekar, G. & Hilty, C. Solution NMR of Polypeptides Hyperpolarized by Dynamic Nuclear Polarization. *Anal. Chem.* **83,** 6054–6059 (2011).
13. Jannin, S., Dumez, J.-N., Giraudeau, P. & Kurzbach, D. Application and methodology of dissolution dynamic nuclear polarization in physical, chemical and biological contexts. *Journal of Magnetic Resonance* **305,** 41–50 (2019).
14. Movellan, K. T. Alpha protons as NMR probes in deuterated proteins. *Journal of Biomolecular NMR* 11 (2019).
15. Zubay, G. In Vitro Synthesis of Protein in Microbial Systems. *Annual Review of Genetics* **7,** 267–287 (1973).
16. Roberts, B. E. & Paterson, B. M. Efficient Translation of Tobacco Mosaic Virus RNA and Rabbit Globin 9S^{43,47,48} RNA in a Cell-Free System from Commercial Wheat Germ. *Proceedings of the National Academy of Sciences* **70,** 2330–2334 (1973).
17. Su, X.-C., Loh, C.-T., Qi, R. & Otting, G. Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective 15N-labelling and production of perdeuterated proteins in H2O. *J Biomol NMR* **50,** 35–42 (2011).
18. Hoffmann, B., Löhr, F., Laguerre, A., Bernhard, F. & Dötsch, V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Progress in Nuclear Magnetic Resonance Spectroscopy* **105,** 1–22 (2018).
19. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role of cis peptide bonds in protein structures. *Journal of Molecular Biology* **214,** 253–260 (1990).
20. Weiss, M. S., Jabs, A. & Hilgenfeld, R. Peptide bonds revisited. *Nat Struct Mol Biol* **5,** 676–676 (1998).
21. Grathwohl, C. & Wüthrich, K. The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides: X-PRO PEPTIDE BOND. *Biopolymers* **15,** 2025–2041 (1976).
22. Grathwohl, C. & Wüthrich, K. NMR studies of the rates of proline *cis* - *trans* isomerization in oligopeptides: rates of *cis* - *trans* isomerization. *Biopolymers* **20,** 2623–2633 (1981).
23. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3,** 619–629 (2007).
24. Namanja, A. T., Wang, X. J., Xu, B., Mercedes-Camacho, A. Y., Wilson, K. A., Etzkorn, F. A. & Peng, J. W. Stereospecific gating of functional motions in Pin1. *Proceedings of the National Academy of Sciences*

108, 12289–12294 (2011).
25. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. Structure-function-folding relationship in a WW domain. *Proceedings of the National Academy of Sciences* **103,** 10648–10653 (2006).
26. Jäger, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. The folding mechanism of a β-sheet: the WW domain. *Journal of Molecular Biology* **311,** 373–393 (2001).
27. Lee, Y. M. & Liou, Y.-C. Gears-In-Motion: The Interplay of WW and PPIase Domains in Pin1. *Front. Oncol.* **8,** 469 (2018).
28. Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell* **89,** 875–886 (1997).
29. Smith, M. H. The Amino Acid Composition of Proteins. *J. Theoret. Biol.* **13,** 261–282 (1966).
30. Grzesiekt, S. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *Journal of American Chemical Society* **114,** 6291–6293 (1992).
31. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99,** 201–207 (1992).
32. Clubb, R. T. A Constant-Time Three-Dimensional Triple Resonance Pulse Scheme to Correlate Intra-residue HN, 15N, and 13C'Chemical Shifts in 15N- 13C-Labeled Proteins. 5
33. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *Journal of American Chemical Society* **128,** 9042–9043 (2006).
34. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance* 7 (2007).
35. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44,** 213–223 (2009).
36. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814,** 1407–1418 (2011).
37. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6,** 277–293 (1995).
38. Riek, R., Wider, G., Pervushin, K. & Wuthrich, K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences* **96,** 4918–4923 (1999).
39. Schirch, V. & Szebenyi, D. M. Serine hydroxymethyltransferase revisited. *Current Opinion in Chemical Biology* **9,** 482–487 (2005).

Introduction

Nuclear Magnetic Resonance (NMR) is a physical phenomenon observed when a nucleus, excited by a radiofrequency pulse, liberates the absorbed energy at a specific frequency. The signal generated offers information about the nucleus' interactions with other nuclei or electrons of its environment. The discovery of this phenomenon was attributed in 1938 to Isidor Rabi[1], who observed absorption and reemission of energy by passing a stream of LiCl salts into a magnetic field (0.2T). Later, in 1945, two independent teams led by Felix Bloch and Edward Mills Purcell demonstrated NMR on water (liquid state) and paraffin (solid state), respectively[2,3]. Since, numerous instrumental and theoretical developments have been made, notably to improve signal quality. NMR spectroscopy has become an immensely versatile tool for the study of molecules, with abundant applications in physical, chemical and biological sciences. In structural biology, solution NMR spectroscopy is now indispensable to elucidate the structure and time-dependent changes in conformations of biomolecules at the atomic level, pushing the limits of our understanding of biomolecular processes. In addition, thermodynamic and kinetic parameters of biomolecular interactions have allowed an improved understanding of structure, dynamic and function of proteins and oligonucleotides.

The greatest limitations when studying proteins, and more generally biomolecules of large molecular mass (>50 kDa) by liquid state NMR, are the low sensitivity and resolution of acquired signals. In the last 3 decades, many improvements have been made to address these drawbacks. Until the 80's, proton NMR was routinely used for protein studies using 2D NMR experiments (COSY, TOCSY, NOESY)[4]. Widespread signal overlap precluded their structural analysis, limiting homonuclear study to small sized proteins[5]. The use of isotopic labeling strategies[6], and the development of polarization transfer techniques[7] led to new heteronuclear experiments such as the two dimensional Heteronuclear Single Quantum Coherence Spectroscopy (HSQC[8]) as well as new experiments offering a third heteronuclear dimension to the traditional ^{1}H-^{1}H TOCSY and NOESY and thus strongly reducing spectral overlaps. Triple resonance experiments such as HNCO or HNCA[9,10], requiring double isotopic labeling to have three active nuclei for NMR, radically improved the understanding of protein structures by offering new spin networks to explore.

Lack of sensitivity and resolution still limited protein study to small proteins (<20 kDa)[11] although a number had had their structure determined. The Transverse Relaxation Optimized SpectroscopY (TROSY) experiments, originally developed for amides[12] in 1997, have greatly expanded the possibilities of protein NMR studies. These 2D Heteronuclear correlation experiments rely on the edition of the slowest relaxing component of a multiplet that result from a two-spin coupled system, whereas the HSQC or HMQC creates a mean of all the multiplet components. After the demonstration on ^{15}N-^{1}H amide pairs, the TROSY concept was applied to many other chemical groups, and notably methyls[13]. TROSY experiments greatly benefit from maximum perdeuteration around the spin system of interest to attain their full potential. In the middle of the 2000's, selective isotopic labeling of chemical groups and perdeuteration pushed the limits of protein study size, improving both sensitivity and resolution of signals with the concomitant use of these particular pulse sequences. Selective isotopic labeling strategies for the groups of interest of the respective TROSY experiments were often developed in parallel to these, so as to enhance to the maximum their signal and resolution gains. Nowadays, combining methyl-TROSY, selective isotopic labeling and very high field NMR spectrometers, researchers are able to obtain well resolved signals of proteins up to 1 mega Dalton[14]. Preserving the $^{15}N^{1}H$ pairs of amides in large deuterated proteins, however, remains somewhat challenging. Thus, the first goal of the thesis is to optimize a protein expression protocol for perdeuteration of a large protein, TET2, of almost half a mega Dalton with selective ^{15}N-^{1}H labeling of amide groups only, whereas the Cα positions remain deuterated as demonstrated through a quantitative analysis of residual protonation after expression, developed in Chapter III.

Among all the TROSY experiments available, one of the least exploited is the methylene TROSY[14]. Compared to the amide and methyl groups, and despite methylenes representing 48% of protons in proteins, there has been no labeling strategy yet published that enables specific labeling of methylenes in perdeuterated proteins. The second goal of this thesis looks at how an optimized expression protocol, allowing the labeling of the two diastereotopic protons of the methylene of glycine in a perdeuterated background, can be beneficial for the CH_2-TROSY experiment, so as to exploit its full potential of signal and resolution enhancement. Optimization of the cell-free expression protocol was undertaken on the Nucleotide-Binding Domain of the P1B-type ATPase HMA8 from *Arabidopsis thaliana*,[15] hereafter named H23 protein. Quantitative NMR measurements have been performed on H23 samples whilst interaction studies that take advantage of these developments were undertaken on Pin1[16], a peptidyl prolyl cis-trans isomerase comprising 15 glycine residues for which an

extensive array of structural and dynamic investigative studies exists. The adapted CH_2-TROSY pulse sequence for glycine residues includes addition of selective pulses and suppression of the CαCβ decoupling, as developed in Chapter I.

As stated before, one of the major drawbacks of NMR for biomolecular investigation is its lack of sensitivity, since the NMR signal is directly proportional to the weak nuclear polarization. To compensate low sensitivity, signal averaging is required but this considerably extends the duration of experiments. However, a technique termed Dynamic Nuclear Polarization (DNP) greatly enhances sensitivity of nuclei by hyperpolarization of nuclear spins, thus increasing signal by several orders of magnitude, offering new possibilities for NMR biological investigations. Hyperpolarization is achieved by transfer of electron spin polarization to nuclear spins. When the experiment was first developed[17] in 1953, it consisted on the transfer of polarization from electron to lithium nuclei by microwave irradiation. They estimated a 100 fold nuclear population difference compared to the unpolarized sample at low field (0.003 Tesla). Because polarization is most intense at cryogenic temperatures, applications remained limited to solid NMR for many years. However, in 2003, a team managed to obtain strongly polarized nuclear spins in solution[18]. They did this by dissolution of the solid sample and transferring it from the polarizer to an NMR spectrometer via a superconducting magnet. Thanks to this methodological improvement, DNP has recently resurfaced as a pioneering method for biological investigations both at the solid and liquid state. However, Dissolution-DNP requires advanced instruments for which few are available commercially, thus stays limited to very few laboratories in the world. Therefore, although some preliminary studies have emerged for protein D-DNP, involving polarization of proteins through hyperpolarized water, there has been no work yet published for polarization of the protons of proteins directly in the polarizer. Luckily, our laboratory is equipped with such an experimental set-up. The final part of this thesis focuses on a labeling strategy that could improve investigative studies of protein interactions by D-DNP. Relaxation of the spins is influenced by their molecular environment, and notably dipolar interactions, which considerably decrease the sensitivity gains that DNP can offer. Developing a sparsely and randomly protonated protein sample in a perdeuterated environment, which could directly be polarized, could offer the relaxation gains needed to take full advantage of the ^1H polarization of proteins for biomolecular investigations. As developed in Chapter V, the small subunit of Pin1, WW, was partially protonated at different densities and used in D-DNP experiments for ligand interaction studies.

The general objective is to use and optimize new labeling strategies to improve the potential of NH/CH_2-TROSY and D-DNP experiments by selective and specific labeling of protons in a perdeuterated background for protein studies. These experiments, including triple resonance experiments for protein assignments, benefit greatly from the removal of dipolar interactions, generated by widespread protons presence, and require high levels of deuteration of the protein samples for maximum efficacy. Three labeling strategies are explored here. Firstly, protein perdeuteration of non-exchangeable sites which limits relaxation of the NH groups of interest and thus greatly enhances resolution of signals. Secondly, specific glycine labeling of proteins for use in CH_2-TROSY experiments to obtain significant gains in signal sensitivity and resolution compared to usual labeling methods. *In vitro* protein expression, also termed cell-free, is a choice strategy in these two cases. Finally, sparse and random proton labeling of proteins in a perdeuterated background for direct polarization of protein nuclear spins, thus limiting relaxation. This labeling strategy does not require sophisticated expression techniques.

Cell-based methods for expression of perdeuterated or selective labeling of proteins present several drawbacks. For selective labeling of NH groups, cell-based methods in H_2O solvent incorporate protons in undesirable sites, notably because of the presence of transaminases. In a deuterated solvent, the protons of amide groups have to be backed exchanged after expression, a considerable feat for large proteins which often requires denaturation steps, thus impacting yields. For selective labeling of glycines, enzymes responsible for amino acid metabolism incorporate labeled isotopes in other amino acids, thus decreasing the amount of labeled glycine. Cell-free protein synthesis is a technique that enables protein expression outside of live cells. The medium is therefore not constrained by complex mechanisms of cell survival, and its open nature enables access and control of the production environment. It has many purposes, notably in biology and biotechnology, for the expression of toxic or membrane proteins. For this research, its optimization enables selective isotopic labeling of the amino acids or groups of interest whilst insuring global perdeuteration. It is an attractive alternative to *in vivo* methods, provided the right additives are used, notably transaminase inhibitors. This technique enables production of perdeuterated proteins protonated on all the amide sites including in hydrophobic cores (for HN labeling) and enables elimination of a great fraction of amino acid scrambling and cross-protonation from the solvent (for HN and glycine specific labeling) resulting in high labeling selectivity.

1. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys. Rev.* **53,** 318–318 (1938).
2. Bloch, F., Hansen, W. W. & Packard, M. Nuclear Induction. *Phys. Rev.* **69,** 127–127 (1946).
3. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69,** 37–38 (1946).
4. Keeler, J. *Understanding NMR Spectroscopy, 2nd Edition*. (Wiley).
5. Wüthrich, K. *NMR of Proteins and Nucleic Acids*. (Wiley, 1986).
6. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **20,** 71–75 (2001).
7. Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **101,** 760–762 (1979).
8. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69,** 185–189 (1980).
9. Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29,** 4659–4667 (1990).
10. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89,** 496–514 (1990).
11. Barrett, P. J., Chen, J., Cho, M.-K., Kim, J.-H., Lu, Z., Mathew, S., Peng, D., Song, Y., Van Horn, W. D., Zhuang, T., Sönnichsen, F. D. & Sanders, C. R. The Quiet Renaissance of Protein Nuclear Magnetic Resonance. *Biochemistry* **52,** 1303–1320 (2013).
12. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94,** 12366–12371 (1997).
13. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced ^{1}H–^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society* **125,** 10420–10428 (2003).
14. Miclet, E., Williams, D. C., Clore, G. M., Bryce, D. L., Boisbouvier, J. & Bax, A. Relaxation-Optimized NMR Spectroscopy of Methylene Groups in Proteins and Nucleic Acids. *J. Am. Chem. Soc.* **126,** 10560–10570 (2004).
15. Mayerhofer, H., Sautron, E., Rolland, N., Catty, P., Seigneurin-Berny, D., Pebay-Peyroula, E. & Ravaud, S. Structural Insights into the Nucleotide-Binding Domains of the P1B-type ATPases HMA6 and HMA8 from Arabidopsis thaliana. *PLOS ONE* **11,** e0165666 (2016).
16. Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380,** 544–547 (1996).
17. Carver, T. R. & Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **92,** 212–213 (1953).
18. Ardenkjær-Larsen, J. H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M. H., Servin, R., Thaning, M. & Golman, K. Increase in Signal-to-Noise Ratio of >10,000 Times in Liquid-State NMR. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 10158–10163 (2003).

Chapter I: Concerns about sensitivity and resolution in solution NMR

To understand the purpose of this work, one has to be familiar with some of the principle concepts of Nuclear Magnetic Resonance[1,2]. This chapter gives a simplistic overview of some of these concepts, issues relating to solution NMR and introduces the purpose of this work.

I.1. Nuclear Magnetic Resonance phenomenon

I.1.1. Generalities

A spin is an intrinsic characteristic of elementary particles. Nuclei of non-zero spin (I), containing an odd number of protons and/or neutrons, possess a nuclear magnetic moment ($\vec{\mu}$), related to an angular momentum ($\vec{I}$) with a quantized orientation in space, and a gyromagnetic ratio (γ) such that: $\vec{\mu} = \gamma \times \vec{I}$. Nuclear Magnetic Resonance is based on the detection and measurement of these magnetic moments. When in interaction with a magnetic field B_0, the magnetic moments will orient themselves parallel or antiparallel according to this field and split into 2I+1 different energy levels, a phenomenon called the Zeeman effect. For given spins I = ½ , which will always be the case in this work, there are two possible energy levels that the spins can populate in a magnetic field B_0 defined by: $E_m = -\gamma\hbar m_I B_0$ (Figure 1.1) with $m_I = +\frac{1}{2}$ (α state) or $m_I = -\frac{1}{2}$ (β state) and with a difference of energy between these two levels corresponding to $\Delta E = \gamma\hbar B_0$ where $\hbar$ is the Plank constant divided by 2π (Figure 1.1).

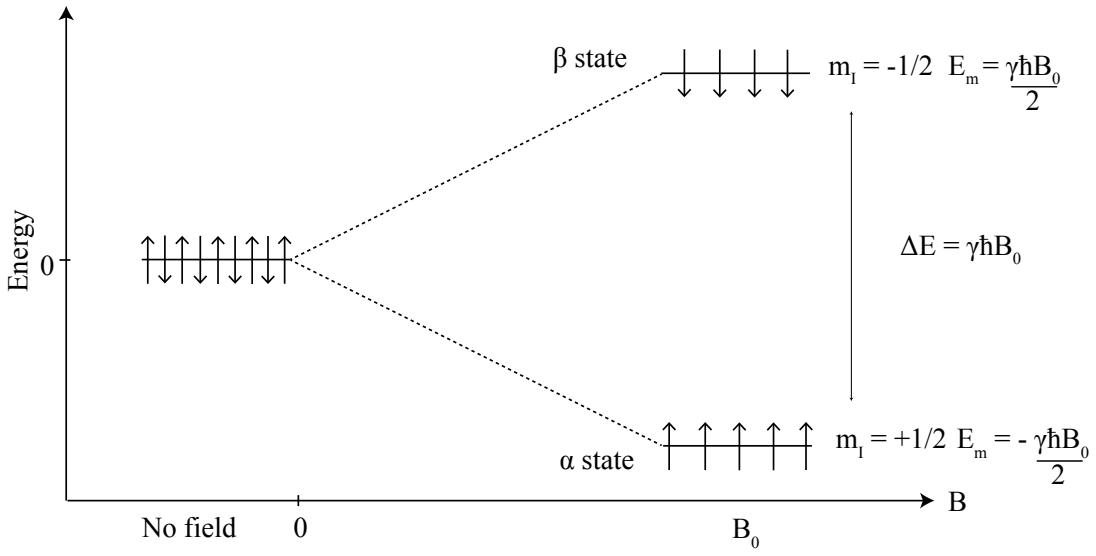


Figure 1.1: Zeeman effect. In a strong magnetic field B_0, the ½ spins will divide themselves into two populations of higher ($m_I = -\frac{1}{2}$) and lower ($m_I = +\frac{1}{2}$) energy state following the Boltzmann distribution.

At thermal equilibrium, the spin populations of the different energy levels follow the Boltzmann distribution, whereby the probability of spin being at a certain energy level $\mathcal{P}_m$ is given by: $\mathcal{P}_m \propto e^{\frac{-E_m}{k_B T}}$ where k_B is the Boltzmann constant and T the temperature. Following this distribution, the lower energy level is slightly more populous. This results in an average net magnetization ($\overrightarrow{M_0}$) defined by the vector sum of all the magnetic moments $\vec{\mu}$: $\overrightarrow{M_0} = (n\alpha - n\beta)\vec{\mu}$. NMR signal sensitivity is conditioned by this net magnetization and a convenient way to account for this sensitivity is to express the polarization P of the sample : $P = \frac{n_\alpha - n_\beta}{n_\alpha + n_\beta}$. The population ratio of the two levels can be expressed as : $\frac{n_\beta}{n_\alpha} = e^{\frac{-\Delta E}{k_B T}}$ with n_α and n_β the number of spins populating the α and β states respectively. For the population of I = ½ spins : $P = \frac{n_\alpha - n_\beta}{n_\alpha + n_\beta} = \frac{1 - {}^{n_\beta}/_{n_\alpha}}{1 + {}^{n_\beta}/_{n_\alpha}} = \frac{1 - e^{-\Delta E/k_B T}}{1 + e^{-\Delta E/k_B T}}$ and with $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$, then $P = \tanh\left[\frac{\Delta E}{2k_B T}\right] = \tanh\left[\frac{\gamma \hbar B_0}{2k_B T}\right]$. In regular NMR conditions where polarization is low, P can be approximated as $P = \frac{\Delta E}{2k_B T} = \frac{\gamma \hbar B_0}{2k_B T}$ following a linear law.

One can see from this equation how the intensity M_0, which influences the NMR signal, is strongly dependent on the gyromagnetic ratio, the field B_0 and the temperature, indirectly influencing average net magnetization. The NMR signal is to a further extent also dependent on the sensitivity of the probe used to detect signals.

In light of this, a quick calculation informs about the first major drawback of NMR spectroscopy. For the highest magnetic field available, at 28.2 T (1.2 GHz), at room temperature, proton polarization reaches $P(^1\text{H}) \approx 0{,}0096\%$, meaning the difference in population between the two states is incredibly small. Although instrumental progresses have been made to generate higher magnetic field strengths, and thus improve polarization, this cannot change the intrinsic characteristic of nuclear magnetic moments and associated gyromagnetic ratios, which are very low compared to that of electrons. A number of advances in NMR, from various labeling strategies to optimized pulse sequences have been in an attempt to enhance NMR signal sensitivity. Dynamic Nuclear Polarization (DNP) is a technique that enables increased polarization of nuclear spins by polarization transfer from electrons. The section 4 of this chapter is dedicated to the topic.

I.1.2. Two types of relaxation

Nuclear relaxation[3] is one of the most fundamental aspects of NMR which includes all the phenomena of energy exchange that drive spins back to thermal equilibrium after a perturbation. The longitudinal and transversal relaxations give information about molecular movements, distance and angles between atoms, enabling analysis of changes in conformational structure.

As stated previously, the vector sum of all the magnetic moments in a magnetic field B_0 creates a net magnetization M_0 at thermal equilibrium, which can be defined solely as magnetization along the z axis (M_z). The magnetic moments precess around the z axis incoherently, such that there is no magnetization in the xy plane ($M_x = M_y = 0$). In pulse NMR, when the system is perturbed by an external pulse of field B_1 perpendicular to B_0, possessing a frequency ν_o corresponding to the difference in energy between the spin states, i.e. a radio-frequency pulse for nuclei or a microwave pulse for electrons, the energy is absorbed by the system. For a given duration of the pulse, the spin populations α and β can be equalized, leading to magnetization in the z axis becoming null. However, such a pulse will also generate phase coherence of the spins, whereby all the spins align in the xy plane, rotating at the same speed around the magnetic field direction and therefore create a M_x and M_y magnetization. The relaxation phenomenon back to thermal equilibrium is composed of the longitudinal evolution of M_z back to its maximum, and the transversal evolution of M_x and M_y back to zero. In 1946, Bloch described relaxation as the evolution of magnetization in the xyz plane for an isolated spin system which can be simply put as:

$$\frac{dM_z(t)}{dt} = -(R_1(M_z(t) - M_0)) \quad (1.11)$$

$$\frac{dM_y(t)}{dt} = -R_2 M_y(t) \quad (1.12)$$

$$\frac{dM_x(t)}{dt} = -R_2 M_x(t) \quad (1.13)$$

Where $R_1 = \frac{1}{T1}$ and $R_2 = \frac{1}{T2}$ are the rates corresponding to the longitudinal and transversal relaxation respectively. M_z follows an exponential recovery towards its equilibrium value M_0 where $M_z = M_0(1 - e^{-t/T_1})$ with a characteristic time constant T_1. On the other hand, the phase coherence of spins created by the radio-frequency pulse will disperse following $M_{xy} =$

$M_0\, e^{-t/T_2}$ with a time constant T_2 (Figure 1.2) and the transverse relaxation including $\sin(\omega_0 t) e^{-t/T_2^*}$ with T^*_2 the time constant of observed relaxation including field inhomogeneities. The associated M_x and M_y magnetization will return to zero following an exponential law.

Whereas longitudinal relaxation (T_1) of the spin system along the z axis represents the transition of spins from an excited state back to thermal equilibrium, thus conditioning the delay between two experiments, transversal relaxation (T_2) along the xy plane represents the randomization, or loss of coherence of spins along in the xy plane by energy dissipation and determines the lifetime of the signal, which will determine the linewidth of the signal.

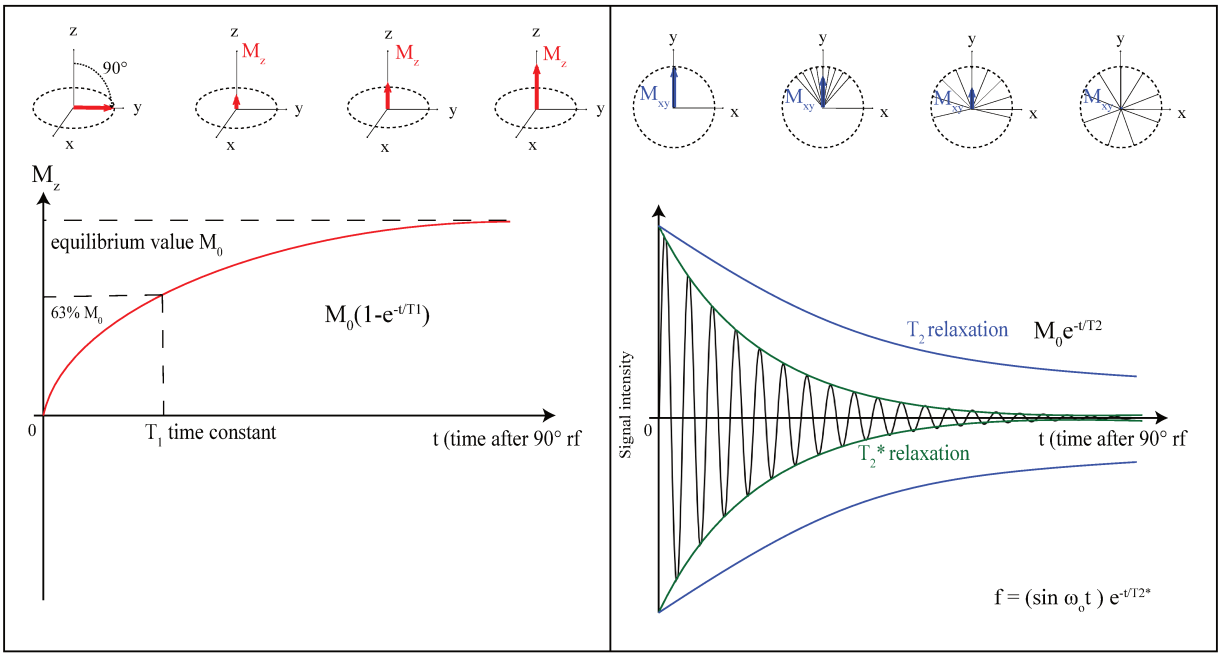


Figure 1.2: Two mechanisms of relaxation of a spin system. The Free Induction Decay (FID), in the time domain is transformed into a signal in the frequency domain by a Fourier Transform.

After this initial pulse, the energy absorbed by the system is exchanged with the environment producing an electromagnetic signal of a certain amplitude rotating around the z axis in the xy plane with a frequency of $\nu_0 = \frac{\gamma B_0}{2\pi}$. But the spins precessing in the xy plane do not all precess with the same frequency, influenced by their chemical environment. The signal emitted by the spins in the xy plane, termed free induction decay (FID) is recorded by the coil probe of the spectrometer. It contains all the precessing frequencies of the different types of spins in the time domain but cannot be interpreted as such[4]. Transformation into the frequency domain by the Fourier transform is required for analysis of the signal.

I.1.3. Different sources of relaxation

There are several factors influencing the relaxation of a spin system of which the four most important are unpaired electrons, chemical shift anisotropy (CSA), quadrupolar and dipolar interactions. Unpaired electrons are an important relaxation factor, as their Landé factor is several orders of magnitude larger than the gyromagnetic ratio of the proton. CSA depends on the electronic distribution around a nucleus, which is generally not spherically symmetrical, but anisotropic. Therefore, the chemical shift of a nucleus will vary with the orientation of the molecule and changes in the magnetic field. The chemical shifts fluctuation due to molecular movements will be a source of relaxation.

Rotational correlation time τ_C (or molecular tumbling corresponds to the average time it takes a molecule to rotate of one radian) describes the time dependent movements of the molecule. The rotational correlation time of a spherical molecule of radius Γ can be defined using the Stockes Einstein formula: $\tau_C = \frac{\eta 4\pi \Gamma^3}{3k_B T}$ where η is the viscosity, k_B the Boltzmann constant and T the absolute temperature. Rotational correlational time is critical to relaxation and in the case of the relaxation of two identical neighboring ½ spins it describes the fluctuation of dipolar interactions and the relaxation rates T_1 and T_2 directly depend on τ_C:

$$\frac{1}{T1} = \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma^4 h^2}{r^6}\left[\frac{2\tau_c}{1+\omega_0^2\tau_c^2} + \frac{8\tau_c}{1+4\omega_0^2\tau_c^2}\right]$$

$$\frac{1}{T2} = \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma^4 h^2}{r^6}\left[3\tau_c + \frac{5\tau_c}{1+\omega_0^2\tau_c^2} + \frac{2\tau_c}{1+4\omega_0^2\tau_c^2}\right]$$ with ω_0 the transition frequency of the spin, *r* the distance between the spins, γ the gyromagnetic ratio of the nuclei, *h* the Plank constant and μ_0 the magnetic moment of the spin. Depending on the value of the τ_C of a molecule, one can see from these equations how a term may take the upper hand for T_1, as $1/T_1$ is dependent of $1/\tau_C$ when $\omega_0\tau_c \gg 1$. When the τ_C increases, the relaxation rate will tend towards 0 thus T_1 relaxation will be very long. Moreover for $\omega_0\tau c \approx 1$ the relaxation rate will reach a maximum and T_1 will be at its minimum. For T_2 however, the dominant term is τ_c and $1/T_2$ depends on τ_c . Thus, the relaxation rate T_2 will only increase with the size of the molecule, and relaxation will be that much more efficacious for slow molecular tumbling, translating an increased speed of coherence loss in the xy plane. This constitutes the second major drawback of NMR spectroscopy, as T_2 relaxation determines the linewidth of the signals. The decreased T_2 relaxation time increases the widths observed for large objects such as proteins or for spectra recorded at lower temperatures, leading to poor resolution of signals. This drawback has been

intensively considered and optimized since NMR emerged, and especially in the context of biomolecular NMR.

T_1 and T_2 relaxation are not only dependent of molecular tumbling of the molecule, but also the magnetic field strength, to a certain degree as $\omega_0 = \gamma B_0$. For T_1, an increased magnetic field strength will decrease relaxation rate, thus improve T_1 relaxation. However, a higher magnetic field for will have little influence over the T_2 for higher τ_C of the molecule. The figure 1.3 below represents the T_1 and T_2 evolution in function of τ_C at two different magnetic field strengths.

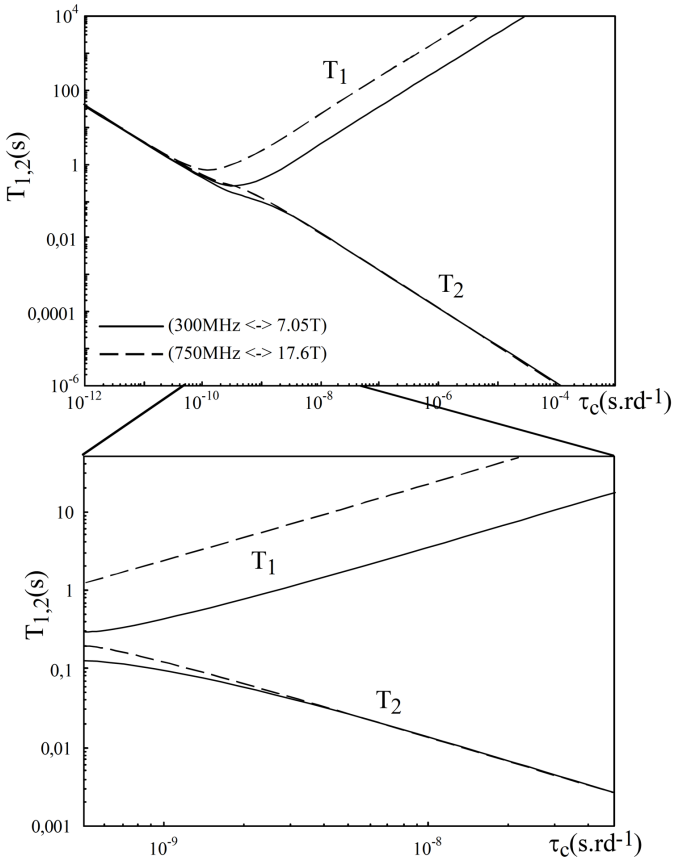


Figure 1.3: Characteristic curves described by the previous equations of T_1 and T_2 relaxation times at 300 MHz (full line) and 750 MHz (dotted line). Calculations were done for two protons of 1.8Å distance. The bottom frame is an enlargement of the rotational correlation times corresponding to proteins.

Dipolar interactions, or the distance between spins, is a contributing factor to T_1 and T_2 relaxation, and notably ^{1}H-^{1}H dipolar couplings. The magnetic moments of nuclear spins interact with neighboring magnetic moments created by other nuclear or electron spins. The molecular movements of these dipoles are random in solution and thus create local, fluctuating magnetic fields which contribute to dissipation of energy. For two neighboring spins S and I,

the local magnetic field B_{loc} created by the dipole-dipole interaction around I is defined by: $B_{loc} \propto \frac{\mu_I}{r_{IS}^6}(3cos^2\theta - 1)$ where θ is the angle between the vector between I and S and the magnetic field B_0, r_{IS} is the distance between the two spins I and S and μ_I is the magnetic moment. The further apart are the two spins, the lower is the local magnetic field of the dipolar interaction. As seen previously, the gyromagnetic ratios of two coupled spins, their magnetic moment and the distance between them are determining factors for the T_1 and T_2 relaxation rates. For spins of larger gyromagnetic ratio, the relaxation rates increase, thus T_1 and T_2 decrease. The transverse relaxation time influenced by dipolar interactions T_1^{DD} and T_2^{DD} can be defined as $\frac{1}{T_1^{DD}} \propto \frac{\gamma_I^2 \times \gamma_S^2}{r^6}$ (1.22) and $\frac{1}{T_2^{DD}} \propto \frac{\gamma_I^2 \times \gamma_S^2}{r^6}$ (1.23). The further the distance between spins, the lower is the dipolar interaction and thus their influence on relaxation rate. Because T_1 and T_2 relaxation time are subject to the dissipation of energy through dipolar interactions, the more nuclear spins are interacting, then the more efficacious is relaxation.

I.2. Reducing T_1 and T_2 Relaxation through deuteration labeling strategies

Fast nuclear spin T_2 relaxation, and T_1 in certain instances, can be major drawbacks to analysis of biomolecules by NMR spectroscopy and diverse strategies have been employed to palliate this problem. Although increasing the volume or concentration of the NMR sample would seem like an efficient way to increase signal, producing high amounts of material, as well as the high cost of using labeled precursors, the problems of shimming big volumes and the issues related to solubilizing high concentrations of proteins means that other strategies have to be employed. Moreover, this does not solve the issues relating to fast T_2 relaxation, and signal overlap due to low resolution precludes detailed study of large molecular objects. As seen previously, a major contributor to the relaxation phenomena are dipolar interactions, created by neighboring spins, that considerably accelerate longitudinal and transversal relaxation of the system by favoring energy dissipation and loss of spin coherence in the xy planes. Thus, protein deuteration has become an efficient technique to limit these dipolar interactions, improving the potential of protein NMR studies.

Perdeuteration of the spin-system environment, by replacing neighboring ^1H, a nucleus with a high gyromagnetic ratio, with ^2H, possessing a lower gyromagnetic ratio by 6.5-fold, is an effective way to limit spin diffusion and improve relaxation properties of proteins. Indeed, the magnetic moment of deuterium being significantly lower than ^1H, it forms a less intense local

magnetic field, thus limits energy dissipation through proton networks[5] (Figure 1.4). Moreover, deuterium is not "seen" on a conventional ^{1}H spectra as its resonance frequency differs considerably from that of the proton. Replacing protons with deuterium eliminates many relaxation pathways, reduces relaxation rates and thus improves resolution and sensitivity of the sample.

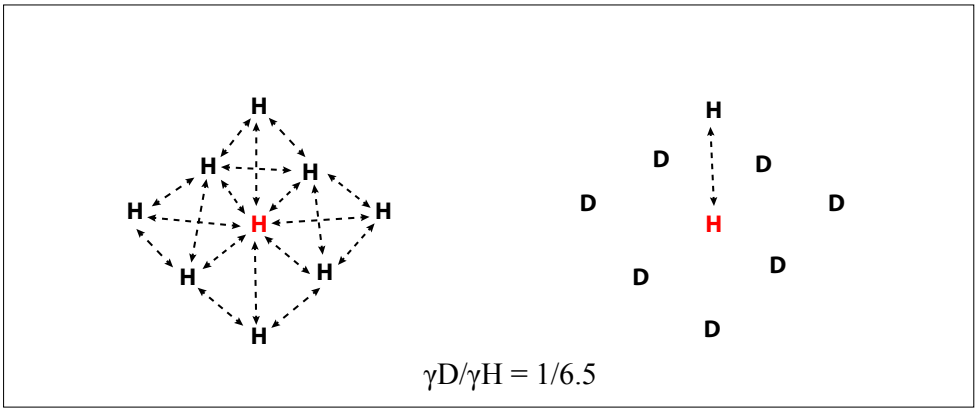


Figure 1.4: Schematic representation of a proton network in a given system, and the equivalent network when the majority of neighboring protons are replaced by deuterium isotopes.

Isotopic enrichment of proteins, and notably their perdeuteration, was initially developed for structural study of proteins[6–8] larger than 20 kDa. Because of fast relaxation rates created by pervasive proton presence, signals were often unresolved, overlapping or below S/N ratio for these large molecular objects, and posed issues for their structural study. Deuteration was an important factor in overcoming these drawbacks. Since the nineties, a number of strategies have emerged and have been perfected over the years for protein deuteration[5,9]. These range from random fractional deuteration[8,10], to broad deuteration of non-exchangeable protons, whilst also including specific residue type, position or group labeling strategies in a perdeuterated background[11,12]. Perhaps one of the greatest achievements to structural study of large proteins was the development of methyl specific labeling. It is possible to specifically label methyls of isoleucine, valine and leucine side chains using ^{13}C α-ketoacid precursors in E.coli[13–19] and notably by the use of cell-free protein expression[15,20], such that now, methyl groups of proteins over half a mega Dalton can be well resolved[21,22]. Together, these labeling strategies have greatly enhanced our knowledge of protein structure, nevertheless, there are still several unresolved issues and unexplored strategies for protein labeling. The next sections introduce the three labeling strategies used in this thesis. Notably, use of cell-free expression for complete deuteration of non-exchangeable protons of large proteins, specific labeling of the methylene of glycine residues in a deuterated background and finally, the use of fractionated deuteration for Dissolution DNP studies.

I.2.1. Deuteration for amide spin system analysis

Deuterium labeling is a key factor to enhanced NMR signals, notably by the removal of dipolar interactions around the spin system of interest. Usually, protein perdeuteration is achieved using a version of the common Marley protocol,[23] which employs unlabeled rich media to start the cell culture, then requires transfer of the cell mass to a small volume of labeled media for protein expression. Using 100% deuterated water and fully deuterated glucose in the labeled media, this approach can attain the very high levels of deuteration required for NH TROSY experiments. Other methods have been described,[24,25] however they all share a common issue for amide analysis by NMR spectroscopy. When expressing medium to large perdeuterated proteins, back exchanging HN protons located within hydrophobic cores is often very long or near impossible by simple solvent exchange and provides incomplete spectra. It is then possible to undertake the denaturation of the expressed protein in H_2O media, and, although ensuring back-protonation of all the amide sites, requires refolding the protein which considerably affects overall yields and sometimes cannot be achieved.

Cell-free protein synthesis is an ideal alternative to these cell-based methods. This technique of *in vitro* protein synthesis employs cell-extracts for protein expression, instead of using live cells. The extract is prepared such that it contains all the necessary enzymes and cofactors for protein expression, whilst remaining an open system in which different additives, some of which toxic for cells, can be added without affecting the synthesis machinery and overall yields to a certain extent. The cell extract is prepared such that very few native amino acids remain, and the amino acids with the experimenter's choice of labeling can therefore be added for protein incorporation. To preserve the presence of ^1H in a perdeuterated background for amide group spectroscopy, the protein of interest can be expressed in H_2O media, supplemented with deuterated amino acids and with a number of additives, of which inhibitors of scrambling enzymes. The perdeuterated amino acids will be incorporated into the protein sequence and conserve the protons on exchangeable sites including in hydrophobic cores. Addition of transaminase inhibitors ensures low levels of back-protonation on Cα sites, enabling us to take full advantage of the TROSY version of various 2 and 3 dimensional experiments without any problems of missing correlations because of un-exchanged amide protons. Such a cell free approach for expressing deuterated proteins on non-exchangeable sites whilst conserving the amide protons has not been demonstrated so far, and has been the subject of a book chapter

where I am co-author[52]. We have demonstrated that this strategy is perfectly adapted for the expression of the large molecular complex TET2.

I.2.2. Deuteration for glycine methylene spin system analysis

Cell-free is also an ideal technique for specifically labeling certain amino acids. Indeed, because the system is closely controlled and amino acids of choice are incorporated into the medium for protein synthesis, it is possible to add only one isotopically labeled amino acid and use deuterated algal extracts to provide the others, with no other isotopic labeling. Addition of transaminase inhibitors ensures limited isotopic scrambling and conservation of the labeled amino acid of interest[26]. For our case, it enables the addition of specifically labeled glycine, with its two diastereotopic protons in a perdeuterated background, for which no other expression protocol has yet been described to our knowledge. For the CH_2-TROSY experiment, insuring perdeuteration around the methylene of glycine can maximize signal and resolution enhancements of the residue, by minimizing T_2 relaxation generated by dipolar interactions. Using different levels of deuterated samples, we show the importance of deuteration to attain the full potential of this experiment, detailed in Chapter IV.

I.2.3. Deuteration for protein Dissolution-Dynamic Nuclear Polarization of protons

Finally, presence of numerous dipolar interactions is not only challenging for signal resolution (fast T_2 relaxation) but also for T_1 relaxation. This is problematic for D-DNP experiments as the duration of hyperpolarization is conditioned by the length of the T_1 (See section 4 of this Chapter). In fact, many D-DNP experiments are undertaken on ^{13}C with partially or completely deuterated samples so as to limit T_1 relaxation. In 2015, Stevanato et al.,[27] managed to conserve a nuclear ^{13}C singlet lifetime of a molecule for more than an hour at room temperature and in solution. However, this was only achievable because the spin pair of interest was well isolated from other spins so as to remove dipolar-coupling-induced relaxation. In 2016, Kurzbach et al.,[28] used hyperpolarized HDO to investigate amide groups of an intrinsically disordered protein. In this way, the HDO acted as a reservoir for hyperpolarized protons, the deuterium preventing fast T_1 relaxation of the proton. Zhang and Hilty[29] suggest using deuteration as a means to extend T_1 times for chemical and biochemical applications of D-DNP.

To date, there has been little published research for direct protein hyperpolarization for D-DNP, and even fewer for proton hyperpolarization of proteins, although one has been published for peptide proton hyperpolarization[30]. A number of reasons are possible for this, of which the difficulty of circumventing fast T_1 relaxation rates observed for proteins. In order to achieve longer T_1 relaxation times, a labeling strategy for random fractional deuteration would be ideal, where protons would be sufficiently isolated from their neighbors, thus offering the T_1 gains necessary for D-DNP experiments, whilst conserving a high enough concentration for hyperpolarization and signal recording. Fractional deuteration was used previously for 3D structure determination studies[31], giving resolution and signal enhancements necessary for large proteins.

The easiest approach for achieving random fractional deuteration is to use bacteria-based expression protocols for over-expression of the protein of interest. The genetically conditioned bacteria, containing a plasmid with the gene to be expressed, are grown in a rich medium and then transferred to minimal media containing the desired amount of D_2O and deuterated glucose. Deuterated water is toxic to cells at a certain concentration, thus, to optimize protein expression, cell cultures can be transferred in several steps to increasing amounts of D_2O based medium during several growth cycles. Slight modifications are necessary to ensure proper cell growth and protein expression, but this protocol is similar to the Marley protocol mentioned previously, and deuteration levels of 80% can easily be achieved.

I.3. Generalities concerning the TROSY experiments

I.3.1. Introduction

2D NMR has had a great impact on structural NMR spectroscopy, as spin correlations provide enhanced structural information for bigger and more complex system analysis. The Heteronuclear Single Quantum Correlation experiment (HSQC) has become a routine 2D experiment in biomolecular solution NMR, where magnetization is transferred between two different kinds of nuclei involved in a scalar coupling (or *J*-coupling). The recorded spectrum will present cross peaks due to coherence transfer through the one bond heteronuclear coupling, with one dimension usually involving the proton and the other another type of nucleus, typically ^{13}C or ^{15}N. When observing the correlation between a proton and another heteronuclear ½ spin, there is a sensitivity advantage to observing the former because of its higher gyromagnetic ratio, that is why signal recording will usually be undertaken on the proton. The HSQC pulse

sequence experiment for two coupled spins H and C, is illustrated below (Figure 1.5). It can be broken down into individual elements of which the basis is the INEPT[32] (Insensitive Nuclei Enhanced Polarization Transfer) sequence, an important building block of heteronuclear experiments, enhancing sensitivity by transferring the magnetization of the proton to one of lower magnetization, such as ^{13}C or ^{15}N, independent of chemical shift (or offset). The HSQC pulse scheme and the vector representation of the spins is displayed on Figure 1.5, taken from Vega-Hernandez & Antuch[33].

Preparation period :

1) The first 90°x pulse on the proton creates a coherence as the spins align together, generating proton transverse magnetization along the -y' axis.
2) The vector, rotating in the xy plane around the z-axis splits depending on the coupling partner of the proton. The system will evolve during a time $1/4^1J_{CH}$, where $^1J_{CH}$ corresponds to the coupling constant between the H and C nucleus. After such a time, the angle between the two components of the vector will be $\pi/2$ radians.
3) The 180°x pulse on the proton will rotate the transverse magnetization in the x'y' plane 180° around the x' axis,
4) The simultaneous 180°x pulse on the carbon will change the net carbon magnetization from z to -z.
5) After another delay of $1/(4^1J_{CH})$ (t), the vector components of the proton form an angle of π radians, thus antiphase magnetization is generated on the H spin
6) A 90°y pulse on the proton brings the vectors to z and -z, depending if they are coupled to a carbon spin β or α respectively
7) A simultaneous 90°y pulse on the carbon will cause coherence transfers of the antiphase signal from the proton to the carbon

Figure 1.5: HSQC pulse sequence for two spins S and I. Pulses are designated by grey bars, with the rotation angles and phase of the pulse above. Red vectors represent the proton whereas blue represents carbon. The arrows represent magnetization and the labels indicate the coupling partner. Underneath is represented the vectoral magnetization evolution during the experiment.[33]

Evolution period :

8) The carbon evolves for a time t_1 according to its offset and depending on the coupling with protons. During this time, the proton transverse magnetization does not evolve as all the magnetization is in the z-axis.
9) At $t_1/2$, a 180°x pulse to the proton will invert the proton magnetization along the z axis, inverting the labels and reversing the direction of the evolution of the carbon vectors.
10) After another $t_1/2$ delay, the carbon vectors will be in antiphase with magnetization along the y'axis. The t_1 time will be incremented during a series of experiments allowing expression of the carbon chemical shifts and construction of the indirect dimension of the spectrum.

11) A 90°x pulse to carbon brings carbon magnetization back to the z axis and invert the population.

Back transfer to proton

12) A subsequent 90°x pulse applied to the proton allows transfer of polarization back to the proton leaving an antiphase signal
13) Proton transverse magnetization will evolve according to the coupling with the carbon, during a time $1/(4^1J_{CH})$ until the components form an angle of $\pi/2$ radians
14) A 180°x pulse to proton inverts the vectors
15) A simultaneous 180°x pulse to carbon inverts the carbon labels
16) The proton vectors continue to evolve for $1/(4^1J_{CH})$ until both components are in-phase and detection can take place (with decoupling of the carbon).

In fine, the HSQC scheme generates a spectrum for a two ½ spin coupled system consisting of a cross peak or four cross peaks if ^{13}C and ^{1}H decoupling pulses are omitted, with different lines widths in both dimensions. This is because for large biomolecular complexes at high magnetic fields, transverse (T_2) relaxation of ½ spins (typically ^{1}H, ^{15}N and ^{13}C nuclei) is dominated by 2 types of relaxation: 1) autorelaxation, caused by dipole-dipole (DD-DD) interactions and chemical shift anisotropy (CSA-CSA) of the individual spins, and 2) cross-correlated relaxation, caused by DD/CSA interactions within spin pairs. Since transverse relaxation is dominated by autorelaxation and cross-correlated relaxation, in a coupled ½ spin system, the individual components of the resulting multiplet will have different relaxation rates, and thus different line widths depending on the efficacy of the relaxation. If an extra spin echo and decoupled acquisition on I are employed (Figure1.5), then there will be only an in phase observable term created which is essentially a mean of all the cross peaks.

The TROSY (Transverse Relaxation Optimized Spectroscopy) experiment was first developed in 1997 to enhance spectral quality of amide groups of proteins by increasing the signal and resolution of correlations[34,35]. TROSY experiments are spin state selective, and are based on the selection and isolation of exclusively the slowest relaxing cross-peak of the multiplet observed in the regular coupled-HSQC scheme, instead of an average, by use of optimized delays, gradients and phase cycles in the pulse sequence.

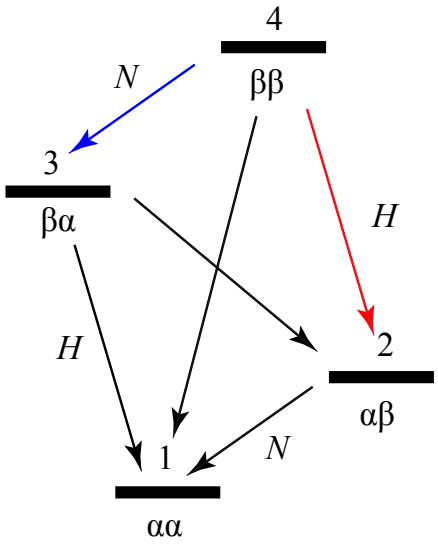


Figure 1.6 : Energy level diagram of a two 1/2 spin coupled system. In blue the most favorable single transition for spin N (nitrogen) and in red for spin H (hydrogen).

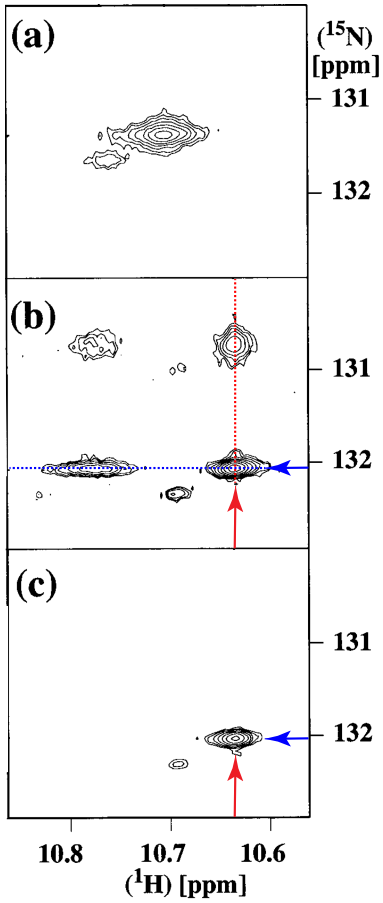


Figure 1.7 : ^{15}N-^{1}H correlations of Trp-48 taken from Pervushin et al., 1997. Panel a) decoupled HSQC, b) fully coupled HSQC and c) TROSY. The blue arrows represent the relaxation from the favorable transition of spin S and the red arrows represent the relaxation from the favorable transition of spin I.

To illustrate, we consider the energy level diagram of a coupled ½ spin system whereby the spins *I* and *S* can either be in the α or β state, and with a scalar coupling constant J_{IS} (Figure 1.6). Each transition is a change in state of one of the spins. The relaxation rates of the individual components of the multiplet can be illustrated as a single transition couple: transitions 2→1 or 4→3 for spin *I* and transitions 4→2 or 3→1 for spin S. A basic spectrum would therefore resemble panel (b) in Figure 1.7 for this two ½ spin system, where all four single-transition couples are visible. As mentioned before, some transitions are more efficient than others, thus components of the multiplet present different relaxation rates, and are differently resolved or intense. In the decoupled HSQC or HMQC spectra, decoupling the insensitive dimension during acquisition creates a mean of all the components of the multiplet (panel (a) Figure 1.7), but in the TROSY spectra, only the slowest relaxing component is selected from the others, therefore only the most favorable single transition couple is visible (panel (c) Figure 1.7).

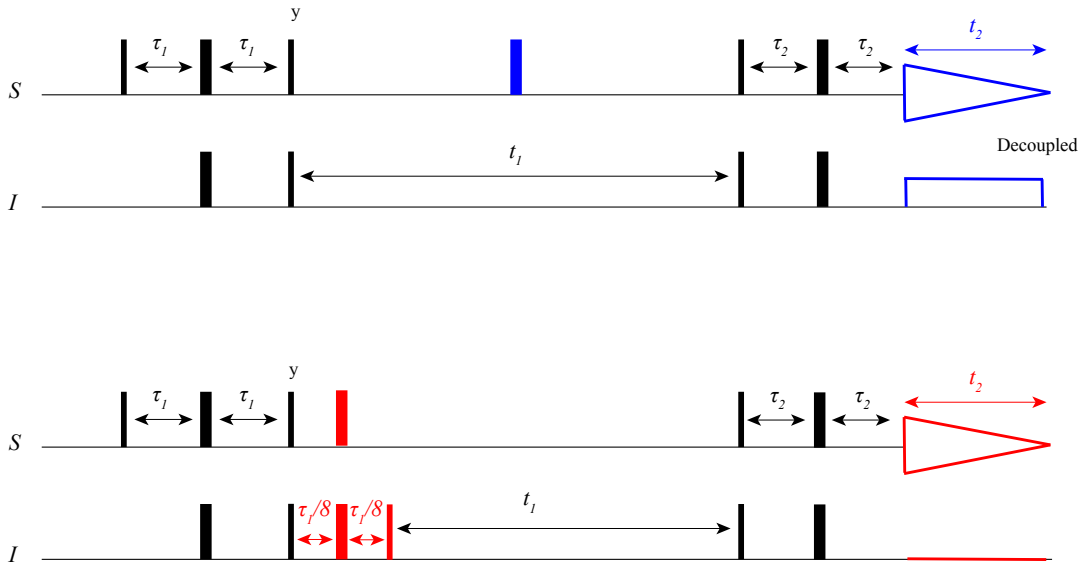


Figure 1.8 : Comparison of a HSQC and TROSY pulse scheme for two ½ spins S and I. In blue, the upper panel, the pulse scheme for a decoupled HSQC and in red, bottom panel the pulse sequence for the TROSY.

To eliminate unfavorable transitions, phase cycles are applied with a pulse of 180° on both the S and I spin, which inverse the sign of certain correlations thus they can be canceled during the acquisition of several scans, and the selection of the slowest relaxing component with a 90° on the I spin separated by optimized delays of $1/16J_{SI}$ (spin state selective excitation S3E).

There have been many different types of 2D TROSY experiments described : for aromatic CH pairs,[36–38] CH_2 methylene groups,[39] CH_3 methyl groups,[40] and more recently, CF_3 trifluorinated methyl groups.[41] TROSY experiments have contributed to pushing the limits of our understanding of biomolecules, and particularly large molecular systems thanks to the enhanced resolution and sensitivity they offer. The NH TROSY was particularly important in this aspect, as it is the base for many 3D triple resonance experiments used nowadays for protein assignments, such as the HNCO[42], HN(CA)CO[43] and HNCACB[44]. But TROSY experiments are heavily dependent on the level of deuteration of the environment of the spin system, so that even low levels of residual protonation will rapidly affect the quality of spectra[45]. This explains why new isotopic labeling methods specific to the different TROSY experiments promptly emerged[13–19]. Methyl-TROSY experiments, in particularly, have occupied a lot of the attention[16,21,40,45]. Methyl groups have extremely favorable dynamic properties as they are generally found at the end of lateral chains thus become degenerate in solution and the dipolar couplings are no longer present. The methyl TROSY is based on an HMQC experiment, and

the resulting spectrum displays some correlations with such unfavorable relaxation properties that they become invisible and only the slowest relaxing component remains. These methyl-TROSY experiments have allowed observation of proteins and complexes of much greater mass. However, for methylenes, and notably methylenes of glycine, these groups are directly on the backbone thus they take the τ_C of the protein and have strong autorelaxation properties as well as being surrounded by numerous other protons. Although methylene-TROSY will not be able to offer the same favorable relaxation properties as the methyl-TROSY, and thus may not be useful for analysis of proteins of large molecular mass, it can still give us precious information about backbone dynamics. This thesis dedicates Chapter IV to a specific glycine labeling strategy, to improve methylene enhancements using the CH_2-TROSY.

I.3.2. Theory of the CH_2-TROSY experiment

The CH_2-TROSY is a spin state selective experiment dedicated to the methylene group which, using optimized transfers and delays, selects the slowest relaxing cross-peak of the sixteen CH_2 multiplet usually observed in a fully-coupled HSQC scheme.

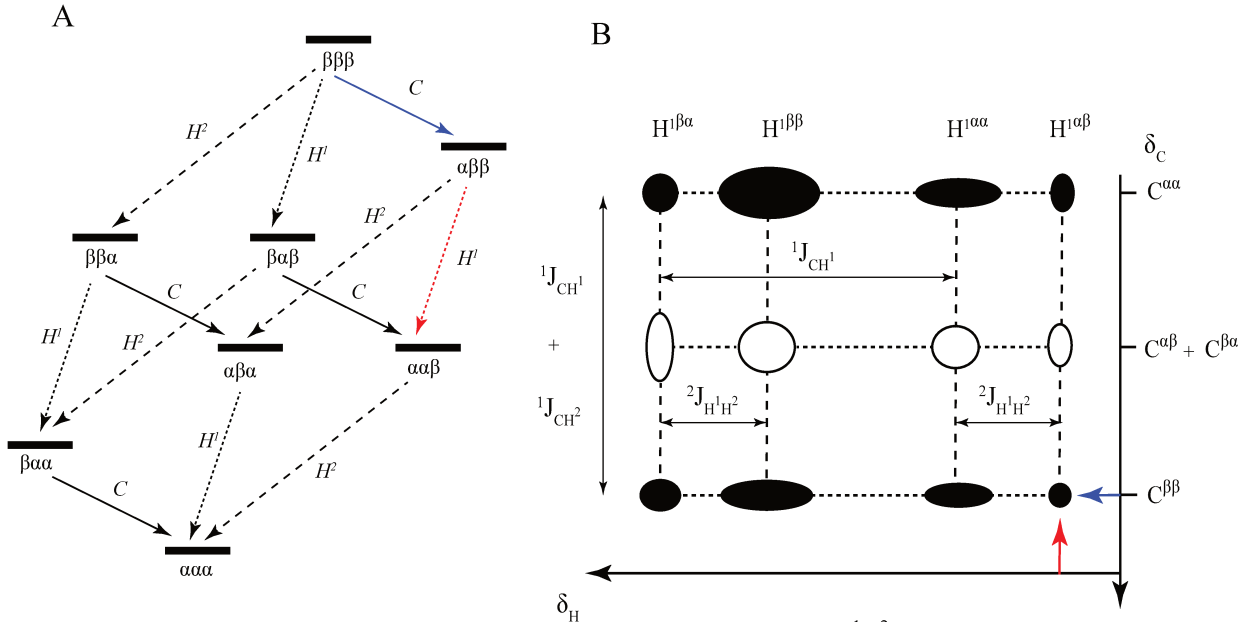


Figure 1.9 : A) Energy level system of three coupled ½ spins of an isolated CH^1H^2 group. Under each bar are indicated three symbols which, from left to right, correspond to the eigenstate of the C, the H^1 and the H^2 atoms of the methylene. Full arrows correspond to C transition, small dotted arrows to H^1 transitions and long lined arrows to H^2 transitions of state. B) Theoretical 2D spectrum for H^1 of the three ½ spin coupled system.

Consider a three ½ spin coupled system CH^1H^2 of an isolated methylene $^{13}CH_2$. The energy level diagram is more complex than for a two-spin system, presenting 8 levels depending on the spin state of C, H^1 and H^2 which can either be in the α (+) or β (-) state (Figure 1.9, panel A). The relaxation rates of the individual components of the multiplet can be illustrated as single

transitions between the different states, where each transition is a change in state of one of the spins. The most favorable transitions are selected and isolated such that on the 2D spectrum, only the slowest relaxing component, and thus the more resolved, will be visible. In this TROSY experiment, to specifically select the best relaxing component of the eight multiplet in a fully coupled HSQC experiment, optimized transfer delays are inserted into the pulse scheme (Figure 1.10). The pulse schemes both start with a selective INEPT transferring magnetization from the proton to the Cα. In the HSQC experiment, the ^{1}H is decoupled during t_1 and the ^{13}C is decoupled during t_2, with a Rance-Kay element at the end of the sequence[46] to select the coherence transfer. In the TROSY experiment, there is no decoupling, but an added spin state selective S^3E element[47] of precisely defined $\frac{1}{16J}$ duration of excitation that selects the most downfield component of the carbon multiplet, corresponding to the slowest relaxing component.

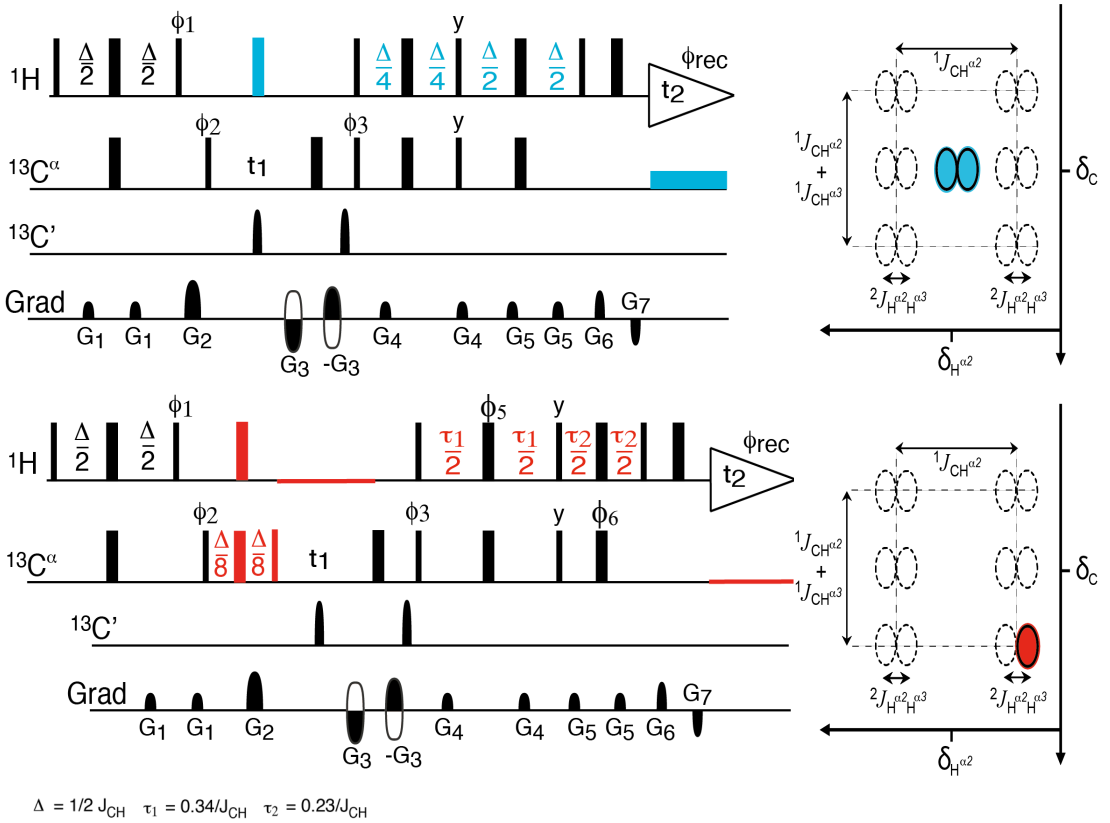


Figure 1.10 : Pulse schemes for HSQC (blue) and TROSY (red) experiments with the corresponding spectra obtained on the right.

After selection of the best relaxing component of the C multiplet (the most downfield, for which both H spin states are in β state), the magnetization can be expressed with four operator terms :

C_x
$2C_xH_z^1$
$2C_xH_z^2$
$4C_xH_z^1H_z^2$

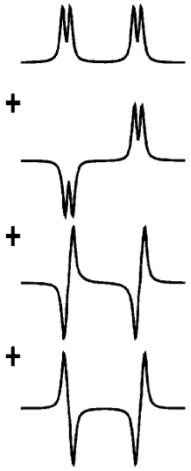
Then, a spin state selective coherence transfer (double INEPT, or Rance Kay element) will transform this into four H^1 product operators. In the spectrum, the four components of the H^1 multiplet will each correspond to a single quantum transition of H^1. Each of these single quantum transitions can be represented as a combination of 4 product operator terms describing in-phase and anti-phase transverse coherences illustrated on Scheme 1.1.

H_x^1

$2H_x^1C_z$

$2H_x^1H_z^2$

$4H_x^1C_zH_z^2$



Scheme 1.1 : Schematic representation of ^{1}H resonances corresponding to in-phase and anti-phase terms generated after the CH_2 spin-state selective coherence transfer. For certain phases ϕ_5 and ϕ_6, only one ^{1}H multiplet component remains when coadding the spectra. Taken directly from Miclet et al.,[39]

The modified Rance-Kay element, CH_2-S^{3}CT, can select distinctly a component of the four multiplet of a methylene proton by change of phases ϕ_5 and ϕ_5 in between τ_1 and τ_2 conserving the H+- component, the most upfield and best relaxing in the case of the TROSY, when coadding the spectra. But depending on the phases, any component(s) of the multiplet can be selected.

The transverse relaxation rates for each transition are described here : in a regular HSQC experiment, the transverse autorelaxation of a given H^1 proton belonging to a CH^1H^2 methylene group surrounded by H^i protons is defined by [39]:

$R_{HSQC} = R_{CH2} + \sum_i \Gamma^{DD,DD}_{H^1H^i,H^1H^i}$

With :

$R_{CH2} = \Gamma^{DD,DD}_{H^1C,H^1C} + \Gamma^{DD,DD}_{H^1H^2,H^1H^2} + \Gamma^{CSA,CSA}_{H^1,H^1}$ the autorelaxation rate with its different terms written as :

$\Gamma^{DD,DD}_{H^1C,H^1C} = \frac{\left(\xi^{DD}_{H^1C}\right)^2 \tau_C S^2}{5}$ and $\xi^{DD}_{H^1C} = \frac{\mu_0 h \gamma_H \gamma_C}{8\pi^2} \times \frac{1}{r^3_{H^1C}}$, relaxation rate of the dipole-dipole interaction between a proton of the methylene and the carbon atom,

$\Gamma^{DD,DD}_{H^1H^2,H^1H^2} = \frac{\left(\xi^{DD}_{H^1H^2}\right)^2 \tau_C S^2}{4}$ and $\xi^{DD}_{H^1H^2} = \frac{\mu_0 h \gamma_H^2}{8\pi^2} \times \frac{1}{r^3_{H^1H^2}}$, relaxation rate of the dipole-dipole interaction between the two protons of the methylene,

$\Gamma^{CSA,CSA}_{H^1,H^1} = \frac{4\left(\xi^{CSA}_{H^1}\right)^2 \tau_C S^2}{45}$ and $\xi^{CSA}_{H^1} = \gamma_H B_0 \Delta\sigma_H$, the relaxation rate of the CSA-CSA interaction of the proton of the methylene with itself. However, as the spin system is never isolated and is surrounded by a multitude of neighboring protons, on has to incorporate another term that takes these into account :

$\sum_i \Gamma^{DD,DD}_{H^1H^i,H^1H^i} = \frac{\sum_i \left(\xi^{DD}_{H^1H^i}\right)^2 \tau_C S^2}{4}$ with $\xi^{DD,DD}_{H^1H^i} = \frac{\mu_0 h \gamma_H^2}{8\pi^2} \times \frac{1}{r^3_{H^1H^i}}$, the sum of the relaxation rates of the dipole-dipole interaction between the proton of the methylene and the neighboring protons.

In the CH_2-TROSY experiment, cross correlation between dipolar and CSA relaxation rates will attenuate the autorelaxation term :

$R_{TROSY} = R_{CH2} + \sum_i \Gamma^{DD,DD}_{H^1H^i,H^1H^i} - \Gamma_{TROSY}$, the transverse autorelaxation subtracting the cross-correlated relaxation rate, with:

$\Gamma_{TROSY} = \Gamma^{CSA,DD}_{H^1,H^1C} - \Gamma^{CSA,DD}_{H^1,H^1H^2} + \Gamma^{DD,DD}_{H^1C,H^1H^2}$, the cross-correlated relaxion rate, composed of the three terms defined as :

$\Gamma^{CSA,DD}_{H^1,H^1C} = \frac{4\xi^{CSA}_{H^1}\xi^{DD}_{H^1C}P_2\left(\cos\theta^{CSA,DD}_{H^1,H^1C}\right)\tau_C S^2}{15}$, the cross-correlated relaxation between the CSA of the proton and the dipole-dipole interaction between the proton and the carbon,

$\Gamma^{CSA,DD}_{H^1,H^1H^2} = \frac{4\xi^{CSA}_{H^1}\xi^{DD}_{H^1H^2}P_2\left(\cos\theta^{CSA,DD}_{H^1,H^1H^2}\right)\tau_C S^2}{15}$, the cross-correlated relaxation between the CSA of the proton and the dipole-dipole interaction between the two protons,

$\Gamma^{DD,DD}_{H^1C,H^1H^2} = \frac{2\xi^{DD}_{H^1C}\xi^{DD}_{H^1H^2}P_2\left(\cos\theta^{DD,DD}_{H^1C,H^1H^2}\right)\tau_C S^2}{5}$, the cross-correlated relaxation between the dipole-dipole interaction between the proton and the carbon and the dipole-dipole interaction between the two protons. In the CH_2-TROSY spectra, the proton line width is diminished by the Γ_{TROSY} rate and the removal of the $^2J_{H1H2}$ coupling, present in the HSQC. Removing dipolar interactions with neighboring protons will also attenuate the autorelaxation rate further, hence the need to isotopically label proteins with deuterium.

In a TROSY experiment, the transverse relaxation of a given ^{13}C carbon belonging to the same CH^1H^2 methylene group surrounded by H^i protons is defined by :

$$R_{TROSY} = \Gamma^{DD,DD}_{CH^1,CH^1} + \Gamma^{DD,DD}_{CH^2,CH^2} + \Gamma^{CSA,CSA}_{C,C} - \Gamma^{CSA,DD}_{C,CH^1} - \Gamma^{CSA,DD}_{C,CH^2} + \Gamma^{DD,DD}_{CH^1,CH^2}$$

In a similar way as for the proton, it is comprised of two general terms, autorelaxation and cross-correlated relation. Autorelaxation of the carbon is defined by the relaxation rates of dipole-dipole interactions and the CSA-CSA interaction of the carbon with itself. The cross-correlated relaxation is composed of the dipole-dipole and CSA interactions of the carbon with the individual protons. However, since the gyromagnetic ratio of the carbon is about 4 times lower than that of the proton, 267.52 against 67.28 10^6 rad·s^{-1}·T^{-1}, the relaxation rate of the carbon is not dominating the relaxation processes in the $^1H^{13}C$ correlation experiments. Hence, only proton relaxation will be analyzed in this study.

We can define the enhancement factor ε_R as the ratio of the ^1H NMR signal linewidth in the HSQC and in the CH_2-TROSY spectra such that when replacing with the relaxation rates we obtain :

$$\varepsilon_R = \frac{R_{CH2} + \sum_i \Gamma^{DD,DD}_{H^1H^i,H^1H^i} + \pi(^2J_{HH} + {}^3J_{HN,H^1})}{R_{CH2} + \sum_i \Gamma^{DD,DD}_{H^1H^i,H^1H^i} - \Gamma_{TROSY} + \pi\left({}^3J_{HN,H^1}\right)}$$

This enhancement will be experimentally determined by performing linewidth ratios in Chapter IV.

I.3.3. S3-Gly experiments

The CH_2-TROSY is a spin state selective experiment that selects the slowest relaxing cross-peak of the CH_2 multiplet. But in fact, it is possible to select specifically any of the components or groups of components of the multiplet by moving the phase $\phi 2$, ϕ_{RK1}, ϕ_{RK2} (figure 1.11) of the radio -frequency pulses to do so. 2D spin state selective experiments thus allow the fine measurement of ^2J or ^1J couplings. In 2005, Miclet et al.,[48] proposed a 3 dimensional experiment for the measurement of 8 scalar or dipolar couplings in a methine-methylene pair by addition of a third dimension, here the carboxyl. Based on this, the S3-Gly experiment, derived from the CH_2-TROSY experiment, proposes to apply the same principal and demonstrate spin state selective sequences dedicated to glycine residues tuned for methylene-amide or methylene-carboxyl pairs. Two sequences are proposed, the s3-HACACO and the s3-HACAN. Only the s3-HACACO will be detailed in this manuscript. It that excites the Hα of the glycine residues and transfer the magnetization to the Cα via a simple INEPT transfer, the antiphase term will then be transferred to the C' nuclei. The chemical shifts are edited in the first dimension without any decoupling and then the magnetization is transferred back to the Cα of the glycine and then to the Hα for recording spin state selective transitions. Thus, in the HαCα planes, a limited number of correlations are retained (8) and in the first dimension, corresponding to the C', several ^1H-^{13}C planes are edited, selecting alternate components of the methylene multiplet.

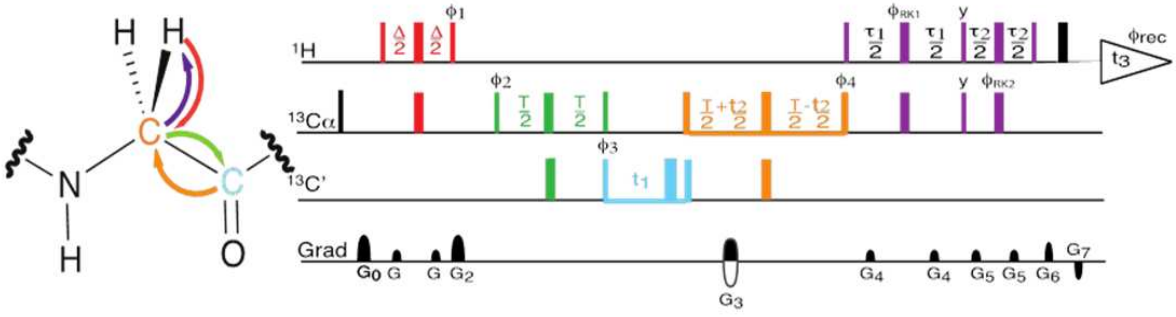


Figure 1.11 : Spin state selective pulse sequence s3-HACACO (top panel). Each color represents a period of the sequence associated to a transfer of magnetization on the glycine residue. The thin bars represent 90° pulses whereas the thick bars represent 180° pulses. When not specified, the phase of the impulsion is x. Delays : Δ= 1/2(*J*CαHα) and T=1/2(*J*CαC').

The pulse sequences for the s3-HACACO spin state selective 3D experiment is given (Figure 1.11), with each color representing a period during which magnetization transfer is undertaken. In red, an INEPT sequence[32] transfers magnetization from the Hα to the Cα with

a selective pulse on the Cα of glycine residues which have particular chemical shifts compared to the Cα of other residues, creating an antiphase term. In green, another INEPT sequence transfers the magnetization from the Cα to the C'. The blue period is the evolution of the magnetization in the first dimension corresponding to the chemical shift of the C' nucleus for a time t_1. There are no decoupling pulses during this time so as to conserve the spin states of neighboring spins. The orange period is the transfer of magnetization from the C' back to the Cα which will evolve for a time t_2, incremented by the displacement of the 180° pulse, thus modulating chemical shift and which will give the second dimension. Finally the purple sequence is the spin state coherence transfer, (modified Rance-Kay) element which brings the magnetization from the Cα back to the proton and selects the components of interest with τ1 and τ2. If all the ^{1}H ^{13}C couplings were expressed during $^{13}C'$ edition, the spectra obtained would be extremely complex. The Rance-Kay element enables us to simplify the spectra by selecting only some coherences of interest acquired on each $^{13}C'$ plane. The signal recording is in black.

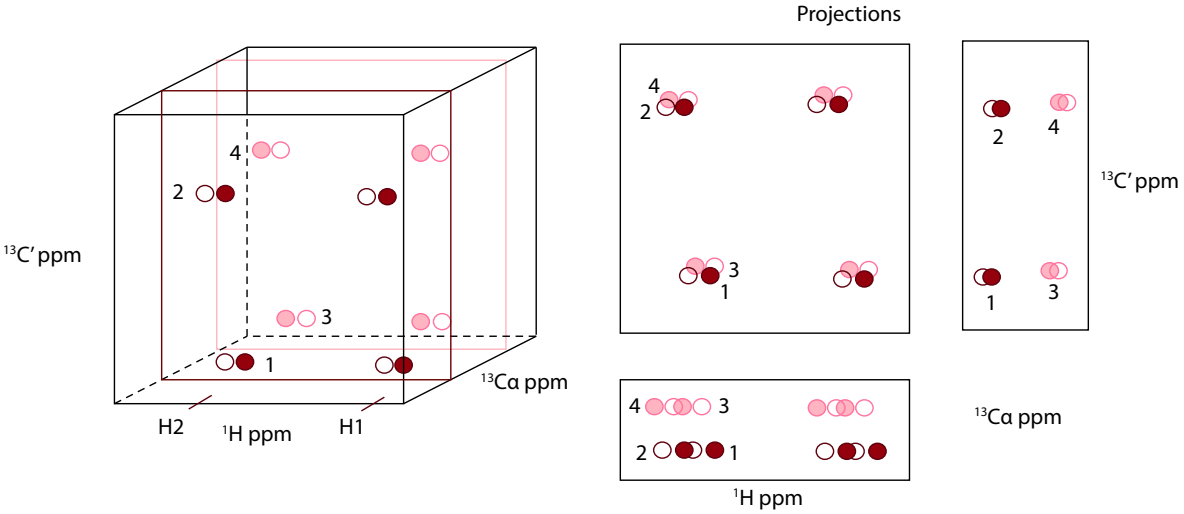


Figure 1.12 : Theoretical spectra for the HACACO experiment completely coupled, and the 2D projections on the right. The individual spin states are represented as full red or pink circles.

Such three-dimensional experiments allow expression of a number of coupling constants (and residual dipolar couplings if the experiment were recorded in anisotropic conditions). Measurements of these couplings may offer precise information about the orientation and dynamics of the glycine residue. In this optic, Chapter IV proposes to look at the structural information such experiments can give when applied to a protein system.

I.4. Dissolution-Dynamic Nuclear Polarization

I.4.1 Introduction

As mentioned before, one of the great drawbacks of NMR for biomolecular investigation is the innate lack of sensitivity of nuclei in large biomolecular systems, due, in part, to their low gyromagnetic ratios, the low population bias at thermal equilibrium and magnetic moment that stems from this. When placed in a magnetic field, non-zero spins will be subject to the Zeeman effect and will follow the Boltzmann distribution which creates longitudinal magnetization along the z axis. The net magnetization created by the segregation of the spins determines the sensitivity of an NMR experiment. Because net magnetization is defined by the polarization (P) of the spins to the different energy states with $P \approx \frac{\Delta E}{2k_B T} = \frac{\gamma \hbar B_0}{2k_B T}$ and because the gyromagnetic ratios of nuclei are low, polarization is very weak.

In the expression of the polarization P, three factors appear that can modulate the polarization: magnetic field B_0, temperature and gyromagnetic ratio. However, increasing the magnetic field does not change polarization in any significant way. For a current field of 9.4 T (400 MHz), proton polarization at room temperature (25°C) is 0.0032%, but for the same temperature using the highest magnetic field available today (28.2 T, 1.2GHz) proton polarization still only reaches 0.0096%. Moreover, modifying temperature is not adapted for liquid NMR, as cryogenic temperatures are needed for any significant impact on polarization. However, even at 1 K, although increasing of a factor of 200 compared to room temperature, polarization still averages less than 1% for the hydrogen atom at 400 MHz.

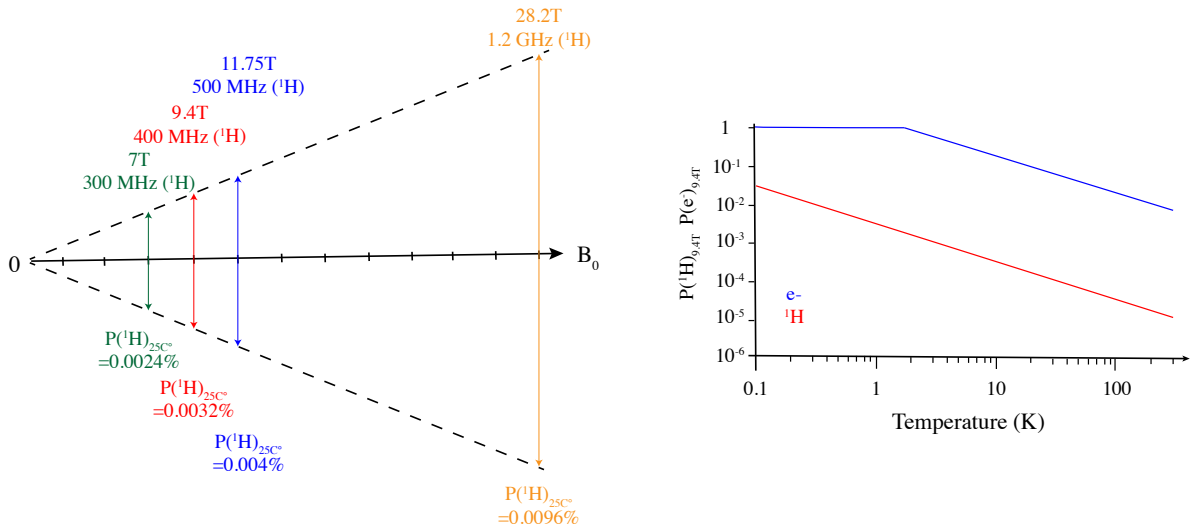


Figure 1.9 : Resulting polarization for the hydrogen atom and electron at room temperature 25°C at different field strengths and polarization at different temperatures for 9.4 T.

Although the intrinsic magnetic moment of an element cannot be changed, polarization transfer from one particle to another is possible. Electrons also present a spin of ½ and their Landé factor is 660 times greater than that of the gyromagnetic ratio of the hydrogen atom. Thus, the energy difference between the two populations in a given magnetic field is 660 times greater, increasing polarization of spin states significantly. Polarization transfer from electron to nucleus is a way to artificially increase signal of nuclei and NMR experiments with such polarization would be extremely sensitive.

The first Dynamic Nuclear Polarization (DNP) experiment was undertaken in 1953[49] after Overhauser had theoretically proven its feasibility. In this experiment, Carver and Slichter polarized the nucleus of Lithium through NOE polarization transfer, described in the next section[7]. Li is paramagnetic as it possesses an unpaired electron on its outer layer. At a low magnetic field, they applied microwaves corresponding to the energy of the electrons to their lithium sample and observed how, by continuous wave NMR, they could record a signal 100 times more sensitive as polarization was transferred to the nucleus. However, DNP remained largely undeveloped because of the lack of sophistication of instruments and remained a solid-state technology for several decades.

But recently, interest in DNP has resurfaced for liquid state NMR, with the improvement of instrumentation. Indeed, in 2003, Ardenkjær-Larsen et al.,[50] managed to conserve hyperpolarization of ^{13}C and ^{15}N atoms of urea in solution after electron transfer polarization at solid state. To manage this feat, they polarized their sample in a polarizer at 1.2 K, irradiating with microwaves their sample, previously mixed with a radical component, the source of electrons. The sample was then dissolved by injection of boiling water and rapidly passed through a magnet, into a spectrometer at room temperature, thus its name Dissolution-DNP (D-DNP). Although loss of polarization during transfer was evident, they conserved a notable percentage of polarization for the ^{13}C and ^{15}N atom and demonstrate that they obtained signal intensities 10,000 times superior than for the same sample without polarization. They opened a gate to numerous solution applications, and D-DNP is now a promising approach for protein studies with numerous recent developments which are looked at in detail in Chapter V.

I.4.2. Electron polarization transfer

Several mechanisms of polarization transfer between electrons and nuclei that are not scalar coupled can occur: solid effect, cross effect, thermal mixing and electron-nucleus Overhauser effect. The latter, commonly known as nOe (nuclear Overhauser[51] effect) transfer, is the best understood as it also occurs between nuclear spins. When two ½ spins are coupled in a strong magnetic field, S the electron and I the nucleus, they can be either in the α or β state. At Boltzmann distribution, there is strong polarization of the electron therefore the majority of spins will be in the α state, this is not the case for the nucleus though, which displays low polarization even at cryogenic temperatures. When the system is perturbed by microwaves matching the electron transition, the majority of electron spins will be transferred to the β state. Relaxation provokes transitions of the spins. Single quantum transitions will include only one spin state change but do not transfer polarization. However, double and zero quantum transitions, where both spins change of state, will transfer polarization of NOE type between two coupled spins, with a positive hyperpolarization for cross relaxation dominated by double quantum transition or negative polarization for zero-quantum transition.

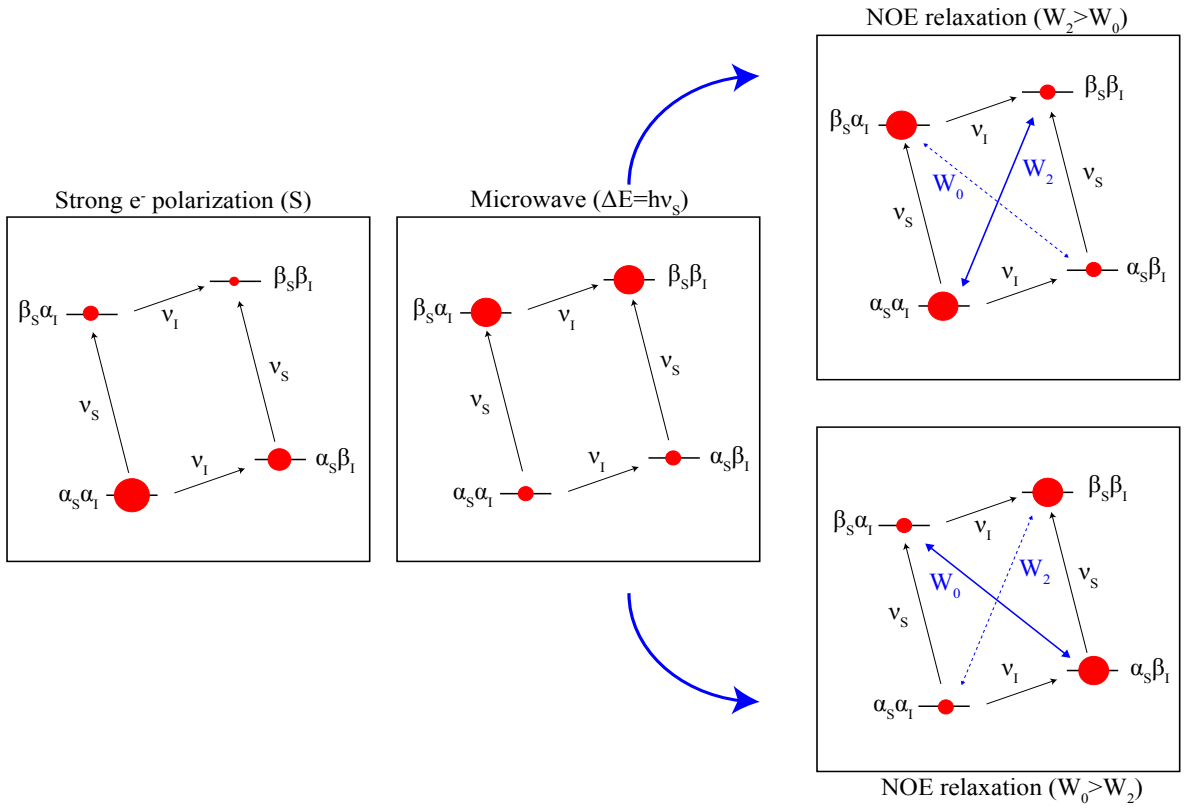


Figure 1.10 Schematic representation of the Nuclear Overhauser Effect (nOe). The bars represent the different spin states, and the red dots the number of spins in each state.

1. Abragam, A. *The principles of nuclear magnetism.* (Clarendon Press, 1961).
2. Keeler, J. *Understanding NMR Spectroscopy*. (John Wiley & Sons, 2011).
3. Bloembergen, N. *Nuclear Magnetic Relaxation*. (W.A. Benjamin, Inc., 1948).
4. Derome, A. E. *Modern NMR techniques for Chemistry Research*. **6,** (1987).
5. Sattler, M. & Fesik, S. W. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4,** 1245–1249 (1996).
6. Gardner, K. H. & Kay, L. E. The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 357–406 (1998).
7. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *J. Am. Chem. Soc.* **116,** 11655–11666 (1994).
8. Venters RA, Huang CC, Farmer BT 2nd, Trolard R, Spicer LD, Fierke CA. High-level 2H/13C/15N labeling of proteins for NMR studies. *J Biomol NMR.* 339-344. (1995).
9. Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V. & Veglia, G. in *Isotope labeling in Biomolecular NMR* (ed. Atreya, H. S.) **992,** 35–62 (Springer Netherlands, 2012).
10. LeMaster, D. M. & Richards, F. M. NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. *Biochemistry* **27,** 142–150 (1988).
11. Metzler, W. J., Wittekind, M., Goldfarb, V., Mueller, L. & Farmer, B. T. Incorporation of ^{1}H/ ^{13}C/ ^{15}N-{Ile, Leu, Val} into a Perdeuterated, ^{15}N-Labeled Protein: Potential in Structure Determination of Large Proteins by NMR. *J. Am. Chem. Soc.* **118,** 6800–6801 (1996).
12. Miyanoiri, Y., Takeda, M., Terauchi, T. & Kainosho, M. Recent developments in isotope-aided NMR methods for supramolecular protein complexes –SAIL aromatic TROSY. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1864,** 129439 (2020).
13. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR †. *Biochemistry* **36,** 1389–1401 (1997).
14. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *Journal of Biomolecular NMR* **13,** 369–374 (1999).
15. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression. *Journal of the American Chemical Society* **136,** 11308–11310 (2014).
16. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. & Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angewandte Chemie International Edition* **49,** 1958–1962 (2010).
17. Religa, T. L., Ruschak, A. M., Rosenzweig, R. & Kay, L. E. Site-Directed Methyl Group Labeling as an NMR Probe of Structure and Dynamics in Supramolecular Protein Systems: Applications to the Proteasome and to the ClpP Protease. *J. Am. Chem. Soc.* **133,** 9063–9068 (2011).
18. Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P. & Boisbouvier, J. An optimized isotopic labelling strategy of isoleucine-γ_2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **48,** 1434–1436 (2012).
19. Ruschak, A. M., Velyvis, A. & Kay, L. E. A simple strategy for 13C,1H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *J Biomol NMR* **48,** 129–135 (2010).
20. Ozawa, K., Headlam, M. J., Schaeffer, P. M., Henderson, B. R., Dixon, N. E. & Otting, G. Optimization of an Escherichia coli system for cell-free synthesis of selectively 15N-labelled proteins for rapid analysis by NMR spectroscopy: Cell-free synthesis of 15N-labelled proteins. *European Journal of Biochemistry* **271,** 4084–4093 (2004).
21. Rosenzweig, R. & Kay, L. E. Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR. *Annu. Rev. Biochem.* **83,** 291–315 (2014).
22. Amero, C., Schanda, P., Durá, M. A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B. & Boisbouvier, J. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *Journal of the American Chemical Society* **131,** 3448–3449 (2009).
23. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **20**, pages71–75 (2001)
24. Cai, M., Huang, Y., Yang, R., Craigie, R. & Clore, G. M. A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *Journal of Biomolecular NMR* **66,** 85–91 (2016).
25. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. & Li, Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **18,** 936–948 (2009).
26. Hoffmann, B., Löhr, F., Laguerre, A., Bernhard, F. & Dötsch, V. Protein labeling strategies for liquid-state

NMR spectroscopy using cell-free synthesis. *Progress in Nuclear Magnetic Resonance Spectroscopy* **105,** 1–22 (2018).
27. Stevanato, G., Hill-Cousins, J. T., Håkansson, P., Roy, S. S., Brown, L. J., Brown, R. C. D., Pileio, G. & Levitt, M. H. A Nuclear Singlet Lifetime of More than One Hour in Room-Temperature Solution. *Angew. Chem. Int. Ed.* **54,** 3740–3743 (2015).
28. Kurzbach, D., Canet, E., Flamm, A. G., Jhajharia, A., Weber, E. M. M., Konrat, R. & Bodenhausen, G. Investigation of Intrinsically Disordered Proteins through Exchange with Hyperpolarized Water. *Angewandte Chemie International Edition* **56,** 389–392 (2017).
29. Zhang, G. & Hilty, C. Applications of dissolution dynamic nuclear polarization in chemistry and biochemistry. *Magn Reson Chem* **56,** 566–582 (2018).
30. Ragavan, M., Chen, H.-Y., Sekar, G. & Hilty, C. Solution NMR of Polypeptides Hyperpolarized by Dynamic Nuclear Polarization. *Anal. Chem.* **83,** 6054–6059 (2011).
31. Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J. & Laue, E. D. An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *J. Am. Chem. Soc.* **118,** 407–415 (1996).
32. Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **101,** 760–762 (1979).
33. de la Vega-Hernández, K. & Antuch, M. The Heteronuclear Single-Quantum Correlation (HSQC) Experiment: Vectors versus Product Operators. *J. Chem. Educ.* **92,** 482–487 (2015).
34. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94,** 12366–12371 (1997).
35. Riek, R., Wider, G., Pervushin, K. & Wuthrich, K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences* **96,** 4918–4923 (1999).
36. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in ^{13} C-Labeled Proteins. *J. Am. Chem. Soc.* **120,** 6394–6400 (1998).
37. Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. & Simorre, J.-P. Improved Sensitivity and Resolution in ^{1} H− ^{13} C NMR Experiments of RNA. *J. Am. Chem. Soc.* **120,** 11845–11851 (1998).
38. Fiala, R., Czernek, J. & Sklenář, V. Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *Journal of Biomolecular NMR* **16,** 291–302 (2000).
39. Miclet, E., Williams, D. C., Clore, G. M., Bryce, D. L., Boisbouvier, J. & Bax, A. Relaxation-Optimized NMR Spectroscopy of Methylene Groups in Proteins and Nucleic Acids. *J. Am. Chem. Soc.* **126,** 10560–10570 (2004).
40. Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L. E. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *J. Am. Chem. Soc.* **126,** 3964–3973 (2004).
41. Rashid, S., Lee, B. L., Wajda, B. & Spyracopoulos, L. Side-Chain Dynamics of the Trifluoroacetone Cysteine Derivative Characterized by ^{19} F NMR Relaxation and Molecular Dynamics Simulations. *J. Phys. Chem. B* **123,** 3665–3671 (2019).
42. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89,** 496–514 (1990).
43. Clubb, R. T. A Constant-Time Three-Dimensional Triple Resonance Pulse Scheme to Correlate Intra-residue HN, 15N, and 13C'Chemical Shifts in 15N- 13C-Labeled Proteins. 5
44. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99,** 201–207 (1992).
45. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced ^{1} H− ^{13} C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society* **125,** 10420–10428 (2003).
46. Kay, L., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114,** 10663–10665 (1992).
47. Parella, T. & Belloc, J. Spin-State-Selective Excitation in Selective 1D Inverse NMR Experiments. *Journal of Magnetic Resonance* **148,** 78–87 (2001).
48. Miclet, E., Boisbouvier, J. & Bax, A. Measurement of eight scalar and dipolar couplings for methine?methylene pairs in proteins and nucleic acids. *Journal of Biomolecular NMR* **31,** 201–216 (2005).
49. Carver, T. R. & Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **92,** 212–213 (1953).
50. Ardenkjær-Larsen, J. H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M. H., Servin, R., Thaning, M. & Golman, K. Increase in Signal-to-Noise Ratio of >10,000 Times in Liquid-State NMR.

Proceedings of the National Academy of Sciences of the United States of America **100,** 10158–10163 (2003).
51. Overhauser, Albert W. Polarization of Nuclei in Metals. Physical Review. 92 (2): 411–5 (1953).
52 Lionel Imbert, Rachel Lenoir-Capello, Elodie Crublet, Alicia Vallet, 4 Rida Awad, Isabel Ayala, Celine Juillan-Binard, Hubert Mayerhofer, Rime Kerfah, Pierre Gans, Emeric Miclet, & Jerome Boisbouvier. In Vitro Production of Perdeuterated Proteins in H_2O for Biomolecular NMR Studies.Chapter 8. Springer Science: Structural Genomics: General Applications, Methods in Molecular Biology, vol. 2199, 127-129. (2021). DOI:10.1007/978-1-0716-0892-0_8,

.

Chapter II : Structural properties of the two proteins studied

Two proteins of similar sizes (16 and 18 kDa) were employed for this work. A model protein termed H23 was used for cell-free synthesis protocol optimization, produced in collaboration with Lionel Imbert and Jérôme Boisbouvier from the Cell-free platform at Institut Biologie Structurale in Grenoble. However, the limited available structural information of this protein were a drawback for the application of novel NMR experiments for structural and dynamic analysis. Applications were therefore undertaken on another model protein, Pin1, a well-documented peptidyl prolyl isomerase for which there exists many NMR structural studies (*vide infra*). Expression of Pin1 and its separate domains by *E.coli* expression were undertaken with the help of Soha Abou Ibrahim, a PhD student at the Laboratoire des Biomolécules. This chapter is dedicated to presenting the structural properties of these two models and their use in this thesis.

II.1. Pin1: a brief overview

II.1.1. Introduction

Phosphorylation of proteins is an essential, reversible post-translational modification, implicating the transfer of a phosphate from a high energy molecule (usually ATP) on an amino acid: tyrosine (1-2%), threonine (4-12%) and serine (84-95%)[1,2]. This activity is undertaken by protein kinases (there are over 518 human kinases described today[3]) and is reversible by phosphatase activity. Phosphorylation of a protein can induce a change in its conformation[2], caused by the high negative charge of the phosphate group, ultimately modifying its function, inducing either its activation, deactivation or interaction with other proteins. It is a key mechanism in the control of different molecular pathways such as cell cycle[4], metabolism[5] and cell signaling[6] amongst many, making them crucial physiological enzymes. The peptide bond between two amino acids can take two torsion angle values, 0° (*cis*) or 180° (*trans*), which will greatly affect the structure of the protein. If the protein is in the inappropriate conformation, then its function can be significantly affected, altering cascades of metabolic reactions that can lead to severe diseases. Because of the high energy barrier for the rotation of such bonds (see section II.1.2), there is the need for proteins that can catalyze this reaction, restoring the correct conformation and activity of the protein.

Pin1 is a human parvulin from the class of peptidyl prolyl *cis-trans* isomerases, catalyzing the isomerization of *cis-trans* and *trans-cis* of phospho-serine/phospho-threonine-proline motifs (pS/pT-P) only. It was discovered in 1996 as a NIMA kinase interacting protein, an enzyme essential to cell-cycle progression[7]. Since, numerous studies have been undertaken on Pin1, bringing to light its regulation of numerous metabolic pathways, with over 70 known partners, and its implication in a number of diseases, notably several cancers[8,9] and neurodegenerative diseases[10–12]. This brief review will present its structure, mechanism of action and some of its most well-known partners as well as the model choice for our study.

II.1.2 Activity, structure and mechanism of action

In proteins, the ω dihedral angle which defines the geometry of the planar peptide bond can take two values: 0° (*cis*) and 180° (*trans*) because of the partial double bond character of this liaison (40%). The *trans* conformation, much more stable energetically, is present 99.85% of the time whilst the *cis* conformation, less stable at 0.15% for secondary amides in polypeptides[13]. These populations fall to 94% and 6% respectively for the *trans* and *cis* conformations for tertiary amide bonds of prolines, the *trans* conformation being less stable[14,15]. This leads to the existence of two possible conformations of the protein with specific activity associated to each of them. However, there is a high activation energy (about 20 kcal/mol[16] in peptides) needed to overcome the rotational barrier between these two states, hence interconversion is slow (<0.002/s at 25 °C)[17]. Pin1 acts by lowering the energy barrier necessary to achieve this isomerization (Figure 2.1) and accelerates the reaction by a factor of 100 to 1000[7,18] for pS/pT-P motifs only.

Human Pin1 is 163 amino acids long (Figure 2.2), with a molecular weight of 18,243 Da (Uniprot) having two domains: WW (named after the two conserved tryptophans W11 and W34) spanning residues M1-G39 which possesses a reported tenfold heightened binding affinity[19,20] to its peptide substrates in vitro compared to the larger PPIase catalytic domain, spanning residues G50-E163[21]. The WW domain recognizes preferentially the *trans*[22,23] conformation of the peptidyl-prolyl motifs but is not necessary for catalytic activity[20], whereas the PPIase domain recognizes preferentially *cis* conformations. Interestingly, increased catalytic activity of full length Pin1 was reported compared to the PPIase domain alone[24]. The two domains only weakly interact in the absence of a substrate[24,25] and are associated by a flexible linker (residues N40-Q49) characteristically rich in serines and glycines[67].

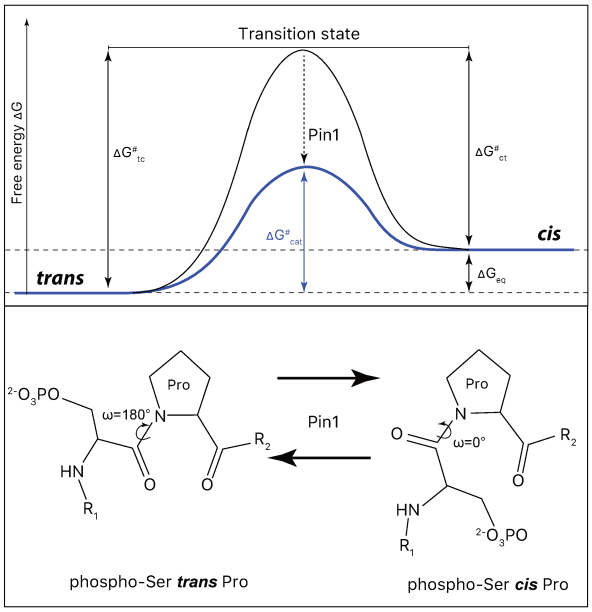


Figure 2.1: Energy level diagram of catalytic activity of Pin1. The *trans* pS-Pro and *cis*-pS-P motifs are illustrated below. Black line : energy path without Pin1, blue line : energy path with Pin1. $\Delta G^{\#}_{tc}$= Energy activation from trans conformation to transition state, $\Delta G^{\#}_{tc}$= Energy activation from transition state to cis conformation. ΔG_{eq}= Difference of free Gibbs energy between the trans and cis conformations, $\Delta G^{\#}_{cat}$= activation energy for catalyzed reaction.

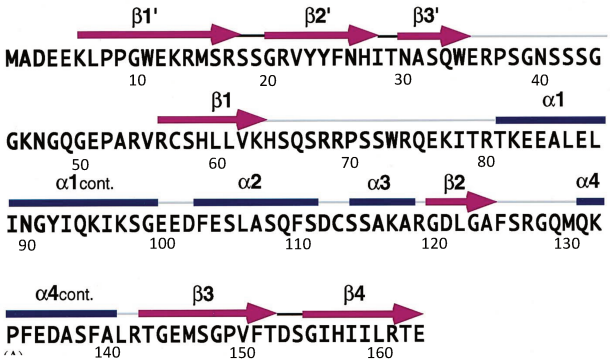


Figure 2.2 : Amino acid sequence of the human peptidyl prolyl *cis-trans* isomerase Pin1. The magenta arrows represent the β-strands whilst the blue bars represent the α-helixes.[20]

Figure 2.3 : Molecular structure of human Pin[26] (PDB : 1PIN) without ligand binding. Structure determined by X-ray diffraction in 1997, resolution of 1.35 Å [21].

The small WW domain has been found in many functionally diverged proteins[27,28]. It consists of three beta strands that form an antiparallel beta sheet and two loops. Loop I situated between S16 and R21 is rate limiting for Pin1 refolding[29,30], and plays an important role in ligand recognition and interaction[20,31], the binding pocket including residues S16, R17, Y23 and W34[23,31]. Indeed, redesign of this loop influences folding kinetics and ligand recognition[29,30], mutations to S16 (documented as a phosphorylation site), R17, Y23 and W34 lead to decreased binding affinity to its substrate[23,31,32] and cells carrying those mutations reportedly are not viable[31]. Loop II situated between H27 and N30 is at the interface with the PPIase domain, displaying chemical shifts perturbations when in contact with a substrate[19,24]. The WW domain shows enhanced thermodynamic stability, with a Tm of 59.0°C[29] (melting temperature at which the globular structure is unfolded), regaining its fold rapidly[30] and conserving its ligand-binding properties after denaturation[30], characteristics that will be used in NMR studies (Chapter VI).

The PPIase domain is constituted of a central β sheet of four anti-parallel strands facing helix α3 in a "gripping-hand" topology. A long loop links strand β1 to helix α1 (K63 -R80) which make up the recognition center, the positively-charged, basic side-chains of K63, R68 and R69 encircling the negatively charged phosphate of the substrate to facilitate binding[21,33,34]. The catalytic site consists of residues H59, S115, C113, L122, M130, F134, T152, and H157[21,22,35] and the interface domain consists of residues S138-R142. The residues implicated in this interface play a significant role in interdomain communication and function regulation of Pin1 following substrate binding[35,36].

There is no general consensus on the mechanism of action of Pin1 upon substrate interaction, although several have been proposed[26,34] (Figure 2.4) highlighting the potentially diverse binding mechanisms needed to bind to the variety of known biological partners and targets. The WW domain being devoid of catalytic activity but heightened biding affinity could suggest that it helps Pin1 in targeting substrates and increase the local concentration of Pin1[24] so as to exert catalytic activity. Thus, the sequential binding model (Figure 2.4 A) proposes that the substrate is first recruited by the WW domain which recognizes and binds the pS/T-P motif, before allowing the PPIase domain to bind to another pS/T-P motif on the same target (multi-site binding)[20]. It has also been proposed that the binding of the WW domain induces a change in conformation which will prompt its displacement in favor of PPIase binding to the same motif for catalytic activity (single-site binding)[37]. The multimeric binding model (Figure 2.4 B)

is based on a multi-complex mechanism of action, whereby a Pin1 substrate is in a complex with an active kinase. In this scenario, the WW domain recruits the kinase on a pS/T-P motif which subsequently phosphorylates the S/T-P motif[24]. The PPIase domain, in close proximity to the newly phosphorylated substrate, can then bind and catalyze the pS/T-P motif[24]. Certain phosphatases, such as PP2A, require the pS/T-P motif to be in the *trans* conformation for catalytic activity. Moreover, previous studies have shown that the PPIase domain is also capable of substrate binding, independently of the WW domain, via the loop created by the basic side chains of residues K63, R68, and R69[21]. The catalysis-first binding model (Figure 2.4 C) therefore suggest that the PPIase domain of Pin1 first binds to the *cis* conformation of the pS/T-P motif to catalyze the isomerization to the *trans* configuration and, as the WW domain binds preferentially the *trans* configuration state[38], its binding would have a stabilizing effect on the substrate to remain in this conformation for dephosphorylation to occur[39,40]. Finally, the simultaneous binding model (Figure 2.4 D) suggest that Pin1 interacts with the substrate when both domains simultaneously bind a pS/T-P motif of the substrate. This model was proposed after noticing that the WW and PPIase domains have different affinities depending on the amino acid following the pS/T-P motif[34].

Figure 2.4 : Proposed models of mechanism of action of the Pin1 protein[26] directly taken from Lee and Liou, 2018 and adapted from Innes et al., 2013[34].

II.1.3. Partners and biological implications

Pin1 determines the fate of phosphoproteins by regulating the *cis/trans* pS/T-P motif configuration. This change in the peptide bond configuration alters the local conformation of the protein, which can, in turn, alter its function, stability and localization which can have important metabolic consequences. Indeed, with over 70 known substrates[34] of Pin1, a number are implicated in cell cycle progression, oncogenesis, neurodegenerative diseases and immune response. The figure below (Figure 2.5[41]) illustrates the numerous metabolic pathways in which Pin1 is implicated via its catalytic activity, with some examples detailed below.

Figure 2.5: Pin1-catalysed prolyl isomerization regulates a spectrum of target activities[41]

In the cell cycle, Pin1 induces a conformational change in Cdc25, an important phosphatase in the control of cell cycle entry in phase S and mitosis which dephosphorylates and activates the cyclinB/Cdc2 complex. The conformational change induces dephosphorylation of Cdc25 and inhibition of its catalytic activity[20,42,43,44]. A number of identified protein substrates of Pin1 are also implicated in initiation and progression of different cancers[9], stabilized by its catalytic activity. Notably, Pin1 isomerizes cyclin D, stabilizing this substrate and promoting its accumulation in the nucleus, where it drives cell cycle progression[45]. In the Ras signaling pathway, Pin1 can isomerize c-Jun which has the effect of stabilizing and increasing its transcriptional activity, inducing expression of genes involved in anti-apoptosis and proto-oncogenesis[46]. The isomerization by Pin1 of transcription factor NF-κB, involved in inflammation-induced carcinogenesis, stabilizes the protein, increasing its

nuclear translocation and inducing constitutive activity in many cancers[47]. This transcription factor then enhances the transcription of genes encoding cyclin D1, c-Myc and Bcl-2[47]. In Alzheimer's disease, the *cis* isomer of the pS/T-P motifs in tau leads to its hyperphosphorylation, aggregation and accumulation in neurons as it is more resistant to dephosphorylation and degradation that the *trans* conformer. Pin1 converts the *cis* to *trans* conformation stimulating its dephosphorylation and degradation[48,49] although this theory has been disputed[50].

The multitude of identified different partners of Pin1 and signaling pathways it is implicated in could explain the variability of residues around the binding motif. The hexapeptide around the pS/T-P motif does not have a consensus sequence, although proline, serine, leucine and alanine were the most common amino acids surrounding it[34]. Moreover, binding sites with a proline at the +1 position, preferentially targeted by the WW domain, were often found to be in disorganized, extended or surface loops in flexible non structured regions which could facilitate a 180° rotation[34] whilst peptides preferentially binding PPIase at the catalytic site were often in curved regions. Finally, Pin1 was often found to recognize several (at least two) sites spaced by 2 to 20 amino acids on a target protein[34], reinforcing the idea that there may be more than one mechanism of action, depending on the target and tissue localization. Two new phosphorylated peptide ligands will be used in this work, one of which contains a trifluoromethyl group (see chapter IV)

II.1.4. Choice of Pin1 model

There is a multitude of available research on the Pin1 protein, notably on its structure and dynamics, with many known partners. Moreover, the numerous glycines (15, of which 3 in the WW domain and 9 in the PPIase domain) well dispersed across its sequence and in different secondary structure elements (G10, G20, G120, G123, G144, G148 and G155 near or included in a β-strand, well dispersed on ^{1}H-^{15}N NMR spectra, G39, G50 and G128 in a loop and G91 and G99 in a α-helix), made Pin1 a good candidate for this study. Indeed, glycine residues are extremely useful for methylene specific NMR experiments. Glycines could enable fine measurements of a number of coupling constants in different secondary structure elements as the methylene is directly on the backbone, with or without the presence of known ligands for which structural and dynamics information already exists. The *J* couplings are averaged in cases of intense dynamic therefore would enable to confirm already existing data. For the CH_2-

TROSY experiments, Pin1 is a good model protein to envision a high molecular weight complex so as to evaluate the gain in resolution and sensitivity obtained depending on the molecular weight of the system. In this aspect, the PPIase domain was used to demonstrate the gains in resolution and sensitivity obtained for CH_2-TROSY spectra compared to ^{1}H-^{13}C HSQC spectra for ligand binding studies, and determine chemical shift perturbations of Cα of glycines which give complementary information for backbone elements, acting as a local probe of the backbone

The WW domain, because of its remarkable folding properties and its small size (5 kDa) is a model candidate for DNP studies. Indeed, the high temperatures reached during the experiment (See Chapter V) require a protein capable of rapid refolding after temperature denaturation. Moreover, the conservation of the ligand binding properties could be useful to probe interactions in screening assays. As a three stranded β-sheet, the WW domain is perhaps not the most rapid folding of proteins, β-sheets being slower to form than helices considering the long distance hydrogen bonds present in its structure. Homeodomains, which contain only α–helixes, would regain their structure in an even more rapid manner, and could also be subject for DNP studies on proteins. Finally, the WW domain can be purified by HPLC and lyophilized easily without concerns about structure loss or degradation which was a key property to undertake DNP studies on this domain.

II.1.5. Expression of different constructs of Pin1 in *E.coli* and assignment

Pin1 and its separate domains PPIase and WW were expressed using the Marley protocol in *E.coli* and purified using different protocols, as detailed in section II.3.1. Using already published attributions, the ^{1}H-^{15}N amide correlations were assigned for each construct. The spectra are available on the next pages (Figures 2.6, 2.7 and 2.8), with the amino acid number next to the assigned correlations. Amino acids for which a correlation is missing are generally situated in flexible regions exposed to the solvent, the pH thus being too basic to limit the H-N exchange. The recording conditions for each construct are detailed in section II.3.4.

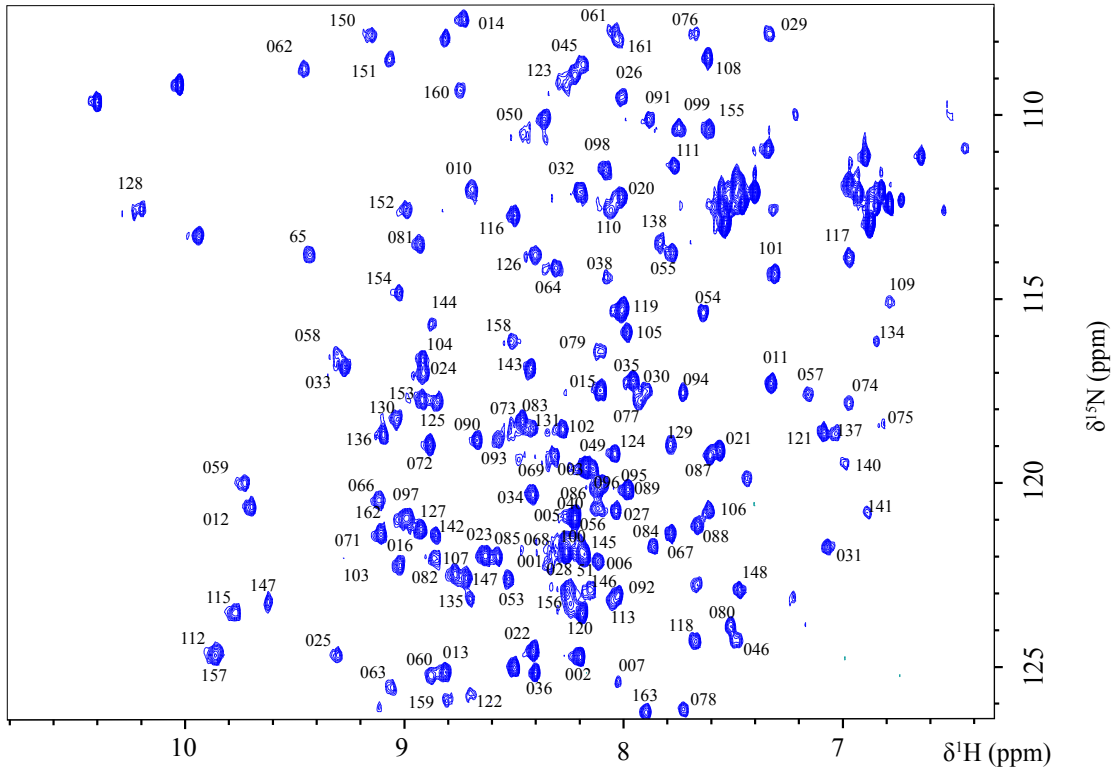


Figure 2.6: ^{15}N-^{1}H HSQC spectra of full-length Pin1 protein at 500 MHz (Bruker Cryoprobe, 150 µM, Tris Buffer 30 mM, pH 7, NaCl 50 mM, 1 mM DTT, 10% D_2O, 1% DSS, 25°C).

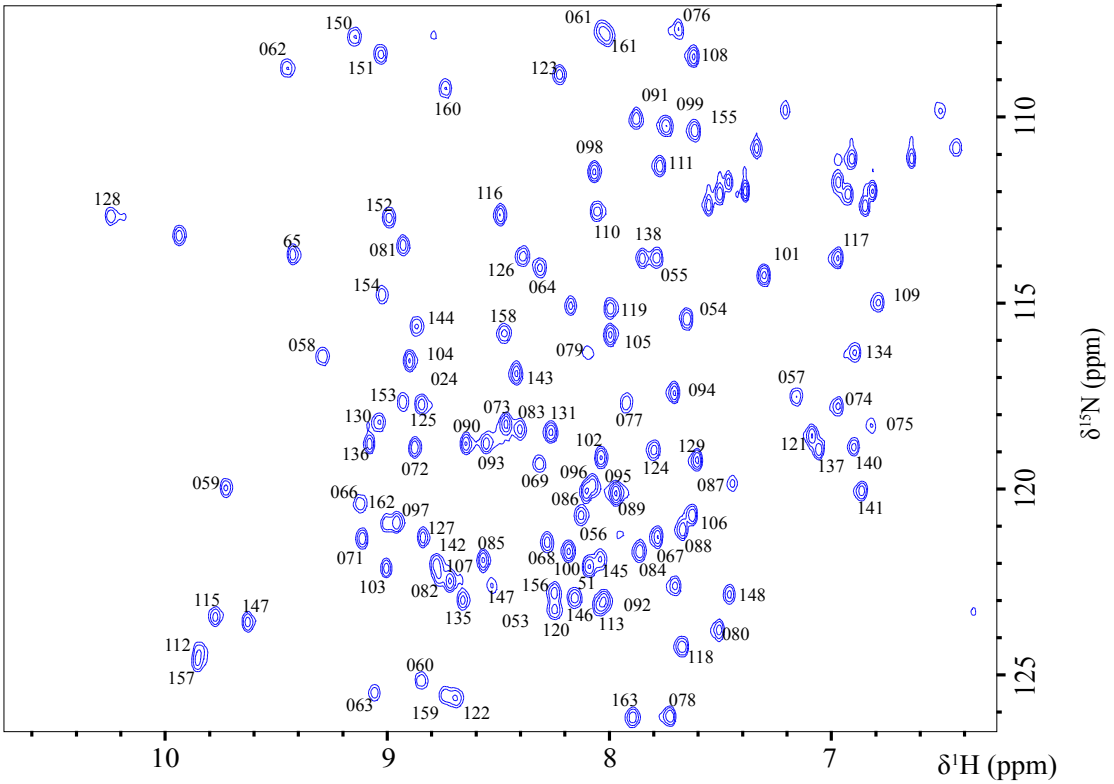


Figure 2.7: ^{15}N-^{1}H HSQC spectra of the PPIase domain of Pin1 protein at 500 MHz (Bruker Cryoprobe, 230 µM, Phosphate Buffer 30 mM, pH 7, KCl 50 mM, 2 mM DTT, 1 mM EDTA, 0.05% NaN_3, 10% D_2O, 1% DSS, 25°C).

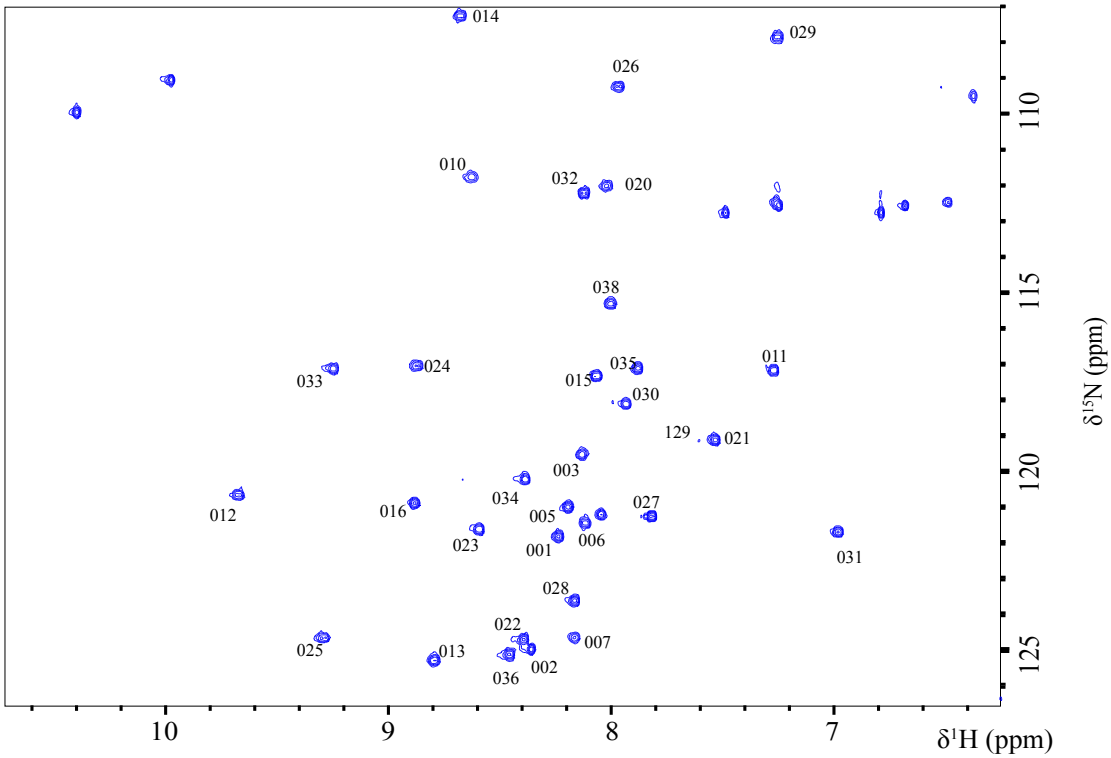


Figure 2.8: ^{15}N-^{1}H HSQC spectra of the WW domain of Pin1 protein at 500 MHz (Bruker Cryoprobe, 150 µM, Tris Buffer 30 mM, pH 6.85, NaCl 50 mM, 1mM DTT, 10% D_2O, 1% DSS, 25°C).

II. 2. NMR Study of H23 protein

II.2.1. Introduction and choice of H23 model

H23 is a small, nucleotide binding subunit of a heavy metal ATPase from *Arabidopsis Thaliana* named HMA8 present in the lumen of thylakoids in chloroplasts and for which there have been few published studies[51–54]. The structure of related proteins was determined by X-ray crystallography previously[55–57]. The X-ray structure of H23, illustrated below, presents six β-strands: The first is from residue 27 to 34 included, the second from residue 75 to 78, the third from residue 81 to 87, the fourth from residue 90 to 95, the fifth from residue 136 to 142 and the sixth from residue 146 to 153. There are also four α-helices : from residue 36 to 47, from residue 53 to 65, from residue 97 to 103 and the last from residue 110 until 121. Residues K123 to S134 are absent in the crystallographic structure, this fragment between α-helix 4 and the β-strand 5 is most probably a very dynamic region in the protein. In the X-ray data, the electron density of R103 corresponds to the superposition of two different backbone conformations. This may reflect the presence of a hinge between α helix 3 and α helix 4 which can adopt two conformations in exchange. The following residues (F104 to D109) do not present a specific secondary structure and may be very dynamic in solution. Helix α2 is

characterized by an important dynamic (mean B factor of 55) compared to the other secondary structure elements (see Graph 1). This helix is itself situated between two particularly dynamic regions (K48-H52 and N66-G74) which could explain the structural differences observed in solution, *vide infra*.

Figure 2.9 *:* X-ray structure of HMA8 with secondary structures illustrated, in red alpha helices, in blue beta strands, in grey beta turns and unstructured regions and in dotted lines unavailable data.[51] Structure solved at 1.75Å resolution.

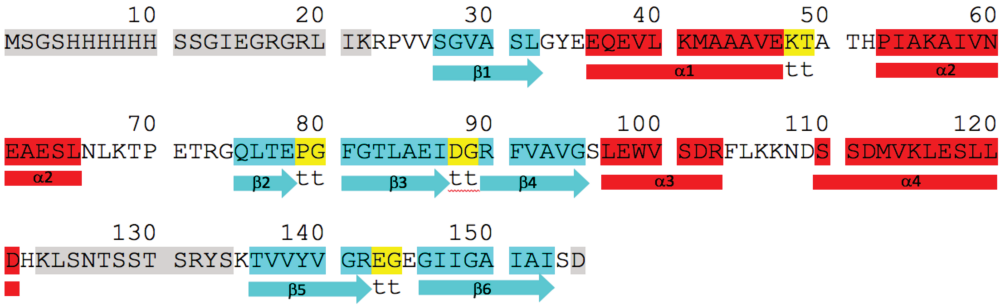
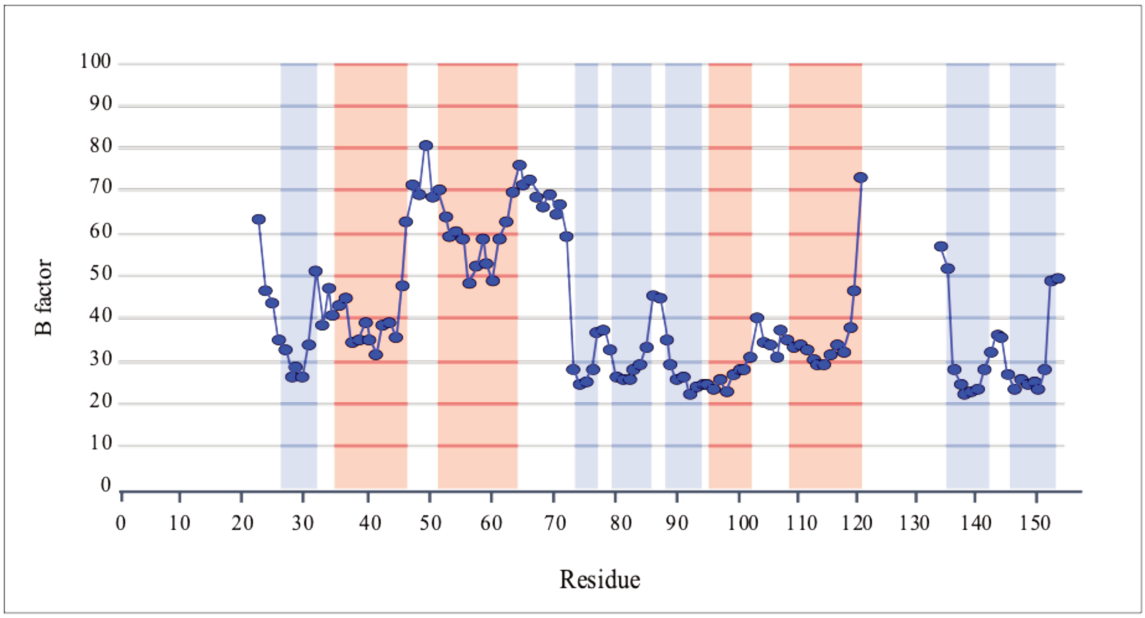


Figure 2.10 : Primary sequence of H23. The structure of H23 (HMA8: n°PDB 5LBK) is comprised of residues R23-S154 included. Highlighted in grey is the His-tag and linker added to our construction for NMR studies with a cleavage site for the Factor-Xa protease: IE/GR. The N terminal tail is not present in the reported X-Ray structure as it was cleaved (two remaining residues are present though in their construct: GA[51]). Residues K123-S134 are also highlighted in grey, representing the unavailable residues by X-ray crystallography. Highlighted in yellow are β-turns, indicated by tt (under). The elements of secondary structure found in the crystallographic structure are indicated under the sequence: red rectangles for α-helices whilst blue arrows represent the antiparallel β-strands.

In the context of this thesis, particular interest was given to this protein as it gave satisfying expression yields in acellular expression media and therefore is choice model for applications in new specific labeling strategies. However, no NMR studies have been undertaken to this date on this protein and, in a preliminary study, assignment was undertaken for ^{1}H, ^{15}N and ^{13}C backbone resonances.



Graph 2.1: Mean B factors (Debye-Waller factors) calculated for each residue of protein H23. Each dot represents the average of the three B factors corresponding to the backbone atoms (N, Cα, C). Individual B factors have been found in the pdb file 5LBK. The secondary structure elements are indicated by blue (β strands) or red (helices α) histograms. Except for helix α2, the secondary elements present in H23 are associated with low B factors (30 on average), which indicates a reduced mobility.

II.2.2. NMR assignment of H23

Assignment of the H23 protein was undertaken using a set of triple resonance experiments: HN(CO)CACB[58], HNCACB[59], HNCO[60] and HN(CA)CO[61], recorded on 700 MHz and 600 MHz Bruker spectrometers in their BEST-TROSY[62,63] version (see II.2.3 for recording conditions) at 25°C on a U-^{1}H-^{15}N-^{13}C labeled H23 sample (See II.2.3 Materials and methods for expression). Assignment of H^{N}, N, C^{α}, C^{β}, C' chemical shifts was achieved for 97% of residues. The residues M1-S11 situated in the His-tag and linker and residues D102-L105 have not be assigned (see Annex 1 for a complete table of assignments). NMRPipe[64] software was used to transform data retrieved from the spectra, whilst Sparky[65] software was employed for peak picking and data analysis. After synchronizing the ladder between spectra and peak-picking the ^{1}H-^{15}N TROSY spectrum, the C'-1, Cα-1 and Cβ-1 were numbered on the HNCO and HN(CO)CACB spectra. Aligning the HN(CA)CO and HNCACB, the C', Cα and Cβ correlations could in turn be numbered (see Figure 2.11). Once the full spectra were peak-picked, the peak lists were computed to compare Co/Cα/Cβ and Co-1/Cα-1/Cβ-1 chemical shifts, starting by characteristic amino acids such as glycine (no Cβ), threonine and serine (Cβ with a higher chemical shift than the Cα). The sequence was reconstructed in this way.

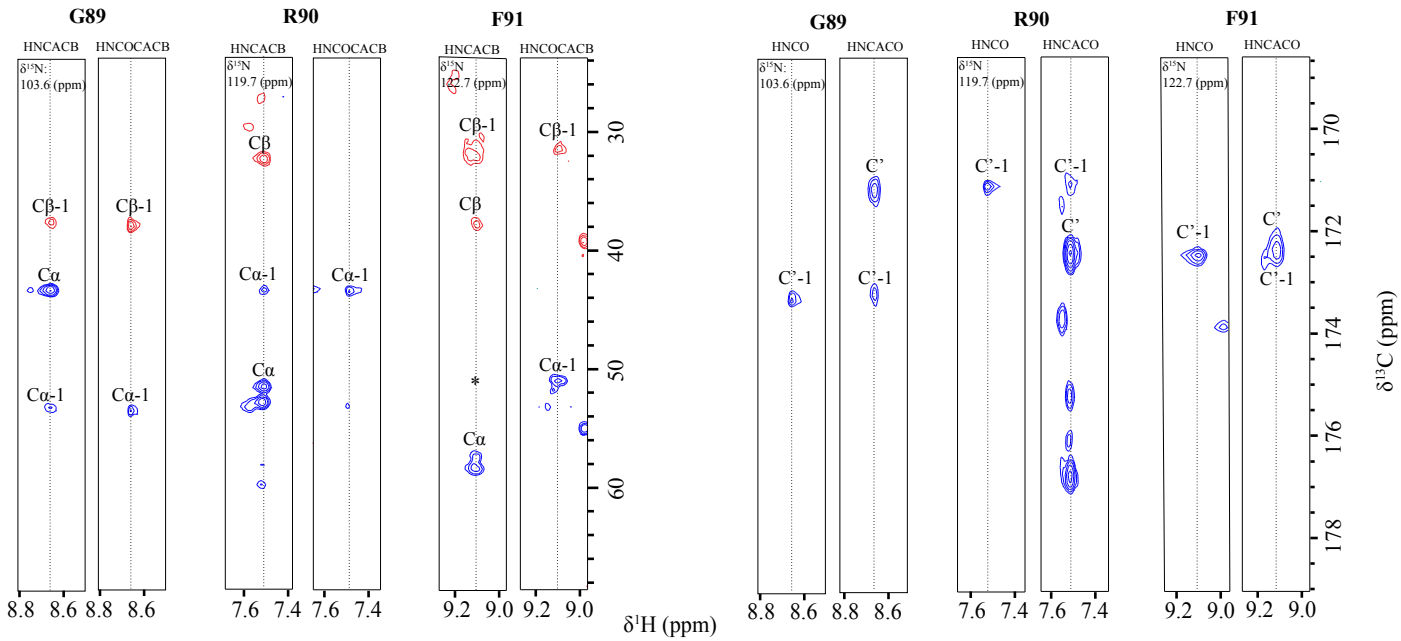
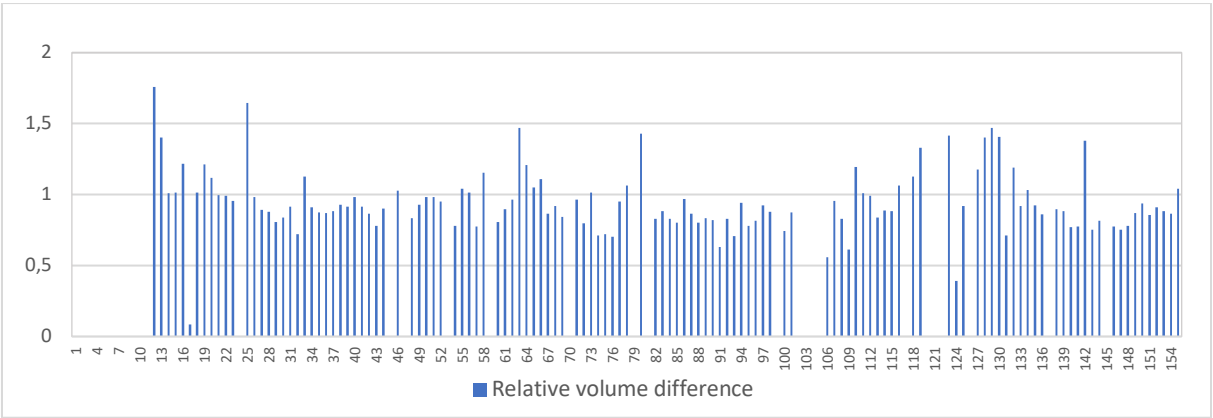
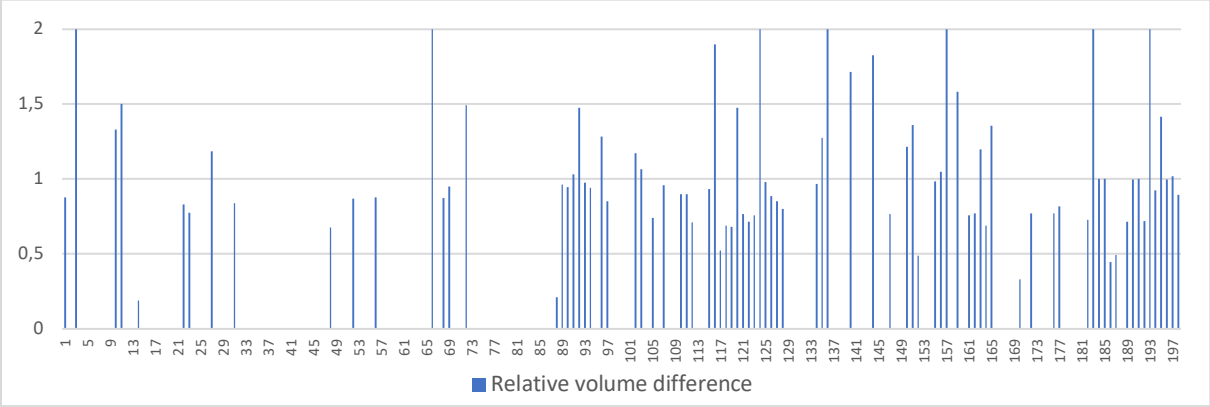


Figure 2.11: Strips taken from BEST-TROSY – HNCACB, HNCOCACB (700 MHz), HNCO and HNCACO (600 MHz) – of G89, R90 and F91 to illustrate the assignment strategy. The asterix(*) represents a correlation below the first contour level chosen for this figure. Sample conditions: 550 µM H23 U ^{1}H-^{15}N-^{13}C Phosphate buffer 20 mM, NaCl 100 mM, pH 6.5, 10% D_2O, 25°C

Assignment was rendered difficult due to a number of supplemental correlations corresponding to a different set of spin systems that also fit in the H23 protein sequence. This suggests that there may have been a certain degree of degradation of the sample. Notably, many other correlations were present in the central part of the spectrum, between 7,5 et 8,5 ppm in the ^{1}H dimension, a region associated with less structured motifs, reinforcing the idea that there could have been degradation. Moreover, an analysis of the integration of these sequences shows an increase or a decrease in intensity with time (after 4 days) (see Graphs 2.2 and 2.3) and it was later possible to get rid of these additional correlations using a fresh sample. The full assignment is illustrated on the ^{1}H-^{15}N spectrum on the next page (Figure 2.12).



Graph 2.2: Relative volume difference (%) between the first TROSY and last taken 4 days later, for each amino acid assigned to H23. The average relative difference calculated is of 0.84. Certain amino acids could not be quantified because their attribution was missing at the time of measurement.



Graph 2.3: Relative volume difference (%) between the first TROSY and last taken 4 days later, for unassigned correlations. The average relative difference calculated is of 1.04. Certain amino acids have an arbitrary volume difference of 2, this is because they were not visible on the first TROSY but appeared in the last.

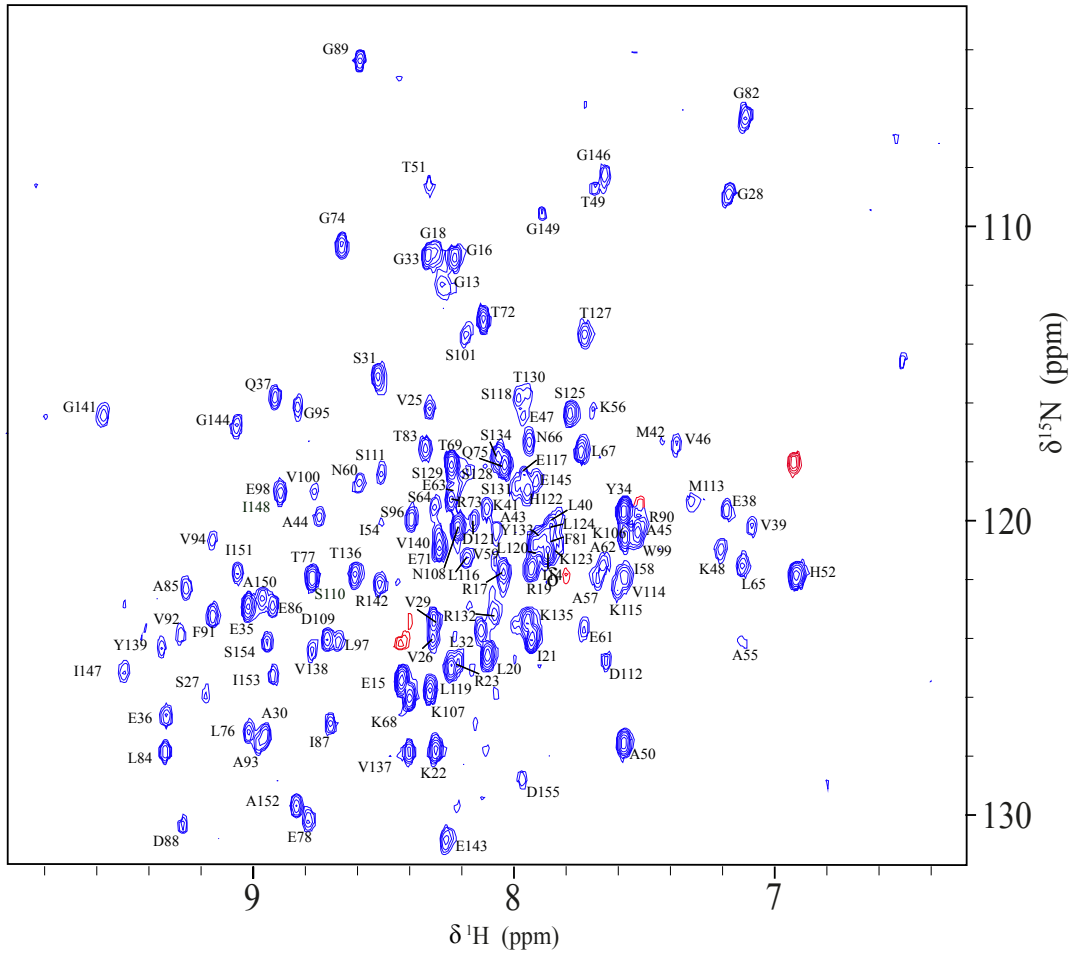


Figure 2.12: TROSY ^{15}N-^{1}H spectra of the H23 protein U-^{2}H-^{15}N-^{13}C labeled (See Chapter III for expression) at 850 MHz (300 µM, Phosphate 20 mM, pH 7, NaCl 100 mM, 25°C). Non assigned correlations are in red.

The TALOS+[66] software enabled us to determine the secondary structure elements of H23 based on the chemical shifts measured for the H^N, N, C^α, C^β and C' atoms with a very good correlation with the secondary structure elements found in the crystal structure. 28% of the predictions were ambiguous. This is often the case for loops and turns, regions of irregular

structure and high flexibility. TALOS+ relies on a data bank of 186 reference proteins, for which both the structure and the chemical shifts are known. The φ and ψ amino acid angles are predicted with the chemical shifts measured for each residue i and the residues i+1 and i-1 that surround it, as similar chemical shifts for resembling sequences are a good indication of shared structural properties. TALOS+ predicts not only the conformation but the dynamic at the residue level, two graphs displaying the RCI derived order parameter, S2. The graph displayed with the length of the bars corresponding to the probability that a residue is in a helix or β-strand.

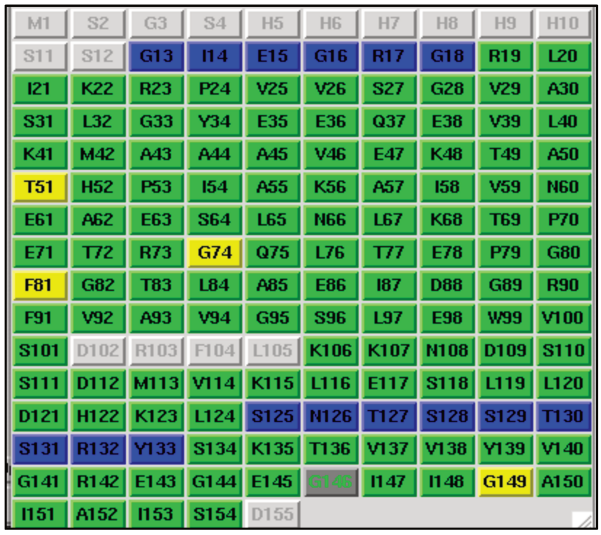


Figure 2.13: Peptide sequence of H23 with the TALOS output color code. Unassigned residues for which no chemical shift was found are in grey. In green are represented residues for which prediction is converging. In yellow, residues for which prediction is ambiguous (divergence of couples of values of φ et ψ found for at least 2 of the 10 related proteins of the data-base). In blue residues for which the prediction indicates a high dynamics (no secondary structure prediction for these regions).

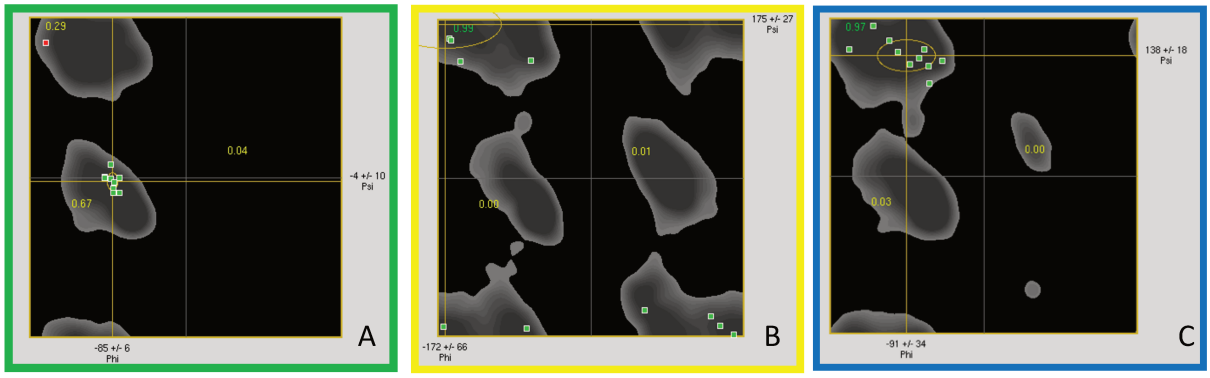


Figure 2.14: TALOS results of assigned H23 residues. A) For T49, one reference protein presents φ et ψ values (red dot) distant from the other references (green dots) for which there is convergence, enabling us to neglect this reference and validate the couple of values. B) For G149, the references are dispersed and do not enable validation of a couple of values. Prediction is therefore ambiguous. C) For S128, there is a relative convergence of values of reference proteins but the regions used for the prediction are dynamic which supposes great flexibility, or random coil of this region.

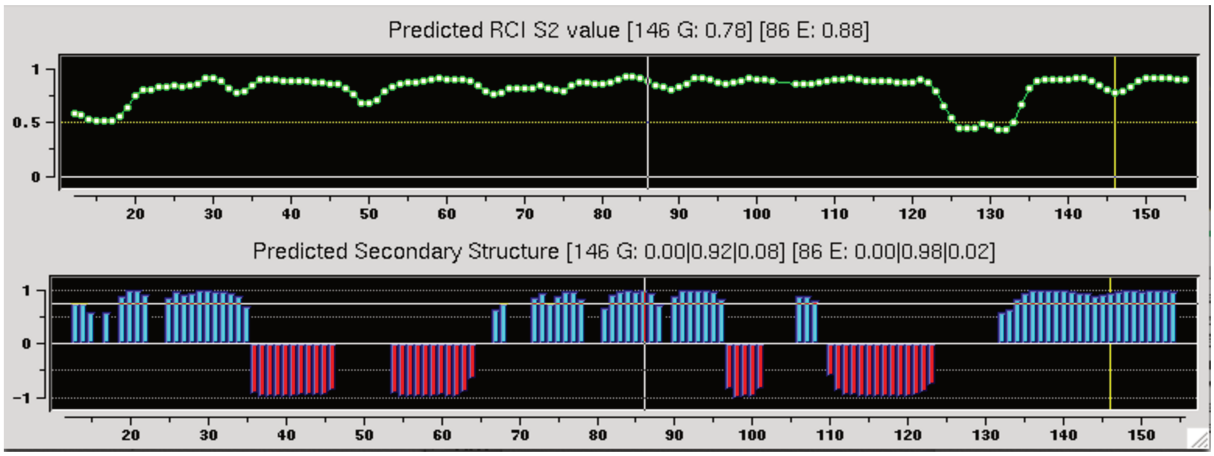


Figure 2.15: Prediction of the dynamics and secondary structures of H23 generated with the chemical shifts obtained by NMR. The upper graph displays, for each residue, variations of S^2 (order parameter) which accounts for the local dynamics. The closer the value is to 1, the more rigid is the residue, inversely, the closer the value is to 0 the more dynamic is the residue. The graph underneath predicts, for each residue, its secondary structure. The blue bars (positive ordinate index) represent a residue included in a β-strand, whereas the red bars (negative ordinate index) represent a residue in an α-helix. The spaces correspond to residues that do not present a particular secondary structure.

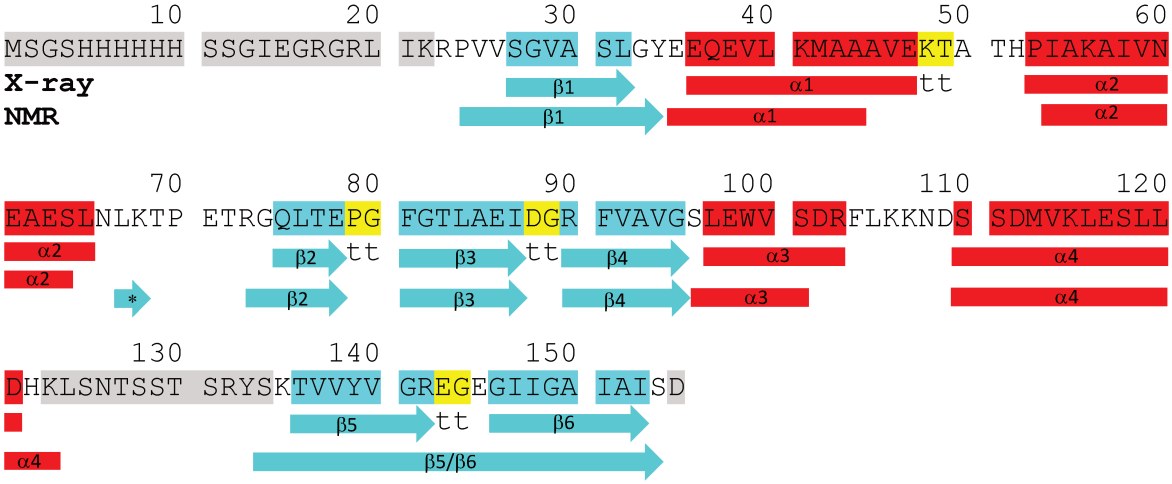
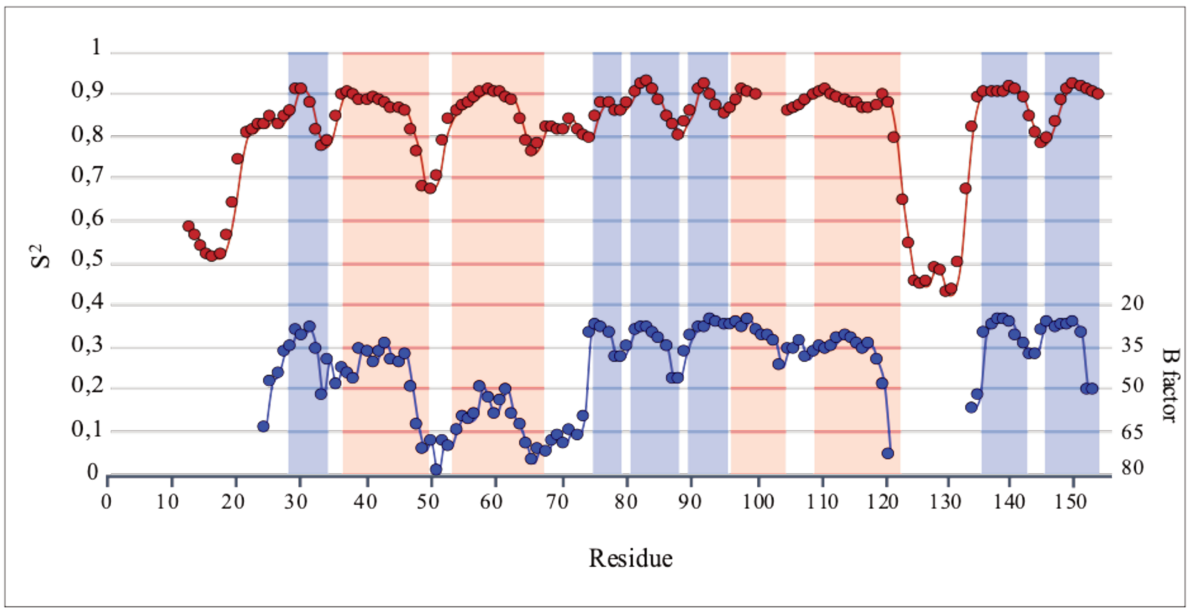


Figure 2.16 : Comparison of secondary structure elements between the NMR and X-ray data of H23. Highlighted in grey are the His-tag+linker sequences for the NMR construction and sequences not present in the crystallographic structure. Highlighted in yellow are the beta-turns found in the X-ray structure. Highlighted in blue the consensus between the two techniques for the beta strands and in red for α helices. The secondary structure elements corresponding for each technique are illustrated underneath, X-ray on top and NMR on the bottom.



Graph 2.4 : Comparison of residue dynamic of H23 between the order parameter (S^2) calculated by TALOS (red) for the set of NMR data and of the average B factor obtained by X-ray (blue). The values for the B factor have been adjusted to fit the ladder. Each dot represents the average of the three B factors corresponding to the backbone atoms (N, Cα, C). Individual B factors have been found in the pdb file 5LBK. The secondary structure elements are indicated by blue (β-strands) or red (α-helices) histograms. Except for helix 2 (B factor 55 on average), the secondary elements present in H23 are associated with low B factors (30 on average), which indicates a reduced mobility.

Most of the secondary structure elements are similar between the X-ray structure and the predicted structure using chemical shifts found by NMR (see Figure 2.15 and Graph 2.4). However, certain β-strands are considerably longer by NMR (β1, β2, β5 et β6, Figure 2.16) than by crystallography and helixes are, contrarily, slightly shorter or displaced by one or two residues (α1, α2 and α3). For the assigned region of the His-tag and linker (M13-P24), we obtained a prediction of values for φ et ψ in the extended region of the Ramachandran plot. This region is also predicted as being very dynamic, ($0,5<S^2<0,75$, see Figure 2.15) which indicates that the N-terminal region is very flexible and random coil. The software predicts two other dynamic regions: T49-K68 and D121-S134, characterized by $0,5<S^2<0,75$, whereas all the other residues display $S^2>0,75$. These two regions include respectively helix α2 and an unstructured region which was not resolved in the crystallographic structure. The region from P53 to L65 corresponding to helix α2 reveal a slightly dynamic region, especially compared to other secondary structure elements of the protein (see Graph 2.3). This could suggest a region transitionally structured and very dynamic in solution. In the solid state, the packing of the crystal could also promote the helix α2 formation through intermolecular contacts. Helix α2 is indeed situated at the periphery of the X-ray structure and exposed to the solvent (see Figure

2.9). The very dynamic parts of the protein have similar boundaries between the two techniques (T49-K68 et L120-K135) and residues not assigned by NMR or not visible on the X-ray structure are included in these regions. In both techniques, we can observe increased flexibility between each secondary structure elements, which are on the contrary much more rigid ($0,85<S^2<0,91$ or $20<B<30$). Overall, the dynamic profiles obtained for each of the two techniques are very comparable, confirming similar global topology of the protein in the liquid or in the solid state. This comparison is valuable for the validation of our chemical shift assignments. The study of the graphs revealed that the helix alpha 2 in X-ray (P53 – P70) is rigid (S2 is much higher, from ($0,75<S^2$), display the same general dynamic than previously. These similarities of secondary structure and dynamic validates the assignment although assignment is still missing for 4 amino acids: D102, R103, F104 and L105, present between alpha helix 3 and 4, so probably also in a very flexible region. As mentioned previously, in the X-ray data, the electron density of R103 corresponds to the superposition of two conformations and interestingly, this residue as well as those surrounding it, were not assigned by NMR, maybe because of an intermediary exchange.

II.3. Materials and Methods

II.3.1. *E.coli* expression and purification conditions for Pin1

4000 mL of LB medium pH 7.2 (10 g casein peptone/L, 5 g yeast extract/L, 10g NaCl/L) was prepared in H_2O, in 500 mL batches and autoclaved. Two 50 mL cultures were inoculated with 1 mL of glycerol stock BL21 (DE3) bacteria, containing the pet15B corresponding plasmids for each sequence, 1mg/mL of ampicillin and 0,3 mg/mL of chloramphenicol. These were left overnight, shaking at 220 rpm, 30°C. The next morning, when DO_{600} reached between 2 and 3, the 500 mL of LB medium were inoculated with the precultures to reach an OD_{600} of 0.05 with 1 mg/mL of ampicillin and 0.3 mg/mL of chloramphenicol. The cultures are left shaking at 200 rpm for three to four hours at 37°C, until the OD_{600} reaches between 0.6 and 0.8. At this stage, if no labeling is desired, cells can be induced with IPTG at 1 mM concentration. If labeling is desired, cells are centrifuged 15 minutes at 3000g 4°C, the supernatant is discarded and cell mass is resuspended in M9 medium.

M9 medium is prepared in 900 mL of autoclaved water with a desired final concentration of 1 L at 50mM of Na_2HPO_4, 25 mM KH_2PO_4, pH 7.2, 10 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.04 mM Thiamine, 1mg/mL ampicillin, 0.3 mg/mL chloramphenicol

and 1,7 g of Yeast Nitrogen Base. YNB needs to be resuspended in water and sonicated before using a 0.22 μm filter to filter-sterilize the solution when adding to the medium. The medium is separated into 3 Erlenmeyers of 333 mL each. Cells are left shaking at 220 rpm 37°C for at 1 hour. At this point, 1% glucose (with desired label), 0,1% ammonium chloride (with desired label) and 1 mM IPTG are added. Cells are returned to the incubator for overnight induction shaking at 220 rpm 25°C.

The next day, cells are centrifuged 20 minutes at 4000g 4°C. Supernatant is discarded, and cells are resuspended in 80 mL of lysis buffer. Lysis buffer is constituted of 30 mM Phosphate buffer pH 7.8, 50 mM NaCl, 30 mM Imidazole, 1mM DTT, 1% Tween, 1 mM PMSF, 1mg/L Pepstatin A and 1 mg/L Leupeptin, left on ice for 15 minutes. Cells are sonicated 20 times, 10 times for 15 seconds and 10 times for 20 seconds, 40 amp with 1 minute on ice between each sonication step. The lysate is then centrifuged 30 minutes at 25000 g 4°C and the supernatant used for purification by His-trap column. The supernatant is filtered in a 0.45 μm filter .The column was equilibrated with phosphate buffer 30 mM pH 7.8, 50 mM NaCl, 30 mM imidazole, 1 mM DTT and eluted with phosphate buffer 30 mM pH 7.8, 50 mM NaCl, 500 mM imidazole, 1mM DTT with a 0-80% gradient in 40 minutes at 1 mL per minute. By monitoring absorbance at 280nm, the protein eluted can be recuperated in fractions. Yields range from 10 to 70 mg/L of culture at this point.

The elution fractions are then diluted by half with Phosphate buffer 30 mM pH 7.8, 50 mM NaCl, 30 mM imidazole, 1mM DTT and incubated with 0.5mg of PreScission enzyme and left overnight at 4°C. EDTA at 1 mM can be added if there is risk of protease activity.

If one is purifying the WW domain, the next two steps can be skipped. The next day, the elution fractions are passed through the His-Trap column again with phosphate buffer 30 mM pH 7,8, 50 mM NaCl, 30 mM imidazole, 1 mM DTT and recuperated directly. Non-specific proteins stay stuck to the column.

The elution fraction is then passed in a size exclusion chromatography column with Phosphate buffer 30 mM pH 7.8, 50 mM NaCl, 1mM DTT and recuperated. The eluates are pooled together and concentrated in an Amicon ® pressure cell concentrator with a 3.5 kDa membrane until the desired concentration.

If one if purifying the WW domain, the elution fractions after incubation with the PreScission enzyme can then purified by high-performance liquid chromatography (HPLC). A C4-300 nm 10-250 mm column was used at a flow rate of 6mL per minute and a acetonitrile gradient of 10 to 100% in 30 minutes, with 0,1% TFA Fractions containing the WW protein elute at 8 minutes (8.5 minutes if the His-tag has not been cleaved) and recuperated. They are shock frozen and lyophilized. Dry powders are weighed the next day and resuspended in D_2O, shock frozen with liquid nitrogen and lyophilized again before keeping at -80°C.

II.3.2. Recording conditions for Pin1

On the Bruker Z gradient equipped Cryoprobe 500 MHz spectrometer were carried out the ^{1}H-^{15}N HSQC for Pin1, PPIase and WW identification:

- Samples of Pin1, 150 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C Tris buffer 30 mM, pH 7, 50 mM NaCl, 1 mM DTT, 1% DSS, 5 mm tube, 500 MHz (Bruker Cryoprobe)

- Sequence parameters: TD F2 2048, F1 128, NS 4, D1 1 sec, D16 0.0002 sec, AQ 0.128 sec 0.0647 sec, SW 16 ppm, 19.5 ppm, O1P : 4.697 ppm, 116.75 ppm, RG 128

- Processing parameters (Topspin): SI 4096 512, sine bell apodisation in F2 and F1, SSB 2.2 2, TDeff 2048

- Samples of PPIase, 230 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C phosphate buffer 30 mM, pH 7, 50 mM KCl, 2 mM DTT, 1 mM EDTA, 0.05% NaN_3, 1% DSS, 5 mm tube, 500 MHz (Bruker Cryoprobe)

- Sequence parameters : TD F2 2048, F1 128, NS 4, D1 1 sec, D16 0.0002 sec, AQ 0.128 sec 0.0647 sec, SW 16 ppm, 19.5 ppm, O1P : 4.697 ppm, 116.75 ppm, RG 128

- Processing parameters (Topspin): SI 4096 512, sine bell apodisation in F2 and F1, SSB 2.2 2, TDeff 2048

- Samples of WW, 150 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C Tris buffer 30 mM, pH 6.85, 50 mM NaCl, 1 mM DTT, 1% DSS, 5 mm tube, 500 MHz (Bruker Cryoprobe)

- Sequence parameters : TD F2 2048, F1 128, NS 4, D1 1 sec, D16 0.0002 sec, AQ 0.128 sec 0.0647 sec, SW 16 ppm, 19.5 ppm, O1P : 4.696 ppm, 116.75 ppm, RG 256

- Processing parameters (Topspin): SI 4096 512, sine bell apodisation in F2 and F1, SSB 2.2 2, TDeff 2048

II.3.3. H23 U-^{1}H-^{15}N-^{13}C labeled expression and purification for assignment recordings

H23 was expressed using Continuous Exchange Cell-free (CECF) expression using a dialysis membrane system. The preparation of the S30 extract is detailed in Chapter III.3. Material and Methods. A pIVEX vector was used for coding H23 protein with a His-tag sequence at N-terminus to facilitate purification, detailed in Chapter III. In the following protocol, all the mixtures were vortexed, except at point 5.

1. Hydrolyzed U-^{1}H-^{15}N-^{13}C Celtone ® is solubilized at 50mg/mL in H_2O. A solution of 23.8 mM Tryptophan is prepared in HCl 1 M, 1 M KOH, 2 M HEPES in H_2O.
2. The dialysis membrane is washed with water and kept in water during the CECF reaction preparation
3. The 10X reaction mix containing HEPES/KOH (55 mM, pH 7.5), DTT (3.4 mM), ATP (1.2 mM), 3', 5'-cyclic AMP (0.64 mM), 0.8 mM of each CTP, GTP and UTP ribonucleotides, folinic acid (68 µM), ammonium acetate (27.5 mM) and spermidine (2 mM) is prepared.
4. The feeding mix with 9,45 mL of the 10X reaction mix, creatine phosphate at 80 mM, Potassium Glutamate at 208 mM, Tryptophan at 0.5 mM, Celtone mix at 3 mg/mL and magnesium acetate at 14.4 mM final is prepared, adding water to 90 mL.
5. The reaction mix is prepared as described for the feeding mix (a factor of ten smaller volume), with addition of 250 µg/mL of creatine kinase, 175 µg/mL of total tRNA E. coli MRE600, 50 µg/mL of T7 RNApol, 40% of the final volume of S30 extract, and 16 µg/mL of the target protein vector. The reaction is not vortexed here.
6. The reaction mix (9 mL) is loaded in the dialysis membrane and the feeding mix (90 mL) in a cylinder and incubated with stirring at a 30°C overnight (16 hours)
7. Next morning, the supernatant is recovered by centrifugation 20 minutes at 10,000 g. It is diluted by a factor 4 to decrease sample ionic strength before loading it to the His-Trap Nickel affinity column.

Amino acid mix	Volume (μL)
Water soluble AA: A, G, H, K, P, R, S, V, T 50 mM concentration for each AA	247.5
In 1 M HCl: C, D, E, L, M, N, Q, W, Y 50 mM concentration for each AA	247.5
In 1 M NaOH: I, F 50 mM concentration for each AA	247.5
RNase-free H_2O	82.5
Total	825.0

10× reaction mix	(μL)
100 mM rCTP	96.0
100 mM rGTP	96.0
100 mM rUTP	96.0
2.0 M HEPES–KOH pH 7.5	330.0
100 mM ATP	144.0
10 mM folinic acid	81.6
100 mM cyclic AMP	76.8
1 M DTT	40.8
1 M spermidine	24.0
9.2 M NH_4OAc	35.9
RNase-free H_2O	178.9
Total	1200

Reaction Mix (RM) and Feeding Mix (FM)	RM (μL)	FM (μL)
10× reaction mix	100.0	1000.0
1 M creatine phosphate	80.0	800.0
Amino acid mix	66.7	666.7
4 M potassium glutamate	52.0	520.0
1.07 M $mg(OAc)_2$	7.9	130.8
17.5 mg/mL MRE 600 tRNA	10.0	0
10 mg/mL creatine kinase	25.0	0
T7 RNA polymerase (1/100e)	10.0	0
Adjust pH of FM to 7.5 with KOH		
S30 extract	400.0	0
Target DNA (16 μg/mL)	16.0	0
RNase-free H_2O	232.5	6882.5
Volume (μL) of mix	1000	10,000

Table 2.1 : Table of cell-free reaction mix and feeding mix preparation for CECF synthesis of proteins

Purification is undertaken by His-trap column. The supernatant was filtered in a 0.45 µm filter. The column was equilibrated with phosphate buffer 20 mM pH 6.5, 100 mM NaCl, 30 mM imidazole, and eluted with phosphate buffer 20 mM pH 6.5, 100 mM NaCl, 500 mM imidazole. By monitoring absorbance at 280 nm, the protein eluted is recuperated in fractions. Yields measured by Nanodrop were of 10mg total after concentrating elution fractions 2,3 and 4 in a Centricon 3 kDa membrane concentrator.

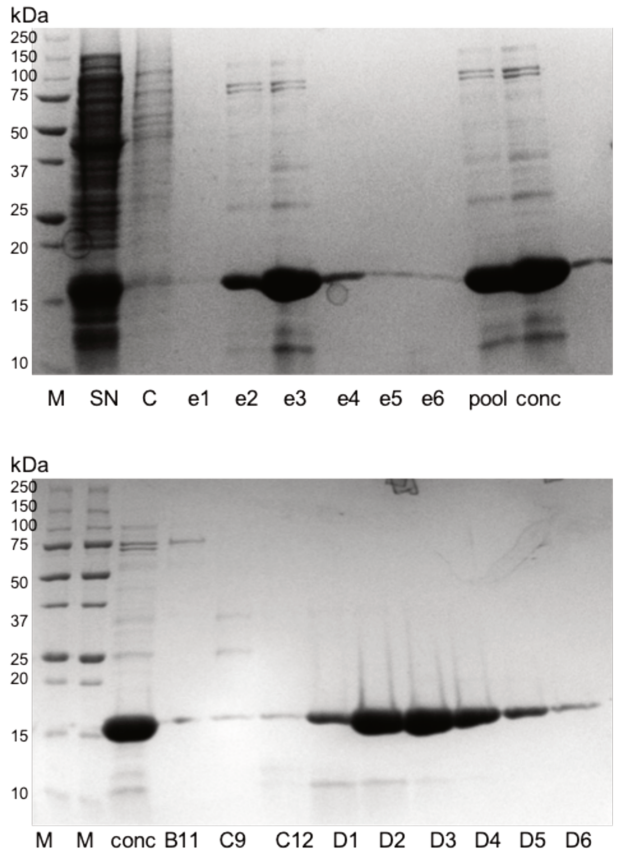


Figure 2.17 : Photographs of the polyacrylamide gels made of the different H23 purification steps. The top panel are the fractions from the passage on the His-trap nickel affinity column, the bottom panel of the exclusion chromatography purification. M : Molecular ladder, SN : supernatant, C : pellet, e1 → e6 : elution fractions 1 to 6 of the His-trap affinity column purification, pool : pool of e2, e3 and e4, conc : concentrated pool. B11, C9, C12: elution fractions from exclusion chromatography, D1→ D6 : elution fractions of H23 from exclusion chromatography.

After this step, the pooled and concentrated elution fractions were passed through a size exclusion chromatography column in phosphate buffer 20 mM pH 6.5, 100 mM NaCl. Elution fractions containing H23 protein were recuperated, pooled and concentrated in a Centricon 3 kDa membrane concentrator for a final yield of 9 mg.

II.3.4. Assignment spectra recording conditions for H23

On the Bruker Z gradient equipped 600 MHz spectrometer were carried out the HNCO and HNCACO assignment spectra for H23:

- Samples of H23, 550 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C phosphate buffer 20 mM, pH 6.85, 100 mM NaCl, 5 mm tube, 600 MHz (Bruker) HNCO

- Sequence parameters: TD F3 1022 F2 128, F1 128, NS 8, D1 0.2 sec, AQ 0.071 sec 0.029 sec 0.028 sec, SW 12 ppm, 36 ppm 15 ppm, O1P : 4.7 ppm, 118 ppm 173 ppm, RG 203

- Processing parameters (NMRPipe): SI 1024 128 128, sine bell apodisation in F3 F2 and F1, offset 0.45 0.5 0.5, end 0.98 1.0 1.0, exponent 2 2 2, Point 1 scale 0.5 0.5 0.5, TDeff 0, linear prediction in F2

- Samples of H23, 550 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C phosphate buffer 20 mM, pH 6.85, 100 mM NaCl, 5 mm tube, 600 MHz (Bruker) HN(CA)CO

- Sequence parameters TD F3 1022 F2 112, F1 112, NS 16, D1 0.2 sec, AQ 0.071 sec 0.029 sec 0.028 sec, SW 12 ppm, 34 ppm 20 ppm, O1P : 4.7 ppm, 118 ppm 173 ppm, RG 203

- Processing parameters (NMRPipe): SI 1024 128 128, sine bell apodisation in F3 F2 and F1, offset 0.45 0.5 0.5, end 0.98 1.0 1.0, exponent 2 2 2, Point 1 scale 0.5 0.5 0.5, TDeff 0, linear prediction in F2

On the Bruker Z gradient equipped 700 MHz spectrometer were carried out the HN(CO)CACB and HNCACB assignment spectra for H23:

- Samples of H23, 550 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C phosphate buffer 20 mM, pH 6.85, 100 mM NaCl, 5 mm tube, 700 MHz (Bruker) HN(CO)CACB

- Sequence parameters : TD F3 1192 F2 148, F1 252, NS 12, D1 0.26 sec D16 0.0002 sec, AQ 0.071 sec 0.029 sec 0.01 sec, SW 12 ppm, 36 ppm 70 ppm, O1P : 4.7 ppm, 118 ppm 39 ppm, RG 203

- Processing parameters (NMRPipe): SI 2048 256 256, sine bell apodisation F3 F2 and F1, offset 0.45 0.5 0.5, end 0.98 1.0 1.0, exponent 2 2 2, Point 1 scale 0.5 0.5 0.5, TDeff 0, linear prediction in F2

- Samples of H23, 550 µM uniformly labeled ^{15}N-^{13}C in 10% D_2O 90% H_2O, 25°C phosphate buffer 20 mM, pH 6.85, 100 mM NaCl, 5 mm tube, 700 MHz (Bruker) HNCACB

- Sequence parameters : TD F3 1194 F2 128, F1 256, NS 16, D1 0.26 sec D16 0.0002 sec, AQ 0.071 sec 0.029 sec 0.01 sec, SW 12 ppm, 36 ppm 70 ppm, O1P : 4.7 ppm, 118 ppm 39 ppm, RG 180.78

- Processing parameters (NMRPipe): SI 2048 128 512, sine bell apodisation F3 F2 and F1, offset 0.45 0.5 0.5, end 0.98 1.0 1.0, exponent 2 2 2, Point 1 scale 0.5 0.5 0.5, TDeff 0, linear prediction in F2

1. Nestler, E. & Greengard, P. in *Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th edition* (Lippincott-Raven, 1999).
2. Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front. Genet.* **5,** (2014).
3. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298,** 1912–1934 (2002).
4. Harper, J. W. & Adams, P. D. Cyclin-Dependent Kinases. *Chem. Rev.* **101,** 2511–2526 (2001).
5. Lim, S. & Kaldis, P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development* **140,** 3079–3093 (2013).
6. Garrington, T. P. & Johnson, G. L. Organization and regulation of mitogen-activated protein kinase signaling pathways. *Current Opinion in Cell Biology* **11,** 211–218 (1999).
7. Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380,** 544–547 (1996).
8. Chen, Y., Wu, Y., Yang, H., Li, X., Jie, M., Hu, C., Wu, Y., Yang, S. & Yang, Y. Prolyl isomerase Pin1: a promoter of cancer and a target for therapy. *Cell Death Dis* **9,** 883 (2018).
9. El Boustani, M., De Stefano, L., Caligiuri, I., Mouawad, N., Granchi, C., Canzonieri, V., Tuccinardi, T., Giordano, A. & Rizzolio, F. A Guide to PIN1 Function and Mutations Across Cancers. *Front. Pharmacol.* **9,** 1477 (2019).
10. Hamdane, M., Smet, C., Sambo, A.-V., Leroy, A., Wieruszeski, J.-M., Delobel, P., Maurage, C.-A., Ghestem, A., Wintjens, R., Begard, S., Sergeant, N., Delacourte, A., Horvath, D., Landrieu, I., Lippens, G. & Buee, L. Pin1 : A Therapeutic Target in Alzheimer Neurodegeneration. *JMN* **19,** 275–288 (2002).
11. Rudrabhatla, P. & Pant, H. C. Phosphorylation-Specific Peptidyl-Prolyl Isomerization of Neuronal Cytoskeletal Proteins by Pin1: Implications for Therapeutics in Neurodegeneration. *JAD* **19,** 389–403 (2010).
12. Lee, T. H., Pastorino, L. & Lu, K. P. Peptidyl-prolyl cis–trans isomerase Pin1 in ageing, cancer and Alzheimer disease. *Expert Rev. Mol. Med.* **13,** e21 (2011).
13. Zhang, J. & Germann, M. W. Characterization of secondary amide peptide bonds isomerization: Thermodynamics and kinetics from 2D NMR spectroscopy. *Biopolymers* n/a-n/a (2011). doi:10.1002/bip.21642
14. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role of cis peptide bonds in protein structures. *Journal of Molecular Biology* **214,** 253–260 (1990).
15. Weiss, M. S., Jabs, A. & Hilgenfeld, R. Peptide bonds revisited. *Nat Struct Mol Biol* **5,** 676–676 (1998).
16. Grathwohl, C. & Wüthrich, K. The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides: X-PRO PEPTIDE BOND. *Biopolymers* **15,** 2025–2041 (1976).
17. Grathwohl, C. & Wüthrich, K. NMR studies of the rates of proline *cis* - *trans* isomerization in oligopeptides: rates of *cis* - *trans* isomerization. *Biopolymers* **20,** 2623–2633 (1981).
18. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3,** 619–629 (2007).
19. Smet, C., Wieruszeski, J.-M., Buée, L., Landrieu, I. & Lippens, G. Regulation of Pin1 peptidyl-prolyl *cis/trans* isomerase activity by its WW binding module on a multi-phosphorylated peptide of Tau protein. *FEBS Letters* **579,** 4159–4164 (2005).
20. Zhou, X. Z., Lu, P.-J., Wulf, G. & Lu, K. P. Phosphorylation-dependent prolyl isomerization: a novel signaling regulatory mechanism. *Cellular and Molecular Life Sciences (CMLS)* **56,** 788–806 (1999).
21. Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89,** 875–886 (1997).
22. Namanja, A. T., Wang, X. J., Xu, B., Mercedes-Camacho, A. Y., Wilson, K. A., Etzkorn, F. A. & Peng, J. W. Stereospecific gating of functional motions in Pin1. *Proceedings of the National Academy of Sciences* **108,** 12289–12294 (2011).
23. Verdecia, M. A., Bowman, M. E., Ping, K., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *nature structural biology* **7,** 5 (2000).
24. Jacobs, D. M., Saxena, K., Vogtherr, M., Bernadó, P., Pons, M. & Fiebig, K. M. Peptide Binding Induces Large Scale Changes in Inter-domain Mobility in Human Pin1. *Journal of Biological Chemistry* **278,** 26174–26182 (2003).
25. Bayer, E., Goettsch, S., Mueller, J. W., Griewel, B., Guiberman, E., Mayr, L. M. & Bayer, P. Structural Analysis of the Mitotic Regulator *h* Pin1 in Solution: INSIGHTS INTO DOMAIN ARCHITECTURE AND SUBSTRATE BINDING. *Journal of Biological Chemistry* **278,** 26183–26193 (2003).
26. Lee, Y. M. & Liou, Y.-C. Gears-In-Motion: The Interplay of WW and PPIase Domains in Pin1. *Front. Oncol.* **8,** 469 (2018).
27. Sudol, M. Structure and function of the WW domain. *Progress in Biophysics and Molecular Biology* **65,** 113–132 (1996).
28. Sudol, M. & Hunter, T. NeW Wrinkles for an Old Domain. *Cell* **103,** 1001–1004 (2000).
29. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. &

Kelly, J. W. Structure-function-folding relationship in a WW domain. *Proceedings of the National Academy of Sciences* **103,** 10648–10653 (2006).
30. Jäger, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. The folding mechanism of a β-sheet: the WW domain. *Journal of Molecular Biology* **311,** 373–393 (2001).
31. Lu, P. Function of WW Domains as Phosphoserine- or Phosphothreonine-Binding Modules. *Science* **283,** 1325–1328 (1999).
32. Lu, P.-J., Zhou, X. Z., Liou, Y.-C., Noel, J. P. & Lu, K. P. Critical Role of WW Domain Phosphorylation in Regulating Phosphoserine Binding Activity and Pin1 Function. *J. Biol. Chem.* **277,** 2381–2384 (2002).
33. Behrsin, C. D., Bailey, M. L., Bateman, K. S., Hamilton, K. S., Wahl, L. M., Brandl, C. J., Shilton, B. H. & Litchfield, D. W. Functionally Important Residues in the Peptidyl-prolyl Isomerase Pin1 Revealed by Unigenic Evolution. *Journal of Molecular Biology* **365,** 1143–1162 (2007).
34. Innes, B. T., Bailey, M. L., Brandl, C. J., Shilton, B. H. & Litchfield, D. W. Non-catalytic participation of the Pin1 peptidyl-prolyl isomerase domain in target binding. *Front. Physio.* **4,** (2013).
35. Wilson, K. A., Bouchard, J. J. & Peng, J. W. Interdomain Interactions Support Interdomain Communication in Human Pin1. *Biochemistry* **52,** 6968–6981 (2013).
36. Namanja, A. T., Peng, T., Zintsmaster, J. S., Elson, A. C., Shakour, M. G. & Peng, J. W. Substrate Recognition Reduces Side-Chain Flexibility for Conserved Hydrophobic Residues in Human Pin1. *Structure* **15,** 313–327 (2007).
37. Lu, K. P., Liou, Y.-C. & Zhou, X. Z. Pinning down proline-directed phosphorylation signaling. *Trends in Cell Biology* **12,** 164–172 (2002).
38. Wintjens, R., Wieruszeski, J.-M., Drobecq, H., Rousselot-Pailley, P., Buée, L., Lippens, G. & Landrieu, I. ^{1}H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* **276,** 25150–25156 (2001).
39. Kops, O., Zhou, X. Z. & Lu, K. P. Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. *FEBS Letters* **513,** 305–311 (2002).
40. Zhou, X. Z., Kops, O., Werner, A., Lu, P.-J., Shen, M., Stoller, G., Küllertz, G., Stark, M., Fischer, G. & Lu, K. P. Pin1-Dependent Prolyl Isomerization Regulates Dephosphorylation of Cdc25C and Tau Proteins. *Molecular Cell* **6,** 873–883 (2000).
41. Lu, K. P. & Zhou, X. Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* **8,** 904–916 (2007).
42. Zhou, X. Z., Kops, O., Werner, A., Lu, P.-J., Shen, M. & Stoller, G. Pin1-Dependent Prolyl Isomerization Regulates Dephosphorylation of Cdc25C and Tau Proteins. *Molecular Cell* 11
43. Shen, M., Stukenberg, P. T., Kirschner, M. W. & Lu, K. P. The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes & Development* **12,** 706–720 (1998).
44. Stukenberg, P. T. & Kirschner, M. W. Pin1 Acts Catalytically to Promote a Conformational Change in Cdc25. *Molecular Cell* **7,** 1071–1083 (2001).
45. Liou, Y.-C., Ryo, A., Huang, H.-K., Lu, P.-J., Bronson, R., Fujimori, F., Uchida, T., Hunter, T. & Lu, K. P. Loss of Pin1 function in the mouse causes phenotypes resembling cyclin D1-null phenotypes. *Proceedings of the National Academy of Sciences* **99,** 1335–1340 (2002).
46. Ryo, A., Suizu, F., Yoshida, Y., Perrem, K., Liou, Y.-C., Wulf, G., Rottapel, R., Yamaoka, S. & Lu, K. P. Regulation of NF-κB Signaling by Pin1-Dependent Prolyl Isomerization and Ubiquitin-Mediated Proteolysis of p65/RelA. *Molecular Cell* **12,** 1413–1426 (2003).
47. Ryo, A., Suizu, F., Yoshida, Y., Perrem, K., Liou, Y.-C., Wulf, G., Rottapel, R., Yamaoka, S. & Lu, K. P. Regulation of NF-□B Signaling by Pin1-Dependent Prolyl Isomerization and Ubiquitin-Mediated Proteolysis of p65/RelA. *Molecular Cell* 14
48. Kimura, T., Tsutsumi, K., Taoka, M., Saito, T., Masuda-Suzukake, M., Ishiguro, K., Plattner, F., Uchida, T., Isobe, T., Hasegawa, M. & Hisanaga, S. Isomerase Pin1 Stimulates Dephosphorylation of Tau Protein at Cyclin-dependent Kinase (Cdk5)-dependent Alzheimer Phosphorylation Sites. *J. Biol. Chem.* **288,** 7968–7977 (2013).
49. Nakamura, K., Greenwood, A., Binder, L., Bigio, E. H., Denial, S., Nicholson, L., Zhou, X. Z. & Lu, K. P. Proline Isomer-Specific Antibodies Reveal the Early Pathogenic Tau Conformation in Alzheimer's Disease. *Cell* **149,** 232–244 (2012).
50. Smet, C., Sambo, A.-V., Wieruszeski, J.-M., Leroy, A., Landrieu, I., Buée, L. & Lippens, G. The Peptidyl Prolyl *cis* / *trans* -Isomerase Pin1 Recognizes the Phospho-Thr212-Pro213 Site on Tau †. *Biochemistry* **43,** 2032–2040 (2004).
51. Mayerhofer, H., Sautron, E., Rolland, N., Catty, P., Seigneurin-Berny, D., Pebay-Peyroula, E. & Ravaud, S. Structural Insights into the Nucleotide-Binding Domains of the P1B-type ATPases HMA6 and HMA8 from Arabidopsis thaliana. *PLOS ONE* **11,** e0165666 (2016).
52. Sautron, E., Mayerhofer, H., Giustini, C., Pro, D., Crouzy, S., Ravaud, S., Pebay-Peyroula, E., Rolland, N., Catty, P. & Seigneurin-Berny, D. HMA6 and HMA8 are two chloroplast Cu+-ATPases with different

enzymatic properties. *Bioscience Reports* **35,** e00201 (2015).
53. Tapken, W., Ravet, K. & Pilon, M. Plastocyanin Controls the Stabilization of the Thylakoid Cu-transporting P-type ATPase PAA2/HMA8 in Response to Low Copper in *Arabidopsis*. *J. Biol. Chem.* **287,** 18544–18550 (2012).
54. Abdel-Ghany, S. E., Müller-Moulé, P., Niyogi, K. K., Pilon, M. & Shikanai, T. Two P-Type ATPases Are Required for Copper Delivery in *Arabidopsis thaliana* Chloroplasts. *Plant Cell* **17,** 1233–1251 (2005).
55. Gourdon, P., Liu, X.-Y., Skjørringe, T., Morth, J. P., Møller, L. B., Pedersen, B. P. & Nissen, P. Crystal structure of a copper-transporting PIB-type ATPase. *Nature* **475,** 59–64 (2011).
56. Tsuda, T. & Toyoshima, C. Nucleotide recognition by CopA, a Cu+-transporting P-type ATPase. *EMBO J* **28,** 1782–1791 (2009).
57. Wang, K., Sitsel, O., Meloni, G., Autzen, H. E., Andersson, M., Klymchuk, T., Nielsen, A. M., Rees, D. C., Nissen, P. & Gourdon, P. Structure and mechanism of Zn2+-transporting P-type ATPases. *Nature* **514,** 518–522 (2014).
58. Grzesiekt, S. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *Journal of American Chemical Society* **114,** 6291–6293 (1992).
59. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99,** 201–207 (1992).
60. Clubb, R. T., Thanabal, V. & Wagner, G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins. *Journal of Magnetic Resonance (1969)* **97,** 213–217 (1992).
61. Clubb, R. T. A Constant-Time Three-Dimensional Triple Resonance Pulse Scheme to Correlate Intra-residue HN, 15N, and 13C'Chemical Shifts in 15N- 13C-Labeled Proteins. 5
62. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *Journal of American Chemical Society* **128,** 9042–9043 (2006).
63. Lescop, E., Rasia, R. & Brutscher, B. Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment. *Journal of the American Chemical Society* **130,** 5014–5015 (2008).
64. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6,** 277–293 (1995).
65. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327 (2015).
66. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44,** 213–223 (2009).
67. Peter M. Steinert, James W. Mack, Bernhard P. Korge, Song-Qing Gan, Susan R. Haynes, Alasdair C. Steven, Glycine loops in proteins: their occurence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins,International Journal of Biological Macromolecules,Volume 13, Issue 3, (1991)130-139,

Chapter III : Optimization and application of the cell free expression for the perdeuteration of proteins

This part of my work was undertaken at the cell-free platform of the Institut de Biologie Structurale de Grenoble. Optimization and application of the cell-free expression protocol were undertaken in collaboration with Lionel Imbert and Jérôme Boisbouvier who have long experience in cell-free protein expression. Their knowledge was invaluable to expressing specifically labeled protein samples. Whereas Lionel and I expressed the H23 protein in differing conditions, I was not directly involved in the expression of the TET2 protein, this work undertaken at IBS. My work here consisted of analysis of the H23 spectra and quantification of residual protonation. This work was the subject of a book chapter: DOI 10.1007/978-1-0716-0892-0_8.

III.1. Introduction

In solution NMR, developments of 2D and 3D heteronuclear experiments for protein assignment and structure determination have called for uniformly[1] and fractionally[2] deuterated samples. Indeed, perdeuteration around the chemical groups of interest is crucial to limit fast transverse relaxation which hinders their signal resolution and sensitivity, especially for larger proteins for which there is increased spectral crowding and line broadening. But as NMR experiments evolved for specific chemical groups, such as the NH[3] and methyl[4] TROSY, so did the need for selectively labeled proteins, with isotopic labels introduced only at certain sites. *In vivo* bacterial expression of isotopically labeled proteins requires addition of labeled carbohydrate and ammonium salts source, which disperse the isotopes to all newly synthesized amino acids and leads to broad scrambling of these isotopes due to the presence of enzymes of amino acid metabolism. This is a useful strategy when looking to uniformly label protein samples and is still used to date. However, selective isotopic expression of an amino acid or chemical group is severely limited using *in vivo* techniques. Understanding bacterial biosynthetic pathways has helped to circumvent these undesirable enzymatic activities and develop procedures that allow production of proteins with the desired labeling pattern[5]. Most notably, specifically labeling methyls with ^{1}H in isoleucine, valine and leucine side chains using ^{13}C α-ketoacid precursors in *E.coli*[4,6–11] and the use of transaminase-deficient and auxotrophic strains of *E.coli* mutants for selective amino-acid labeling of proteins *in vivo* has also been

explored for methyl groups[12] for amide groups[13] and for amino acid residues[14].

However, enzymes responsible for amino acid metabolism cannot be entirely removed from *E.coli* organisms and inevitably lead to scrambling of labels between different amino acids, this is especially true for production of samples with maximum deuteration of all non-exchangeable protons or selectively ^{15}N labeled on a certain residue. *In vivo* perdeuterated expression is generally undertaken in a deuterated medium, requiring the transfer of the newly synthesized protein to a fully protonated medium for back-protonation of amide sites. However, amide protons located in hydrophobic cores of large proteins are difficult to back-exchange, often generating incomplete ^{1}H-^{15}N spectra. The complete H/D exchange of such proteins then requires a denaturation step in a protonated solvent, followed by renaturation which is not always possible. Some purified proteins are incapable of regaining their original structure in the absence of protein partners and the yields and quality of spectra are severely affected when the expression protocol requires denaturation/renaturation steps. Moreover, *in vivo* expression of selectively labeled proteins is limited by the numerous enzymes present in cell systems that can exchange isotopically labeled-atoms during reactions[5]. Among these reactions, transamination reversibly transfers amino groups from amino acids to keto acids. Following an oxidation step in the presence of the Pyridoxal 5'-phosphate cofactor (PLP), the Cα is reduced and recuperates a proton from the solvent at the Hα2 position (pro*R*). It had been previously observed that high levels of residual protonation on Cα and Cβ positions arose when expressing perdeuterated proteins in water[15]. This is one of the reason why D_2O solvent is preferred over H_2O when the purpose is to obtain high deuteration yield, even upon supplementation of deuterated amino acids[16].

In vitro protein expression can be particularly advantageous in these situations as amino acid metabolism is strongly reduced in such systems[17,18]. Although some enzymatic activity persists *in vitro*, particularly for glutamate, aspartate, glycine and serine, central to amino acid metabolism, label scrambling of deuterated amino acids in a water based solvent can be minimized by addition of inhibitors as was previously undertaken for selective ^{15}N labeling[19] and perdeuteration of backbone and side chains[18]. The inhibitors usually react with PLP conjugates often found in amino acid metabolism, but not necessary for transcription and translation. The cell machinery thus employs the provided amino acids for protein expression whilst enzymes responsible for isotopic scrambling remain inactive, producing the protein of

interest with the desired labeling, selective of ^{15}N amides or perdeuterated on the backbone and side-chains. We screened different cell-free protocol conditions by expressing the H23 protein in modified culture conditions, of which the addition of enzymatic inhibitors and the expression mode. The samples obtained were analysed with a particular attention to minimal Cα back protonation.

Technical developments have made cell-free protein expression more accessible and versatile, becoming a central tool in synthetic biology. Cell-free expression system is an open, *in vitro* protein synthesis approach that exploits the fraction of the cell machinery required for protein expression, discarding mechanisms of cell viability and growth, optimizing synthesis for exclusively the protein of interest. Transcription and translation machineries are extracted from prokaryotic[20] or eukaryotic[21] lysates, original DNA and RNA are removed, and the extract is supplemented with T7 RNA polymerase, amino acids, energy sources, and a DNA (or RNA) template. Compounds necessary for improving the protein of interest's expression and viability, of which cofactors, various stabilizers, inhibitors of RNAse and proteases for sensitive proteins, can be added at any point during expression. This approach facilitates control of the production environment to optimize synthesis of the protein of interest, leading to lower protein expression costs and times whilst increasing yields compared to *in vivo* techniques (Figure 3.1). There are a variety of applications of cell-free, notably for the synthesis of large or membrane proteins and the overexpression of enzymes that are toxic for cells. As far as the membrane proteins are concerned, their purification can prove difficult as detergents are needed to separate them from inclusion bodies or membranes, denaturing proteins and decreasing yields considerably. Cell free permits expression of membrane proteins in satisfying quantities and in the desired membrane mimetic environment.[22]

Figure 3.1: Schematic representation of an *in vivo* protein expression and cell free protein synthesis (CFPS). Cell-free systems are more cost effective than in vivo expression techniques as a 12 L cell growth makes enough cell extract for 50 to 200 productions, with the protein of interest ready in 3 days, whereas *in vivo* expression takes almost an entire week for each production. The figure was taken directly from Carlson et al., 2012[152] [23]

III.1.1. General principle of the cell free expression and set up

DNA-dependent cell-free expression is based on coupling transcription and translation *in vitro* using the cell machinery of a prokaryotic or eukaryotic cell-extract and a DNA vector (usually a plasmid optimized for cell-free expression, see section III.3.1) in which the gene of interest has been inserted. The key component in this synthesis is the cell extract which provides ribosomes and the majority of the necessary enzymes and biomolecular complexes for gene expression, such as tRNA-aminoacyl synthetase. The gene of interest is transcribed into messenger RNA by an added T7 RNA polymerase (RNA Polymerase of 98 kDa derived from the bacteriophage T7) and then translated by ribosomes into a peptide sequence. In Figure 3.2,

a double stranded DNA template is transcribed by the T7 RNA polymerase added to the cell lysate. This polymerase is commonly used in cell-free for its low error rate (average error frequency of $1/2*10^4$ [24]) its capability to synthesize long transcripts without additional proteins, and its specificity to the T7 promoter,[25] hence the need for a DNA vector containing the T7 promoter. The T7 RNA polymerase requires addition of cofactor Mg^{2+} to function correctly and its activity is stimulated by spermidine, a polyamine commonly found in cells. The newly synthesized mRNA will then be translated by the ribosomes of the cell lysate into a new polypeptide, with tRNA-aa delivering the correct amino acids, which will later be degraded for ATP regeneration. For ATP and nucleotide regeneration, the necessary molecules need to be added to the reaction mixture (as they are removed during cell extract preparation), as well as creatine phosphate and the enzyme creatine kinase, not present in sufficient quantities from the basal cell extract. Chaperone proteins present in the cell extract will ensure post-translational modifications and folding of the peptide sequence into an active protein.

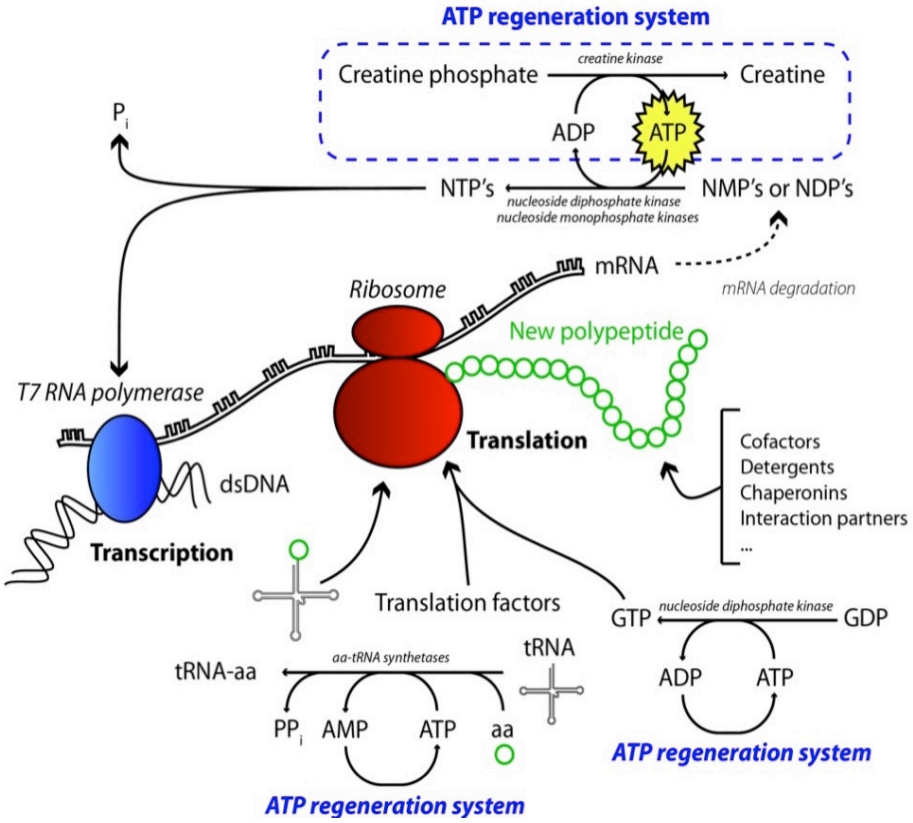


Figure 3.2: General scheme of cell free expression. (Adapted from Schneider et al., 2010[26] and Pederson et al., 2011[27])

III.1.2. Configuration of the system

There are two main configurations for cell-free systems which are used depending on the experimental applications (see Figure 3.3).

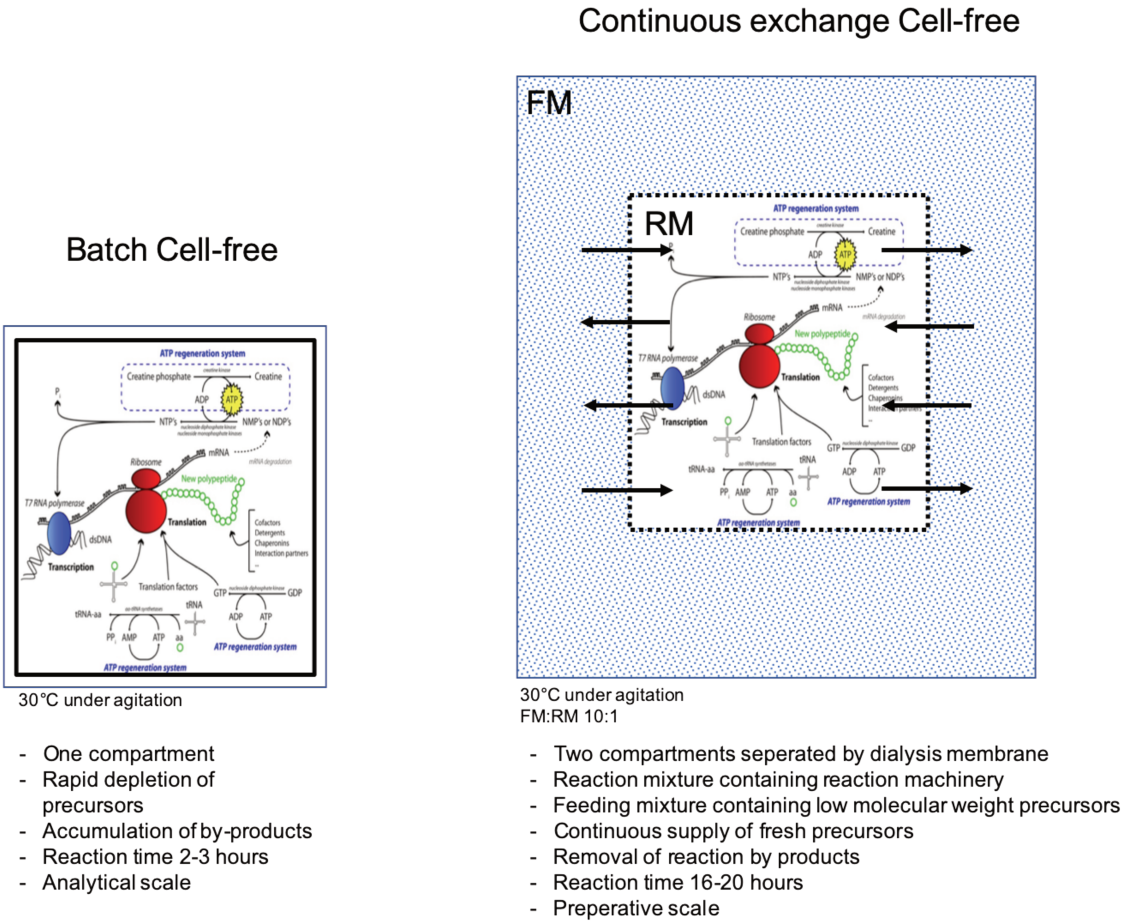


Figure 3.3 : Configuration of cell free systems. On the left the batch system, on the right the continuous exchange system. Adapted from Mus-Veteau et al., 2010.[26]

In the batch system, the reaction takes place in a single compartment, consuming compounds and accumulating reaction by-products which will change reaction conditions. Due to the rapid depletion of precursors, the duration of expression is limited to about 3 hours and reaction volumes are in the tens of microliter range, with yields up to 0.5 mg of protein per mL of reaction mixture, depending on the expressed protein. This set-up is used for analytical scale protein expression and a single batch will rarely yield the quantities necessary for NMR analysis. The continuous exchange system consists of two compartments separated by a semi-permeable membrane (8 kDa cut off). In the inner compartment is contained the reaction mixture (RM) consisting of the cell extract and other necessary components such as various proteins, enzymes and the expression vector which are too large for dialysis across the

membrane. The feeding mixture (FM) in the outer compartment provides prolonged supply of fresh precursors, such as nucleic acids, energy molecules and amino acids, and removal of reaction by-products of low molecular weight from the reaction mixture. Reaction times can be multiplied by a factor of 8 compared to the batch system. Moreover, this configuration is more efficient for protein expression with yields that can attain several mg of protein per mL of reaction mixture depending on the protein of interest, quantities sufficient for NMR studies.

III.1.3. S30 cell extract from *E.coli*

The cell lysate is indispensable for cell-free protein expression as it provides many of the enzymes and other molecules necessary for protein expression. The most common cell-lysates are derived from *E.coli* strains, optimized for high throughput enabling expression of up to several mg of protein per mL of reaction. Cell-lysates derived from yeasts, wheat germ[21] or mammalian cells also exist and are used for different purposes and shall not be discussed here. The exact composition of the *E.coli* cell extract composition is still not fully known and can vary greatly from one production to another. Fractioning the proteome to conserve only protein expression metabolic pathways implies partial or complete removal of many other metabolic pathways, resulting in a subjective composition of the S30 lysate. Some authors have identified the presence of 821 different proteins[28] found from the *E.coli* strain A19. The tables 3.1, 3.2 and 3.3 detail some of the different proteins isolated, such as transcription and translation related proteins, chaperones and proteases.

There are a multitude of protocols available for S30 lysate preparation[11–17] but a general schematic representation is given (see Figure 3.4). A full protocol for preparation of the S30 extract is presented at the end of the Chapter (Chapter III.3. Materials and methods). The protocol can be divided into three big steps: the cell culture, the lysate harvest and the processing of the extract. The *E.coli* lysate preparation begins with a cell culture and inoculation of medium with a fresh preculture of the chosen *E.coli* strain in LB medium. The cells are incubated in a fermenter at 37°C under stirring conditions and appropriate airing. Fermenters are practical for ensuring proper growth conditions, but it is possible to achieve this using flasks. When the cells reach an optimal optical density (OD_{600}), which corresponds to mid log growth phase ($OD_{600} \approx 4$), the fermenter is left to cool to room temperature to prevent further cell replication. The lysate harvest can begin with a series of centrifugation and wash steps. Cells are centrifuged at 6000 g for 20 minutes at 4°C and washed 3 times with Tris buffer. The pellet

is then resuspended in Tris buffer and cells are ruptured using mechanical force. Finally, extract processing includes two steps to retrieve the fraction of biomolecules according to size: centrifugation removes the bigger biomolecules whilst dialysis removes the smaller molecules. Centrifugation of the lysate is undertaken at 30 000g for 30 minutes and the supernatant is recovered for a second identical centrifugation step. The S30 denomination comes from the lysate centrifugation speed, 30 000 g, which determines a certain cut-off for the compounds retained. The supernatant is finally dialyzed (membrane of 12-14 kDa cut off) against an ice cold buffer twice, once for 3 hours and once overnight. The lysate in then centrifuged one last time at 30 000 g for 30 minutes to remove precipitates and the clarified lysate is shock-frozen with liquid nitrogen and stored at -80°C.

Transcription and translation related proteins			
	UniProt #	**Description**	**Gene ID**
RNAP	P0A7Z4	DNA-directed RNA polymerase subunit alpha	rpoA
	P0A8V2	DNA-directed RNA polymerase subunit beta	rpoB
	POA8T7	DNA-directed RNA polymerase subunit beta'	rpoC
	P0A800	DNA-directed RNA polymerase subunit omega	rpoZ
	P60240	RNA polymerase-associated protein	rapA
σ-factors	P00579	RNA polymerase sigma factor	rpoD
	P24255	RNA polymerase sigma-54 factor	rpoN
	P24251	Sigma factor-binding protein.	crl
Termination	P0A6W5	Transcription elongation factor	greA
	P0AG30	Transcription termination factor	rho
	P0AFF6	Transcription termination/antitermination protein	nusA
	P0AFG0	Transcription termination/antitermination protein	nusG
Cold-shock	P0A9 x 9	Cold shock protein	cspA
	P0A9Y6	Cold shock-like protein	cspC
	P0A972	Cold shock-like protein	cspE
	P0A978	Cold shock-like protein	cspG
Initiation	P69222	Translation initiation factor IF-1	infA
	P0A705	Translation initiation factor IF-2	infB
	P0A707	Translation initiation fcator IF-3	infC
Elongation	P60785	Elongation factor 4	lepA
	P0A6M8	Elongation factor G	fusA
	P0A6N4	Elongation factor P	efp
	P0A6N8	Elongation factor P-like protein	yeiP
	P0A6P1	Elongation factor Ts	tsf
	P0CE47	Elongation factor Tu 1	tufA
	P0A9W3	Energy dependent translational throttle protein	ettA
	P0AFX0	Ribosome hibernation promoting factor	hpf
	P0A832	SsrA-binding protein	smpB
	P45748	Threonylcarbamoyl-AMP synthase	tsaC
Termination	P0A710	Peptide chain release factor 1	prfA
	P07012	Peptide chain release factor 2	prfB
	P0A714	Peptide chain release factor 3	prfC
	P0A805	Ribosome recycling factor	frr

Table 3.1: List of S30 lysate transcription and translation related proteins

Chaperones			
	UniProt #	**Description**	**Gene ID**
cytoplasmic	P0A850	Trigger factor	tig
	P0A6F9	10 kDa chaperonin	groS
	P23869	Peptidyl-prolyl cis-trans isomerase B	ppiB
	P0A6Y8	Chaperone protein	dnaK
	P0A6F5	60 kDa chaperonin	groL
	P0A9L5	Peptidyl-prolyl cis-trans isomeraase C	ppiC
	P0A9K9	FKBP-type peptidyl-prolyl cis-trans isomerase	slyD
	P0A9L3	FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase	fklB
	P25522	tRNA modification GTPase	mnmE
	P77395	Uncharacterised protein	ybbN
	P10408	Protein translocase subunit	secA
	P0AEMO	FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase	fkpB
	P08622	Chaperone protein	dnaJ
periplasmic	P0AEG4	Thiol disulfide interchange protein	dsbA
	P45523	FKBP-type peptidyl-prolyl cis-trans isomerase	fkpA
	P0ABZ6	Chaperone protein	surA
	P31697	Chaperone protein	fimC
	P0AFL3	Peptidyl-prolyl cis-trans isomerase A	ppiA
	P0AEU7	Chaperone protein	skp
	P0AEG6	Thiol disulfide interchange protein	dsbC

Table 3.2: List of S30 lysate chaperones.

Proteases			
	UniProt #	**Description**	**Gene ID**
	P0A6G7	ATP-dependent Clp protease proteolytic subunit	clpP
	P0ACF4	DNA-binding protein HU-beta	hupB
	P27298	Oligopeptidase A	prlC
	P0A7B8	ATP-dependent protease subunit	hslV
	P0A9M0	Lon protease	lon
	P0A7C2	LexA repressor	lexA
	P12008	Chorismate synthase	aroC
	P23865	Tail-specific protease	prc
	P0C0V0	Periplasmic serine endoprotease	degP
	P0AAI3	ATP-dependent zinc metalloprotease	ftsH
	P05458	Protease 3	ptrA
	P0A898	Endoribonuclease	ybeY
	P24171	Peptidyl-dipeptidase	Dcp
	P0AEB2	D-alanyl-D-alanine carboxypeptidase	dacA

Table 3.3: List of S30 lysate proteases.

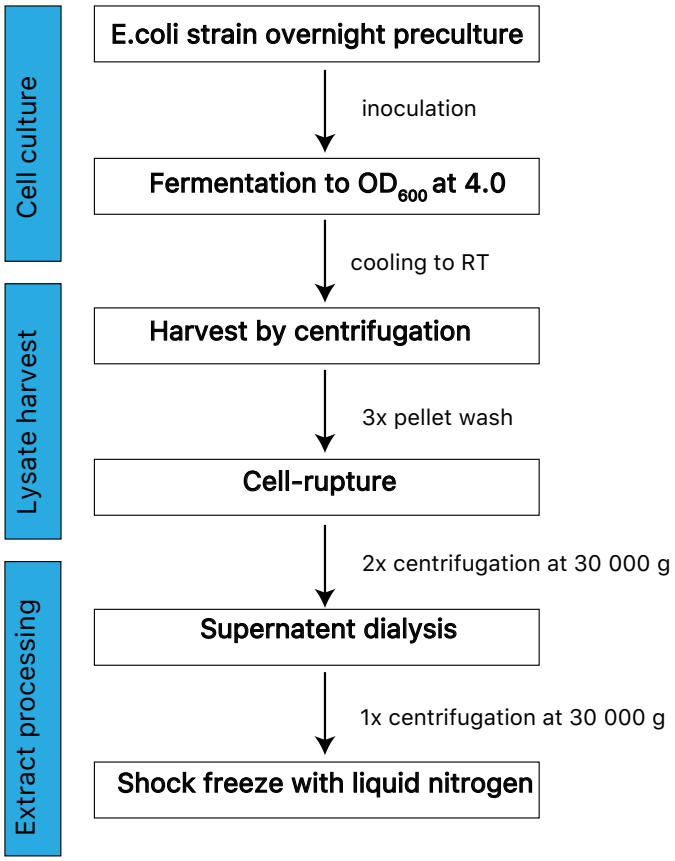


Figure 3.4: Representation of the major steps for S30 lysate preparation adapted from Foshag et al., 2018.[28]

III.1.4. Amino acids supply

The cut off of the dialysis performed for the S30 lysate production is large enough (12-14 kDa cut off) that native nucleic acids and amino acids do not remain in sufficient quantities to allow for protein expression. However, even though the lysate endures 2 dialysis during its preparation, it is possible that trace amounts of amino acid stay present, the dialysis membrane only diluting their presence by several factors. Their presence, minor as it could be, would be problematic for the specific labeling required for the application we are considering in this thesis work (see section III.1.5. Isotopic scrambling during cell-free expression)). The interest of using cell-free for our protein expression is to obtain the incorporation of specifically labeled amino acids, notably glycine, in a completely perdeuterated background. Different amino acid mixes were added to our production system to optimize the desired labeling; either added individually from commercially available powders or collectively from algae lysates which provide almost all amino acids uniformly labeled. These lysates, sold under brands such as

Celtone ® or Isogro ®, are produced by growing algae (no information found about the type of algae) with stable isotopes. Amino acids are then extracted, with other compounds, and lyophilized into powder form. The table below indicates the general composition of Isogro ® Growth Media.

Composition:

Salts	30%
Water	3%
Glucose	2%
Amino acids/Peptides	65%

Amino Acid Analysis:

Ala	13%	Ile	4%	Pro	4%
Arg	3%	Leu	8%	Ser	4%
Asp	14%	Lys	6%	Thr	5%
Glu	10%	Met	3%	Tyr	3%
Gly	12%	Phe	4%	Val	6%
His	1%				

Table 3.4 : ISOGRO composition and amino-acid – approximative values as there are variations between batches. Values are given as a percentage of dry weight.

Isotopic labeling of amino acids from these growth media are available in a number of different types, U-^{13}C, U-^{13}C-^{15}N, U-^{13}C-^{15}N-^{2}H, U-^{15}N-^{2}H from different manufactures and labeling ranges around 98%, although this can vary considerably depending on batches and manufactures. There are four amino acids missing from the growth media: cysteine, tryptophan, asparagine and glutamine. Asparagine and glutamine can easily be regenerated from aspartic and glutamic acids with the action of asparagine and glutamine synthetases, respectively, which are present in the S30 extract. However, cysteine and tryptophan have to be added individually if required, since the amino-acid-synthesis-enzymes don't have the necessary substrates in the cell-free extract to regenerate these two amino acids. Cysteine synthase requires dihydrogen sulfide to synthesize cysteine whereas tryptophan synthesis requires chorismic acid, precursor to indole derivatives of which tryptophan. Finally, treatment of the algae mixes with Pronase® before use (a commercially available mix of nonspecific endo and exo-proteases isolated from *Streptomyces griseus* that digest proteins and peptides into single amino acids) increases protein yields by generating more available amino acids for protein expression.

III.1.5. Isotopic scrambling during the cell-free expression

As *in vivo*, the S30 extract contains enzymes involved in metabolic pathways other than protein synthesis, notably amino acid metabolism[35]. A number of these enzymes are pyridoxal-phosphate (PLP) dependent enzymes, catalyzing a variety of reactions on substrates containing an amino group, of which racemization, transamination and decarboxylation[36]. However, many of these enzymes are source of isotopic scrambling. The following table (Table 3.5) lists some of the enzymes identified in the S30 extract responsible for metabolizing amino acids, notably transaminases such as aspartate aminotransferase (aspC, Figure 3.5), and phosphoserine aminotransferase (serC), amongst others not mentioned, problematic for ^{15}N selective amino acid labeling (applicable to the labeling strategy exposed in Chapter IV) and for perdeuteration of amino acids in H_2O. Serine hydroxymethyl transferase (glyA) will reversibly exchange the serine side chain for a proton from the solvent, creating protonated glycine (see Chapter IV). Some of these enzymes, such as asparagine synthetase B (asnB, Figure 3.7) and glutamine synthetase (glnA) are necessary to replenish in asparagine and glutamine from aspartic acid and glutamic acid if algae mixes are used. The other enzymes listed (speA, cysK, cysM, gltB, gltD, gshA, gor, ansA, sdaA, metK, cysE, trpA and tnaA) don't intervene directly in the isotopic scrambling that we wish to avoid here, as there is no need for selective amide labeling.

Amino acid modifying enzymes		
UniProt #	**Description**	**Gene ID**
P22106	Asparagine synthetase B (glutamine-hydrolyzing)	asnB
P00509	Aspartate aminotransferase	aspC
P21170	Biosynthetic arginine decarboxylase	speA
P0ABK5	Cystein synthase A	cysK
P16703	Cystein synthase B	cysM
P09831	Glutamate synthase (NADPH) large chain	gltB
P09832	Glutamate synthase (NADPH) small chain	gltD
P0A6W9	Glutamate-cystein ligase	gshA
P0A9C5	Glutamine synthetase	glnA
P06715	Gluthatione reductase	gor
P0A962	L-asparaginase 1	ansA
P16095	L-serine dehydratase 1	sdaA
P23721	Phosphoserine aminotransferase	serC
P0A817	S-adenosylmethionine synthase	metK
P0A9D4	Serine acetyltransferase	cysE
P0A825	Serine hydroxymethyltransferase	glyA
P0A877	Tryptophan synthase alpha chain	trpA
P0A853	Tryptophanase	tnaA

Table 3.5 *:* List of enzymes responsible for amino acid scrambling adapted from Foshag et al., 2018[28].

Transaminases reversibly catalyze transamination reactions between amino acids and keto-acids, exchanging the amine function between two amino acids, via PLP. The general mechanism is presented below (Figure 3.5). The reaction can be described in 4 steps (Figure 3.6). During step 1, a keto acid will react with the amide of pyridoxamine phosphate, forming a ketimine between the nitrogen and the Cα of the keto-acid. The ketimine is then converted via asymmetric hydrogen shift by the enzyme in steps 2 and 3, forming an aldimine intermediate with the substrate and Schiff base between the substrate and the PLP. During the reduction of the keto-acid, a proton from the solvent intervenes on the Cα. In step 4, the aldimine is hydrolyzed to an amino acid and pyridoxal phosphate. The pyridoxamine phosphate can be regenerated from pyridoxal phosphate and another amino acid via a reverse process in step 5.

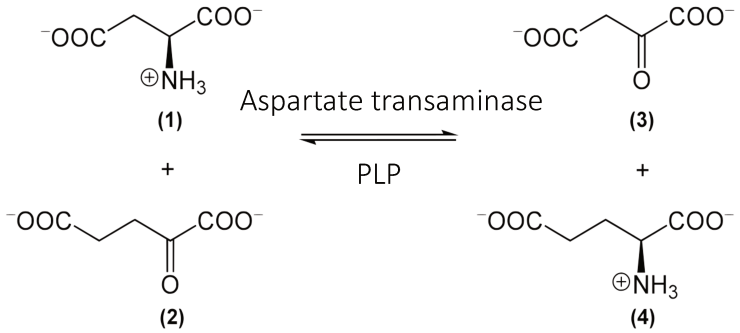


Figure 3.5: General mechanism of action of an aspartate aminotransaminase. Asparagine (1) and glutamate (2) are catalyzed into aspartate and glutamine.

Figure 3.6: Step by step mechanism of the transamination process.

Figure 3.7: Reaction catalyzed by the asparagine synthetase

Asparagine synthetase functions by converting aspartate and glutamine to asparagine and glutamate in an ATP-dependent reaction. In this reaction, the carboxyl of the aspartate is activated by an ATP molecule, forming a β-aspartyl-AMP intermediate. The subsequent deamidation of the glutamine leads to the release of ammonia, which then performs a nucleophilic attack on the β-aspartyl-AMP to produce asparagine and glutamate.

The major issue with the presence of transaminases is the protonation of Cα positions by recuperation of a proton from the solvent on the Hα2 (*proR*) position for deuteration labeling strategies which occur in water (see Figure 3.8). Our expression medium, therefore, requires addition of certain inhibitors of these transaminases to prevent amino-acid scrambling and limit the risk of incorporating protons from the solvent on Cα positions of deuterated amino acids. There is also another issue with the presence of transaminases in the medium, not applicable here but in the next Chapter, which is the transfer of the labeled amide to other amino acids, for selective labeling strategies, resulting in its presence on various other amino acids.

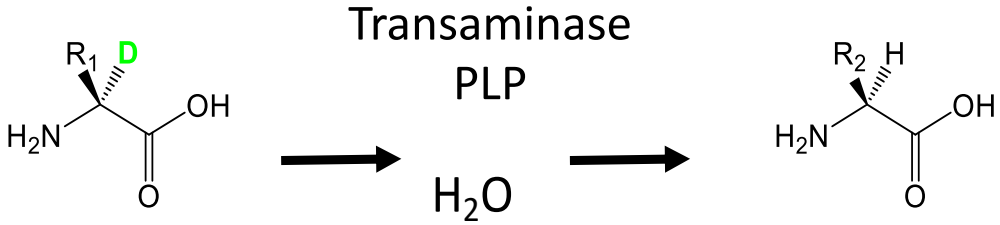


Figure 3.8: Schematic overview of the isotope scrambling resulting from the transaminase activity, present in the S30 extract.

III.1.6. Inhibition of PLP dependent enzymes

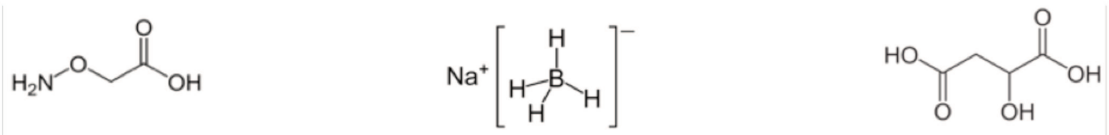
One of the most useful aspects with the cell-free protocol is that it is an open system, enabling addition of specific reagents. There are many inhibitors available with different efficacy to avoid amino acid metabolism during the cell free expression[18,19] for selective

isotopic labeling. As stated before, a number of enzymes involved in amino acid metabolism are PLP dependent, such as transaminases.

Figure 3.9: Pyridoxal phosphate free or bound by a Schiff base to either the ε-amino group of a lysine residue of the enzyme or to the amino group of a substrate.

In PLP-dependent enzymes, the carbonyl group of the cofactor binds to the ε-amino group of a lysine residue of the catalytic site of the enzyme to form an internal aldimine. During the reaction, the amino group of the amino acid substrate replaces the ε-amino group of the lysine at the site of the Schiff base to form an external aldimine. Two of the three inhibitors used in our cell-free protocols (Figure 3.10) inhibit this cofactor either by targeting it directly or irreversibly reducing the Schiff-base linkage it makes with the enzyme lysine. Aminooxyacetic acid (AOA) attacks the Schiff base linkage by forming an oxime type complex[37], whereas sodium borohydride ($NaBH_4$), a reducing agent, reduces this Schiff base in the PLP-enzyme complex to form a chemically stable secondary amine[19].

L-aspartate ammonia-lyase (AAL) reversibly deaminates Asp to form fumarate and ammonia, fueling the Tricarboxylic Acid cycle (TCA), and leading potentially to formation of other amino acids from the isotopically labelled amide. The addition of D-malic acid (DM : D-malate), as a competitive inhibitor of AAL, to the cell-free medium can prevent this isotopic scrambling[38,39].



AOA :Aminooxyacetic acid $NaBH_4$: sodium borohydride DM : D-malic acid (D-malate)

Figure 3.10: Illustrations of the three inhibitors used in the cell-free protocol described here, from left to right, AOA aminooxyacetic acid, NaBH4 : sodium borohydride, DM : D-malic acid. The first two inhibitors react on PLP, cofactor used by many amino acid metabolism enzymes, DM reacts specifically with L-aspartate ammonia lyase, as described in the main text. Underneath is represented the inhibition of PLP-dependent enzyme by the $NaBH_4$ by irreversible reduction of the Schiff-base linkage, taken directly from Su et al.[19]

An issue raised with the use of inhibitors is whether asparagine and glutamine would not be replenished in sufficient quantities in the reaction mixture if algae mixes are used as the amino acid source. If amino acid scrambling is completely inhibited, they may not be in sufficient amounts in the expression media which could limit protein expression, affecting yields. However, asparagine synthetase is not PLP dependent, and addition of inhibitors targeting that cofactor do not affect the enzyme's activity. In any case, issues relating to missing amino acids in the cell-free mixture can easily be solved by separate addition of those amino acids in the cell-extract. Note that if cysteine and tryptophan are present in the protein sequence to be expressed, those two amino acids also have to be added to the reaction mixture separately, as the precursors are in insufficient amounts for replenishment and these amino acids are not available in algal extracts (cf. II.1.4 Amino acid supply).

Another issue raised is that use of $NaBH_4$ to inhibit PLP-dependent enzymes can severally diminish protein yields, sometimes by several factors, because as a strong non-specific reducer it may also affect the integrity of other proteins not PLP dependent. The S30 extract is treated with $NaBH_4$ before use in cell-free reactions and is eliminated in two dialysis steps, though trace amounts can remain which could also affect the cofactors added after treatment. However, that is a compromise that has to be made between quality and quantity of the protein samples. Optimization of the cell-free protocol for optimum sample quality was undertaken and applied to other proteins, Tet2 (this Chapter) and the protein Pin1 (Chapter IV).

Moreover, although these inhibitors ensure limited amino acid isotopic scrambling on the Cα of residues, Toneli et al.,[18] have reported that addition of AOA in the cell free mix did not prevent partial exchange at certain side-chain methyl and methylene groups : Ala-Hβ, Asn-Hβ, Asp-Hβ, Gln-Hγ, Glu-Hγ, and Lys-Hε, noting that this exchange should be of interest for the study of large proteins and biomolecular complexes.

III.1.7. Plasmids for cell-free expression

There are specific plasmids that can be used for cell-free expression: pIVEX vectors, (plasmid for *In Vitro* Expression), with a His_6-tag fusion on the N-terminus of the protein. Designed by Roche, these vectors are optimized for high throughput in cell-free systems, containing regulatory elements necessary for *in vitro* expression using T7 polymerase in *E.coli* lysate, although, in practice, higher expression rates are not necessarily guaranteed. The Figure 3.16 represents the pIVEX 2.4d vector used to insert the gene coding for the H23 protein. It is designed with an origin of replication (ORI) as well as an antibiotic resistance gene (AmpR), here against ampicillin, for cloning. It also includes a T7 promoter comprising the gene 10 translation enhancer,[45] and a T7 terminator that helps prevent 3'-terminal degradation of the mRNA by exonucleases[46], a prokaryotic Shine-Dalgarno sequence (a ribosomal binding site in bacterial messenger RNA, located usually 8 bases upstream of the start codon AUG), a Multiple Cloning Site for insertion of the target DNA sequence, and a His-tag sequence at the N or C-terminus to facilitate purification of the expressed protein,

Figure 3.16: Representation of the pIVEX 2.4d plasmid used for H23 cell-free expression. An origin (Ori) of replication, a gene for antibiotic resistance (here ampicillin AmpR), a T7 promoter and a multiple cloning site.

III.2. Cell-free protocol for the perdeuteration of proteins in H_2O

Protocols were optimized on the H23 protein, a small subunit of a heavy metal ATPase found in *Arabidopsis thaliana*[40] (See Chapter II.2). Long of 155 amino acids with no cysteine and a single tryptophan in its sequence, and with a mass of 16,728 kDa (Protparam HMA8), this protein was chosen for protocol optimization as it is efficiently expressed by cell-free.

Demonstration of the utility of the optimized cell-free expression protocol for perdeuteration of large proteins in $^{1}H_2O$ was undertaken on the TET2 protein assembly. TET2 (pdb 6R8N) is a 468 kDa homododecameric tetrahedral (TET) aminopeptidase that hydrolyzes small peptides into amino acids in prokaryotes and Archaea and bacteria[41], subject of study by our partners at the Institut de Biologie Structurale of Grenoble[42].

Protocol optimization was necessary to ensure the efficient deuteration for all amino acids and minimal residual protonation at the non-exchangeable Cα positions. A deuterated background is important to avoid the spread of ½ spins which would increase the transverse relaxation of the NH groups in the NH TROSY experiment. For this purpose, uniformly perdeuterated H23 protein was produced in batch mode or continuous exchange cell-free (CECF) mode, in H_2O based solvent, with or without adapted quantities of a cocktail of enzyme inhibitors (aminooxyacetic acid (AOA), borohydride ($NaBH_4$), and D-malic acid (DM)) to ensure that there was limited amino acid scrambling activity, which would result in incorporation of proton at the Hα2 position (pro*R*). Perdeuterated amino acids were supplemented either from algal extracts U-^{2}H-^{13}C-^{15}N (batch) or from a mix (CECF mode). After expression, reaction mixtures were centrifuged and supernatant purified to collect the protein of interest. A complete protocol for the cell-free expression system is given in section III.3. Materials and Methods.

III.2.1. Quantifying Cα protonation using the optimized protocol

Quantification of residual protonation was obtained by measuring the volume of ^{1}H-N-Cα(^{1}H)and ^{1}H-N-Cα(^{2}H) correlations from 3D HNCA spectra (see Chapter III.3.6. Materials and Methods for recording and processing conditions) using the integration module of the NMRPipe software [43]. To distinguish the Cα–^{1}H from Cα–^{2}H, the deuterium was decoupled during ^{13}C edition, while the ^{1}H decoupling was omitted. Percentage of ^{1}H or ^{2}H isotopes on Cα positions was deduced from the volume ratio of the correlations corresponding to the ^{13}C-^{1}H and the ^{13}C-^{2}H pairs

Coupling constants between the ^{13}Cα and ^{1}Hα of amino acids are usually in the 135 Hz to 145 Hz range. When the proton is replaced by a deuterium, an isotope effect leads to an

upfield chemical shift of 0.34 to 0.5 ppm for the carbon resonances[4] which would represent a shift of about -80 Hz on the spectrometers used (600 and 850 MHz). As deuterium was decoupled during ^{13}C edition, Cα bearing an $^{2}H\alpha$ would appear as a single peak, as ^{1}H was not decoupled, Cα bearing $^{1}H\alpha$ would appear as a doublet (or triplet for Gly residues bearing two ^{1}H although this possibility is unlikely and was not seen in practice on our spectra). Due to the isotope effect on the ^{13}C chemical shift, the decoupled Cα-{$^{2}H\alpha$} singlet is very close to the upfield component of the Cα-$^{1}H\alpha$ doublet. The Figure 3.11 illustrates theoretical resonances that are observed in the ^{13}C dimension for C$\alpha^{2}H$ and C$\alpha^{1}H$ moieties of amino acids.

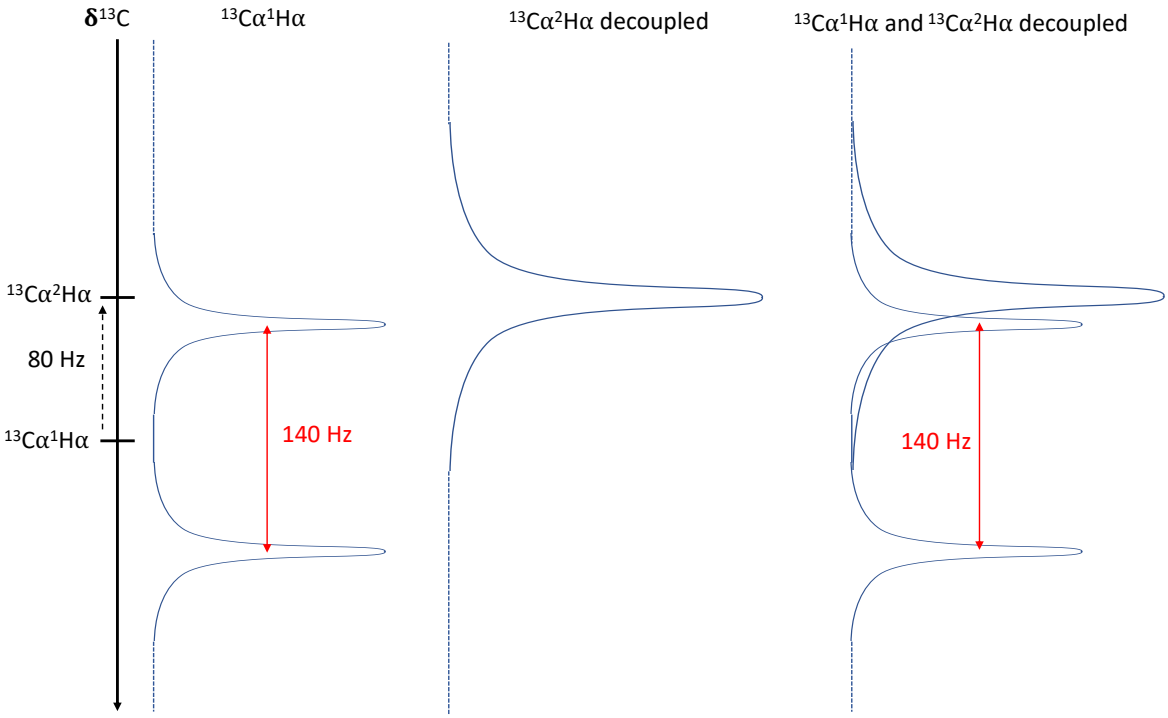


Figure 3.11: Schematic illustration of resonances corresponding to the C$\alpha^{2}H$ and C$\alpha^{1}H$ coupling in the carbon dimension. The C$\alpha^{1}H\alpha$ coupling is about 140 Hz. The C$\alpha^{2}H$ resonance is shifted about 80 Hz upfield. This means that, when both the C$\alpha^{1}H\alpha$ and C$\alpha^{2}H\alpha$ forms are present, there is only a small difference between the upfield component of the C$\alpha^{1}H\alpha$ doublet and the C$\alpha^{2}H\alpha$ resonance, they would therefore not be resolved enough considering the peak linewidth (about 5 Hz).

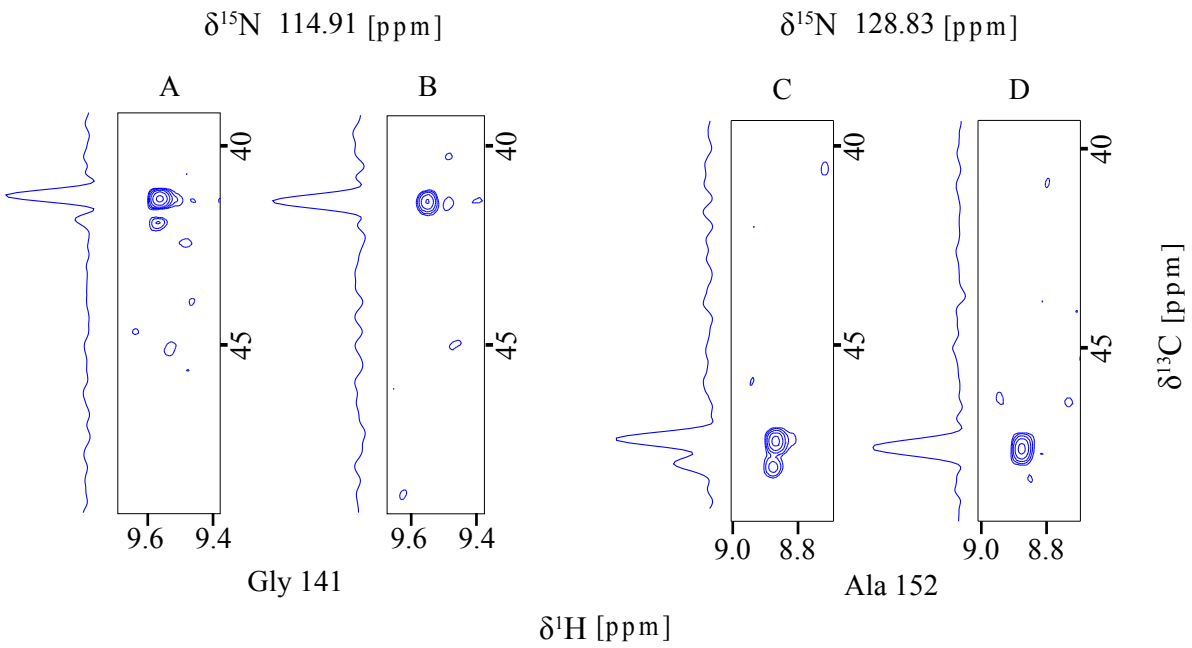


Figure 3.12: 2D strips extracted from 3D HNCA spectra of H23 produced in cell free dialysis mode of Gly 141 and Ala 152, uniformly ^{2}H-^{15}N-^{13}C- labeled, in H_2O, without (panel A and C) and with (panel B and D) three transaminase inhibitors (AOA, DM and $NaBH_4$). 850 MHz, Tris buffer 20 mM pH 7.5, 100 mM NaCl, 300 µM without inhibitors, 100 µM with inhibitors, 25°C. The presence of the major and minor peaks testifies the presence of protonated Cα carbon of the amino acids.

For each residue type, the results displayed on Figure 3.13 are the percentage of residual protonation calculated from the mean of 2 to 7 amino acids (exceptions are for histidine in both production modes, phenylalanine and tryptophan in batch mode and proline in CECF mode, where only one amino acid was quantifiable). Percentage of residual protonation was calculated by dividing the volume found for the minor peak (V_m) by the sum of the volume of the minor peak and major peak (V_M) : $\%^{1}H = V_m/(V_m+V_M)$ for each peak and averaging for each type of residue. Quantification was not undertaken for methionine and aspartic acid because correlations were overlapping. Tryptophan is absent in CECF mode (B) because it was added in an unlabeled form. The data for those amino acids, denoted by an asterix (*), were therefore taken from Su et al.[19]. There are no quantifications for cysteine as this protein does not have any in its sequence.

Residual protonation on Cα sites in proteins produced *in vitro* using 3 inhibitors is reduced by several factors, with a maximum residual protonation of 10 % for few amino acids, and an overall mean of less than 5% residual protonation for both production modes, in batch or in CECF mode.

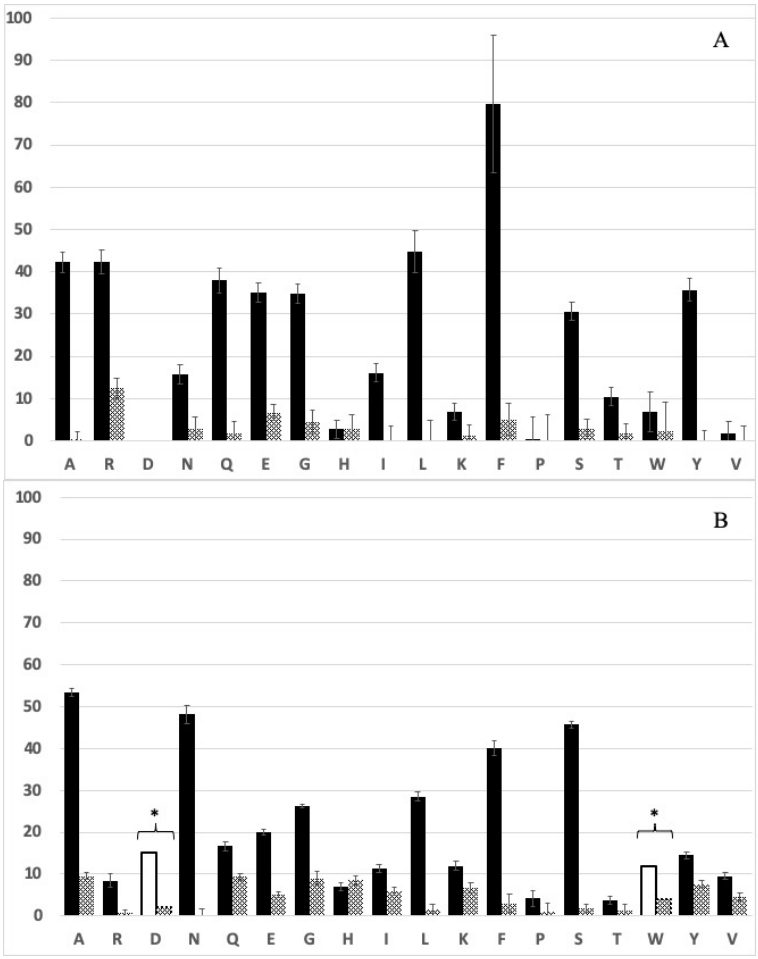


Figure 3.13: Residual protonation levels (%) detected on Cα-position of the H23 protein expressed *in vitro* using $^{1}H_2O$ and perdeuterated amino acids. Cell-free expression was undertaken following protocols in Material and Methods III.3. hydrolyzed U-^{2}H-^{13}C-^{15}N labeled algal amino acid mix (Celtone ®) complemented with unlabeled tryptophan using the batch mode (A), or the CECF mode containing 20 U-^{2}H-^{13}C-^{15}N amino acids (Merck) was used (B. Without inhibitors (black) or with inhibitors (20 mM AOA, 20 mM D-malate and $NaBH_4$ S30 extract treatment) (grey). *There are no values found for the Asp residue in batch mode, expressed with or without inhibitors as correlations were not identifiable or overlapping and none could be found in the literature for these expression conditions. The values are therefore taken from Su et al.,[19]

Su et al.,[19], undertook a different approach to analyze residual protonation : they assessed directly the ^{2}H–^{1}H exchange of the 20 natural ^{2}H-^{15}N-^{13}C-labeled amino acids after 7 h of cell-free reaction using reduced and conventional S30 extract without addition of DNA in CECF mode, by the integral of a ^{13}C-HSQC cross-peak of the amino acid in the feeding mixture. Likewise, they also observed less intense correlations for cell extracts not treated with NaBH4 estimating at 5% on average the rate of ^{2}H–^{1}H exchange at any position and less than 1% for the Cys, His, Ile, Leu, Pro, Thr and Val residues when using the $NaBH_4$ treatment, which is in fair agreement with our own findings.

Analysis was only undertaken for the Cα positions of amino acids, although Cβ positions are also subject to ^{2}H-^{1}H exchange[44]. Moreover, addition of inhibitors of PLP dependent enzymes does not take into account all the other enzymes involved in amino acid metabolism which do not get inhibited and whose activity can still partially protonate undesirable positions of the amino acids. However, for the purpose of ^{1}H-^{15}N TROSY on high

molecular mass proteins, the presence of Hα is the more deleterious for NMR signal since it is next to the amide protons.

III.2.2. Demonstrating the utility of the optimized cell-free protocol for the expression of the TET2 protein

In order to appreciate the usefulness of this labeling strategy, the application of the cell-free protocol to the protein TET2 was undertaken and the spectra available on Figure 3.14. Spectrum A makes it possible to clearly visualize ^{1}H-^{15}N correlations in the 9-10 ppm region in the ^{1}H dimension, which usually brings together amide groups from highly structured regions. Spectrum B corresponds to the same region for a sample produced in D_2O. Dialysis in H_2O allows the D/H exchange for the amide groups accessible to the solvent, certain correlations are however not visible on this spectrum which indicates that the corresponding amides are not in contact with the solvent, even after a 10 days incubation time. Such a slowing of the H/D exchange is common in the hydrophobic cores of high molecular weight proteins. To visualize these resonances again, a denaturation step in H_2O would be necessary, and would of course involve a renaturation step, with uncertain outcome. Finally, Image C corresponds to the TET2 protein expressed in H_2O light water but dialyzed against a D_2O solvent. This is ultimately a "negative" spectrum compared to the previous one where all the correlations that were not visible are now the only ones present. This is because the amino acids accessible to the deuterated solvent then disappear from the ^{1}H-^{15}N spectrum. Conversely, protonation is conserved in hydrophobic cores despite the use of heavy water. Superimposing images B and C would allow us to find image A. We therefore understand the interest of cell free labeling for the deuteration of high molecular weight proteins. This work has been the subject of a book chapter to be published in Structural Genomics: General Applications, Methods in Molecular Biology, vol. 2199 (paper doi.org/10.1007/978-1-0716-0892-0_8)

III.2.3. Conclusions

These spectral data clearly illustrate the utility of this cell-free culture protocol optimized for deuteration. First, the use of transaminase inhibitors reduces the residual protonation on the Cα to less than 5% on average. This very high deuteration rate is essential in order to limit the relaxation induced by the protons as much as possible and to obtain quality ^{1}H-^{15}N spectra for a protein of such large molecular mass as TET2. Expression in *E. coli* using

H_2O as solvent would not have made it possible to obtain a satisfactory rate of deuteration, even by adding a source of deuterated amino acids, due to the numerous enzymatic conversions within the cells, in particular the activities of transaminases. Culture in *E. coli* using D_2O as a solvent would have ensured complete deuteration but would not have achieved H/D exchange of all of the amide groups. The cell-free protocol described in this thesis, on the other hand, allows the expression of highly deuterated proteins in water. This approach guarantees observation of all the NMR signals of the amide groups, while preserving a very high level of deuteration at the level of the carbon chains, which is necessary for high molecular weight objects.

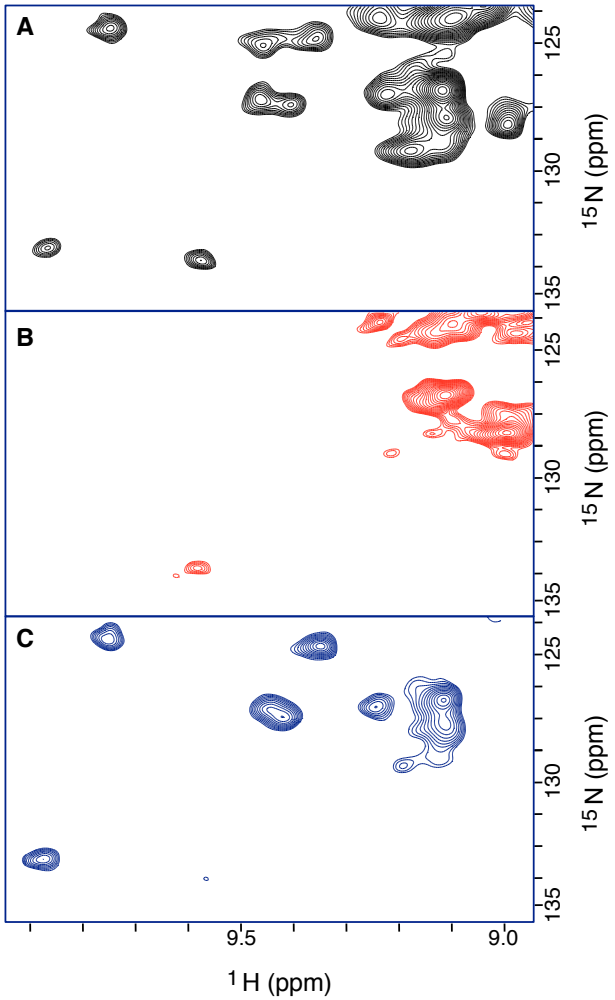


Figure 3.14: 2D-(^{15}N, ^{1}H) extracts from solution NMR spectra of the TET2 protein using the CRINEPT-HMQC-TROSY experiment (Riek PNAS 1999) 950 MHz Bruker Avance III HD spectrometer equipped with 5 mm cryogenically cooled pulsed-field-gradient triple-resonance probe at 50 ° C. U- [^{2}H-^{15}N] TET2 was produced in vitro (CECF mode), (200 µL at 10 µM) 20 mM TRIS buffer (pH 7.4) 20 mM NaCl, 4 mm shigemi tubes. (A) The sample used to acquire the spectrum was produced using $^{1}H_2O$ as a solvent during cell-free protein synthesis and a mixture of amino acids from hydrolyzed U- ^{2}H-^{15}N Celtone® supplemented with unlabeled W and C amino acids. Sample (B) was produced using the same source of amino acids but D_2O as a solvent during synthesis, and the purified protein was then dialyzed in H_2O buffer and stored at room temperature for 10 days to allow substitution. ^{2}H-> ^{1}H exchangeable. Sample (C) was produced as described for sample (A), but the purified protein was dialyzed in D_2O buffer and stored at room temperature for 10 days to allow ^{1}H-> ^{2}H substitution to detect signals of the amide positions protected by the solvent. These experiments were recorded at the IBS of Grenoble by Jérôme Boisbouvier.

III.3. Materials and methods

Protocols for S30 extract preparation and cell-reaction were taken directly from Imbert et al., 2020, DOI 10.1007/978-1-0716-0892-0_8. The H23 protein and TET2 produced for

this Chapter were done following the expression protocol described here. The purification protocol for H23 is detailed in Chapter II.

III.3.1. *E.coli* extracts preparation (S30 extract)

S30 extracts were prepared in a fermenter. The protocol detailed below corresponds to a 12 L-fermenter culture of *E. coli* BL21 DE3 strain, enabling the preparation of 200 mL of S30 extract. RNAses contaminations can occur anytime during the protocol, so strict cleaning procedures with RNAse remover should be applied for glassware and RNAse free plastic consumables should be used. Moreover, as ribosomes are temperature sensitive, S30 extract preparation must be performed on ice with chilled glassware/plastics.

1. Fermenter preparation: The fermenter tank is filled with 9 L of H_2O and 3 L of Z-media 4X (165 mM KH_2PO_4, 664 mM K_2HPO_4, and 40 g/ L yeast extract) and autoclaved *in situ.* We used a Techfors-S fermenter (Infors HT) equipped with a steam generator allowing *in situ* sterilization. Approximatively 1 L of water is lost by evaporation during the autoclave of the culture medium. 400 mL of LB medium are inoculated with 50 µL of BL21 DE3 cells (glycerol stock) at 37°C overnight.

2. *E. coli* culture: In the morning, the medium is warmed to 37°C, with the stir set at 550 rpm and the airflow at maximum level. 1 L autoclaved glucose-thiamine solution (240 g of glucose and 12 mg of thiamine per liter) is added with a sterile funnel and the fermenter is inoculated with the 400 mL overnight culture at an initial OD_{600} of 0.1. pH, pO_2 and OD_{600} is followed every 30 minutes. A decrease of pO_2 is expected together with an increase of the optical density. Culture should be stopped when the value of pO_2 is close to zero or the OD_{600} equal to 3, usually occurring 3 to 4 hours after the beginning of the culture. The temperature must be quickly reduced from 37 to 16°C, and the OD_{600} should not exceed 3.2. The cells are harvested by centrifugation for 15 minutes, 5,000 *g*, 4°C.

3. Preparation of S30 extracts.
   Wash: 3 washing steps of the cell pellets are undertaken with cooled S30 buffer, the first one with 2 liters, the second one with 1 liter and finally with 0.5 liter. At each step, the

cells are pelleted at 4°C for 15 minutes at 5000 *g*. The wet cells are weighed and stored overnight on ice in a cold room. The total mass of wet cells for a 12 L culture is expected to be around 150 g.

Lysis: The next day, cells are resuspended in cold S30 buffer (1.27 mL of buffer per gram of wet cells, 10x S30 buffer composition: 100 mM HEPES–KOH, pH 7.5, 600 mM $C_2H_3O_2K$ (potassium acetate); 140 mM $C_4H_6MgO_4$ (magnesium acetate), fresh DTT at 10 mM) and then disrupted using French press (only one pressure cycle). The supernatant is clarified by two centrifugation steps at 4°C in six 70 ml ultracentrifuge tubes, using a 45 Ti rotor (30,000 *g*, 30 minutes). After each centrifugation step, only the supernatant is recovered (corresponding to 80 % of the initial volume).

Maturation: The endogenous mRNA is processed by incubating the S30 extract supernatant at 42°C for 45 minutes after the addition of a 5 M NaCl solution to reach a final concentration of 400 mM.

Dialysis: Glassware (a 2 L cylinder as dialysis tank, a 500 mL one for measuring the S30 volume), S30 buffer (12 L prepared without DTT) and a 12 kDa cut-off dialysis are stored in advance at 4°C for dialysis. S30 extract is dialyzed in several steps each time against 2 L of S30 buffer at 4°C. The DTT (1 mM) is added at the beginning of each dialysis step. Two successive 1h-dialysis are performed, one overnight dialysis bath and then three extra 1h-dialysis steps the next morning. The S30 extract is centrifuged in 50 ml tubes at 5000 *g* and 4°C. 2 mL safe-lock tubes are prepared and the supernatant aliquoted in 1 mL fractions and frozen immediately in liquid nitrogen. These S30 extract aliquots can be stored for several years at -80°C. If the protocol includes addition of inhibitors to avoid isotopic scrambling, the S30 extract has to previously be incubated with a solution of $NaBH_4$[19]. As $NaBH_4$ is a strong reducing agent activated by water, the stock solution must be prepared in DMF. $NaBH_4$ attacks the aldehyde group of PLP and reduces the Schiff base occurring between PLP and NH group of transaminases. H_2 release is accompanied by the formation of foam. Work under a fume-hood. It then has to be dialyzed out twice against 1:1000 volume ratio as it can be quite harsh for the activity of other enzymes. Before assembling the mix, the S30 extract has to be incubated at least 15 minutes on ice with 20 mM each of AOA and D-malic acid.

Quality control: Expression yields are very dependent on Mg^{2+} concentration and the precise Mg^{2+} concentration already present in the S30 extract can vary from batch to batch. For each S30 extract batch, the Mg2+ concentration has to be optimized by performing expression tests, in duplicate, using a model protein. We usually perform six *in vitro* expression tests in small volumes, with additional magnesium concentration in the reaction and feeding mixes ranging from 5 to 15 mM. This optimization is routinely performed using GFP as a model protein, and we determine the optimal concentration of Mg2+ from highest fluorescence intensity.

III.3.2. Cell-free expression of proteins

Here is presented the Continuous Exchange Cell-free (CECF) reaction using a dialysis membrane system. The cut-off should be 12 kDa or less depending on the size of synthetized protein. This protocol is optimized for the large-scale expression of unlabeled proteins typically used in structural biology projects. This section describes the preparation of the reaction mix (1 mL) and the feeding mix (10 mL) required to synthesize *in vitro* the target protein. The DNA vector should be prepared in large scale (*c.a.* 500 µg) using a commercial plasmid preparation kit and eluted in RNAse free water at a concentration of 1 mg/mL. In the following protocol, all the mixtures are vortexed, except at point 6.

1. The amino acid mixtures are solubilized in three 50 mL tubes, the concentration of each amino acid being 50 mM. Alanine, Arginine, Glycine, Histidine, Lysine, Proline, Serine, Threonine, and Valine are resuspended together in water. Acidic soluble amino acids (Asparagine, Aspartate, Cysteine, Glutamine, Glutamate, Leucine, Methionine, Tryptophan and Tyrosine) are resuspended in HCl 1 M. Isoleucine and Phenylalanine are resuspended in KOH 1 M. The sources of labeled amino acids (AA) with different isotopic labeling designs can be found from many providers (CIL, Merck, Cortecnet, …). They are available as individual amino acids, 16 AA mix, or 20 AA mix ready-to-use for cell-free expression. The main advantage of individual amino acids is that users can adjust the concentration of each amino acid in accordance with the protein target sequence. Algal extract (Isogro® or Celtone®) can also be used as a cheap source of 16 labeled amino acids but C, W, N and E amino acids are degraded during Algal extract preparation and need to be added in the amino acid mix for cell free expression. Algal

extracts contain c.a. 60% of these 16 AA at different concentrations, in the form of free amino acids or small peptides. These small peptides can be enzymatically hydrolyzed by a protease (pronase from *Streptomyces griseus*) in order to be incorporated during in vitro protein synthesis. To hydrolyze 1 g of algal extract in a volume of 20 mL (Tris 20 mM pH 7.5), 33 mg of this protease is added and incubated under stirring at 37°C overnight. The next morning, a heat shock (95°C for 30 minutes) can inactivate the protease. The solution is centrifuge (5000 g, 15 minutes) and the supernatant containing the free amino acids can be recovered. Hydrolyzed algal extract should be used at a final concentration of 3-5 mg/mL in the feeding and reaction mixtures.

2. Wash extensively the dialysis Gebaflex tube with water and keep it in water during the CECF reaction preparation.
3. The amino acid solution (AA mix) is prepared by adding 247.5 µL of each amino acid mix (water-soluble, acid-soluble and base-soluble amino acids), and complemented with 82.5 µL RNase free water.
4. The 10X reaction mix containing HEPES/KOH (55 mM, pH 7.5), DTT (3.4 mM), ATP (1.2 mM), 3', 5'-cyclic AMP (0.64 mM), 0.8 mM of each CTP, GTP and UTP ribonucleotides, folinic acid (68 µM), ammonium acetate (27.5 mM) and spermidine (2 mM) is prepared.
5. The feeding mix is prepared with 1,05 mL of the 10X reaction mix, creatine phosphate at final concentration of 80 mM, Potassium Glutamate (208 mM), AA mix (1 mM for each amino acid), and magnesium acetate (14.4 mM).
6. The reaction mix is prepared as described for the feeding mix, with further addition of 250 mg/mL of creatine kinase, 175 µg/mL of total tRNA E. coli MRE600, 50 µg/mL of T7 RNApol, 40% of the final volume of S30 extract and 16 µg/mL of the target protein vector. The reaction must not be vortexed at this stage. (T7 RNA polymerase (T7RNApol) needs to be purified in RNAse free conditions at 4°C. It is an essential component of the Cell-free reaction and high concentrations are needed for the reaction. It can be purchased from different providers or produced in-house by overexpression in *E. coli* and a one-step ion exchange purification. If the T7 RNApol has a His-tag, it will co-elute with the His-Tagged target protein during affinity purification. It has to be taken into account when the purification strategy is defined.)
7. The reaction mix (1 mL) is loaded in the dialysis Gebaflex and the feeding mix (10 mL) in a 25 mL cylinder. The Gebaflex is put in the cylinder and incubated overnight with

stirring at a temperature optimal for the target protein (16 hours at 30°C here for H23 and TET2). A synthesized version for a 1:10 CECF is given on the Table 3.6 which recapitulates the different compounds that are used for the different mixes.

8. Next morning, the supernatant is recovered by centrifugating 20 minutes at 10,000 *g* at 4°C. Purification can proceed according to the standard protocol of the protein. A dilution by a factor 4 is usually required in order to decrease sample ionic strength before loading it on a purification column.

In order to reduce the amount of labeled amino acids, and reaction time, the cell-free reaction can be performed in Batch mode. Synthesis in Batch mode is performed using only the reaction mix described. Cell-free reaction by-products inhibit the protein synthesis and therefore the reaction is usually limited to 2-3 hours. This protocol reduces quantities of labeled amino acids required by a factor of 10 but also decreases the protein synthesis yield by a factor of 3-5.

Amino acid mix	**Volume (µL)**
Water soluble AA: A, G, H, K, P, R, S, V, T 50 mM concentration for each AA	247.5
In 1 M HCl : C, D, E, L, M, N, Q, W, Y 50 mM concentration for each AA	247.5
In 1 M NaOH: I, F 50 mM concentration for each AA	247.5
RNAse free H_2O	82.5
TOTAL	**825.0**

10x reaction mix	**(µL)**
100 mM rCTP	96.0
100 mM rGTP	96.0
100 mM rUTP	96.0
2.0 M HEPES-KOH pH 7.5	330.0
100 mM ATP	144.0
10 mM Folinic Acid	81.6
100 mM cyclic AMP	76.8
1 M DTT	40.8
1M spermidine	24.0
9.2 M NH_4OAc	35.9
RNAse free H_2O	178.9
TOTAL	**1200**

Reaction Mix (RM) and Feeding Mix (FM)	**RM (µl)**	**FM (µl)**
10X reaction mix	100.0	1000.0
1M Creatine Phosphate	80.0	800.0
Amino acid mix	66.7	666.7
4 M Potassium Glutamate	52.0	520.0
1.07 M $Mg(OAc)_2$	7.9	130.8
17.5 mg/ml MRE 600 tRNA	10.0	0
10 mg/ml Creatine Kinase	25.0	0

T7 RNA polymerase (1/100e)	10.0	0
Adjust pH of FM to 7.5 with KOH		
S30 extract	400.0	0
Target DNA (16 µg/mL)	16.0	0
RNAse free H_2O	232.5	6882.5
Volume (µL) of mix	**1000**	**10000**

Table 3.6: Preparation of reaction mix and feeding mix for CECF *in vitro* synthesis of protein

III.3.3. 3D HNCA recording conditions for H23 in batch and CECF modes

3D HNCA experiments were recorded on uniformly labeled H23 samples expressed either in batch mode with a 20 amino acid mix (Merk) or in CECF mode with a 16 amino acid algal extract (Celtone) and added unlabeled tryptophan. Both were expressed in H_2O, with or without the three inhibitors of amino acid scrambling treatment: $NaBH_4$, 20 mM AOA and 20 mM DM.

For the cell-free batch expression mode of H23 in H_2O, the sample conditions were as follows : Tris buffer 20 mM, pH 6.5, NaCl 100mM, recorded on a 600MHz Cryoprobe Bruker spectrometer, 25°C:

- shigemi 4 mm tube, concentration of 100µM of H23 expressed without inhibitors

- BEST-TROSY HNCA sequence parameters : TD F3 1022 F2 80 F1 208, NS 4, D1 1 sec, AQ 0.07 0.018 0.023 sec, SW 12 36 30 ppm, O1P : 4.7 118 54 ppm, RG 203

- BEST-TROSY HNCA processing parameters : SI 2048 160 416, sine bell apodisation in F3 F2 and F1, offset 0.45 end 0.98 power 2 c 0.5 ; 0.5 0.9 1; 0.5 1 1, TDeff 0

- 3 mm tube concentration of 400µM of H23 for the expression conditions with inhibitors

- BEST-TROSY HNCA sequence parameters : TD F3 1022 F2 70 F1 204, NS 8, D1 1 sec, AQ 0.07 0.016 0.022 sec, SW 12 36 30 ppm, O1P : 4.7 118 54 ppm, RG 203

- BEST-TROSY HNCA processing parameters : SI 2048 140 408, sine bell apodisation in F3 F2 and F1, offset 0.45 end 0.98 power 2 c 0.5 ; 0.5 0.9 1; 0.5 1 1, TDeff 0

For the cell-free dialysis expression mode of H23 in H_2O, the sample conditions were as follows : Tris buffer 20 mM, pH 7.5, NaCl 100mM, recorded on 850MHz Cryoprobe Bruker spectrometer, 25°C :

- shigemi 4 mm, concentration of 300µM of H23 expressed without inhibitors

- BEST-TROSY HNCA sequence parameters : TD F3 1448 F2 100 F1 300, NS 4, D1 1 sec, AQ 0.07 0.016 0.023 sec, SW 12 36 30, O1P : 4.7 118 54 ppm, RG 203

- BEST-TROSY HNCA processing parameters : SI 2896 200 600, sine bell apodisation in F3 F2 and F1, offset 0.45 end 0.98 power 2 c 0.5 ; 0.5 0.9 1; 0.5 1 1, TDeff 0

- shigemi 4 mm, concentration of 100µM of H23 for the expression conditions with the three inhibitors

- BEST-TROSY HNCA sequence parameters : TD F3 1448 F2 78 F1 300, NS 8, D1 1 sec, AQ 0.07 0.012 0.023 sec, SW 12 36 30 ppm, O1P : 4.7 118 54 ppm, RG 203

- BEST-TROSY HNCA processing parameters : SI 2896 156 600, sine bell apodisation in F3 F2 and F1, offset 0.45 end 0.98 power 2 c 0.5 ; 0.5 0.9 1; 0.5 1 1, TDeff 0

1. Markley, J. L., Putter, I. & Jardetzky, O. High-Resolution Nuclear Magnetic Resonance Spectra of Selectively Deuterated Staphylococcal Nuclease. *Science* **161,** 1249–1251 (1968).
2. LeMaster, D. M. in *Methods in Enzymology* **177,** 23–43 (Elsevier, 1989).
3. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94,** 12366–12371 (1997).
4. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR †. *Biochemistry* **36,** 1389–1401 (1997).
5. Hoogstraten, C. G. & Johnson, J. E. Metabolic labeling: Taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concepts Magn. Reson.* **32A,** 34–55 (2008).
6. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. 6
7. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression. *J. Am. Chem. Soc.* **136,** 11308–11310 (2014).
8. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. & Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angewandte Chemie International Edition* **49,** 1958–1962 (2010).
9. Religa, T. L., Ruschak, A. M., Rosenzweig, R. & Kay, L. E. Site-Directed Methyl Group Labeling as an NMR Probe of Structure and Dynamics in Supramolecular Protein Systems: Applications to the Proteasome and to the ClpP Protease. *J. Am. Chem. Soc.* **133,** 9063–9068 (2011).
10. Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P. & Boisbouvier, J. An optimized isotopic labelling strategy of isoleucine-γ 2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **48,** 1434–1436 (2012).
11. Ruschak, A. M., Velyvis, A. & Kay, L. E. A simple strategy for 13C,1H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *J Biomol NMR* **48,** 129–135 (2010).
12. Monneau, Y. R., Ishida, Y., Rossi, P., Saio, T., Tzeng, S.-R., Inouye, M. & Kalodimos, C. G. Exploiting E. coli auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *J Biomol NMR* **65,** 99–108 (2016).
13. Waugh, DavidS. Genetic tools for selective labeling of proteins with ?-15N-amino acids. *J Biomol NMR* **8,** (1996).
14. McIntosh, L. P. & Dahlquist, F. W. Biosynthetic Incorporation of ^{15} N and ^{13} C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins. *Quart. Rev. Biophys.* **23,** 1–38 (1990).
15. O'Brien, E. S., Lin, D. W., Fuglestad, B., Stetz, M. A., Gosse, T., Tommos, C. & Wand, A. J. Improving yields of deuterated, methyl labeled protein by growing in H2O. *J Biomol NMR* **71,** 263–273 (2018).
16. Löhr, F., Katsemi, V., Hartleib, J., Günther, U. & Rüterjans, H. A strategy to obtain backbone resonance assignments of deuterated proteins in the presence of incomplete amide 2H/1H back-exchange. 21
17. Ozawa, K., Headlam, M. J., Schaeffer, P. M., Henderson, B. R., Dixon, N. E. & Otting, G. Optimization of an Escherichia coli system for cell-free synthesis of selectively 15N-labelled proteins for rapid analysis by NMR spectroscopy: Cell-free synthesis of 15N-labelled proteins. *European Journal of Biochemistry* **271,** 4084–4093 (2004).
18. Tonelli, M., Singarapu, K. K., Makino, S., Sahu, S. C., Matsubara, Y., Endo, Y., Kainosho, M. & Markley, J. L. Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR* **51,** 467–476 (2011).
19. Su, X.-C., Loh, C.-T., Qi, R. & Otting, G. Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective 15N-labelling and production of perdeuterated proteins in H2O. *J Biomol NMR* **50,** 35–42 (2011).
20. Zubay, G. In Vitro Synthesis of Protein in Microbial Systems. *Annual Review of Genetics* **7,** 267–287 (1973).
21. Roberts, B. E. & Paterson, B. M. Efficient Translation of Tobacco Mosaic Virus RNA and Rabbit Globin 9S RNA in a Cell-Free System from Commercial Wheat Germ. *Proceedings of the National Academy of Sciences* **70,** 2330–2334 (1973).
22. Sobhanifar, S., Reckel, S., Junge, F., Schwarz, D., Kai, L., Karbyshev, M., Löhr, F., Bernhard, F. & Dötsch, V. Cell-free expression and stable isotope labelling strategies for membrane proteins. *J Biomol NMR* **46,** 33–43 (2010).
23. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnology Advances* **30,** 1185–1194 (2012).

24. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by Wild-Type and Mutant T7 RNA Polymerases: Identification of Interactions That Reduce Misincorporation Rates by Stabilizing the Catalytically Incompetent Open Conformation †. *Biochemistry* **39,** 11571–11580 (2000).
25. Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *International Journal of Biological Macromolecules* **118,** 49–56 (2018).
26. *Heterologous Expression of Membrane Proteins*. **601,** (Humana Press, 2010).
27. Pedersen, A., Hellberg, K., Enberg, J. & Karlsson, B. G. Rational improvement of cell-free protein synthesis. *New Biotechnology* **28,** 218–224 (2011).
28. Foshag, D., Henrich, E., Hiller, E., Schäfer, M., Kerger, C., Burger-Kentischer, A., Diaz-Moreno, I., García-Mauriño, S. M., Dötsch, V., Rupp, S. & Bernhard, F. The E. coli S30 lysate proteome: A prototype for cell-free protein production. *New Biotechnology* **40,** 245–260 (2018).
29. Fujiwara, K. & Doi, N. Biochemical Preparation of Cell Extract for Cell-Free Protein Synthesis without Physical Disruption. *PLOS ONE* **11,** e0154614 (2016).
30. Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A. & Yokoyama, S. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics* **5,** 63–68 (2004).
31. Schwarz, D., Junge, F., Durst, F., Frölich, N., Schneider, B., Reckel, S., Sobhanifar, S., Dötsch, V. & Bernhard, F. Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nature Protocols* **2,** 2945–2957 (2007).
32. Shrestha, P., Holland, T. M. & Bundy, B. C. Streamlined extract preparation for *Escherichia coli* -based cell-free protein synthesis by sonication or bead vortex mixing. *BioTechniques* **53,** 163–174 (2012).
33. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. in *Structural Proteomics: High-Throughput Methods* (eds. Kobe, B., Guss, M. & Huber, T.) 257–268 (Humana Press, 2008). doi:10.1007/978-1-60327-058-8_16
34. Kim, T.-W., Keum, J.-W., Oh, I.-S., Choi, C.-Y., Park, C.-G. & Kim, D.-M. Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of Biotechnology* **126,** 554–561 (2006).
35. Hoffmann, B., Löhr, F., Laguerre, A., Bernhard, F. & Dötsch, V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Progress in Nuclear Magnetic Resonance Spectroscopy* **105,** 1–22 (2018).
36. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1814,** 1407–1418 (2011).
37. Beeler, T. & Churchich, J. E. Reactivity of the phosphopyridoxal groups of cystathionase. *J. Biol. Chem.* **251,** 5267–5271 (1976).
38. Yokoyama, J., Matsuda, T., Koshiba, S. & Kigawa, T. An economical method for producing stable-isotope labeled proteins by the E. coli cell-free system. *J Biomol NMR* **48,** 193–201 (2010).
39. Falzone, C. J., Karsten, W. E., Conley, J. D. & Viola, R. E. L-Aspartase from Escherichia coli: substrate specificity and role of divalent metal ions. *Biochemistry* **27,** 9089–9093 (1988).
40. Mayerhofer, H., Sautron, E., Rolland, N., Catty, P., Seigneurin-Berny, D., Pebay-Peyroula, E. & Ravaud, S. Structural Insights into the Nucleotide-Binding Domains of the P1B-type ATPases HMA6 and HMA8 from Arabidopsis thaliana. *PLOS ONE* **11,** e0165666 (2016).
41. Appolaire, A., Colombo, M., Basbous, H., Gabel, F., Girard, E. & Franzetti, B. TET peptidases: A family of tetrahedral complexes conserved in prokaryotes. *Biochimie* **122,** 188–196 (2016).
42. Macek, P., Kerfah, R., Erba, E. B., Crublet, E., Moriscot, C., Schoehn, G., Amero, C. & Boisbouvier, J. Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci. Adv.* **3,** e1601601 (2017).
43. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6,** 277–293 (1995).
44. Etezady-Esfarjani, T., Hiller, S., Villalba, C. & Wüthrich, K. Cell-free protein synthesis of perdeuterated proteins for NMR studies. *J Biomol NMR* **39,** 229–238 (2007).
45. Olins, P. O. & Rangwala, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. *J. Biol. Chem.* **264,** 16973–16976 (1989).
46. Rogé, J. & Betton, J.-M. Use of pIVEX plasmids for protein overproduction in Escherichia coli. *Microb Cell Fact* **4,** 18 (2005).

Chapter IV : Optimization of the cell free expression for the specific labeling of Gly residues in proteins and application in NMR Spectroscopy

H23 expression was undertaken at the cell-free platform of the IBS Grenoble, in collaboration with Lionel Imbert and Jérôme Boisbouvier. PPIase expression was undertaken at the Laboratoire des Biomolécules, Sorbonne Université Paris, with the help of Soha Abou Ibrahim. The ligands used for titration experiments were synthetized at the Laboratoire de Chimie Biologique (LCB) from Université de Cergy Pontoise.

IV.1. Introduction

NMR greatly benefits from Transverse Relaxation Optimized (TROSY) experiments, which correspond to spin state selective spectra editing only the slowest relaxing components. The TROSY pulse schemes allow the acquisition of well-resolved and intense resonances compared to classic experiments, especially for large biomolecular systems. Moreover, spin state selective experiments derived from the TROSY NMR allow the measurements of precise scalar and dipolar coupling constants. TROSY have been described for NH amide groups,[1,2] aromatic CH pairs,[3–5] CH_2 methylene groups,[6] CH_3 methyl groups,[7,8] and more recently, CF_3 trifluorinated methyl groups.[9] Coupled to the development of new isotopic labeling methods for obtaining high level of deuteration[10–14], these TROSY experiments have pushed forward the frontiers of the liquid state biomolecular NMR. By specifically labeling methyls with ^{1}H in isoleucine, valine and leucine side chains using ^{13}C α-ketoacid precursors in *E.coli*[15–21], well resolved ^{1}H-^{13}C spectra of methyl groups have thus been obtained for proteins over half a mega Dalton[22,23]. Thus, specific and robust labeling strategies have been a focus of researchers in order to take full advantage of the TROSY experiments. However, the quality of the TROSY experiments rapidly degrades for even weak levels of residual protonation, since neighboring protons considerably reduce relaxation times of the observed spin systems by creating dipolar interactions. Many routes have been described for obtaining very specific protonation in highly deuterated background for ^{1}H-^{15}N or the methyl TROSY. Among all the TROSY experiments developed, one of the least exploited is the CH_2-TROSY[6], even though 46% of protons in proteins are in methylene groups. When it was first published, the resolution gains for the CH_2-TROSY, compared to a regular Rance-Kay HSQC, averaged to 1.6 for the residues of the IgG-binding domain from Streptococcal Protein G (a small protein of 61 residues, 6.641 kDa) at 10 °C and 800 MHz. In these conditions, no

improvement was observed on average in terms of sensitivity but a closer inspection of the data revealed that spectral characteristics differed considerably depending on the residue type. Indeed, whereas the poorest gains were obtained for the C^{β} methylene group of the methionine, glutamine, glutamate, lysine and arginine residues, for which the CH_2-TROSY experiment was frequently less sensitive than the regular HSQC, the highest gains were attained for the C^{α} methylene group of glycine residues. This was expected since the methylene of glycine residues is isolated from proton-rich side chains that will greatly affect relaxation through dipole-dipole interactions. This part of the manuscript focuses on the use of the chemical properties of glycine, where pairs of diastereotopic protons are naturally isolated, coupled to their specific labeling in a perdeuterated background, to fully exploit the CH_2-TROSY experiment. Such a sample was produced by cell-free, as it prevented profuse amino acid isotopic scrambling from the glycine residue and insured high deuteration levels.

To fully appreciate the value of the CH_2-TROSY experiment, and quantitatively evaluate the importance of a completely perdeuterated sample, different levels of protein isotopic labeling were tested by decreasing the number of protons neighboring the methylene. This was easily achieved by using either a H_2O or D_2O based solvent, and a uniformly or specifically labeled sample. The utility of the CH_2 TROSY experiment for biological investigation is then demonstrated for a titration experiment using the uniformly labeled catalytic domain of a Peptidyl-Prolyl isomerase, Pin1, using a novel fluorinated peptide ligand. Whereas glycine-specifically labeled samples would undoubtedly be required for titration of larger systems, the CH_2-TROSY experiment gave substantial spectral improvements for this medium-sized protein, using uniformly ^{13}C-labeled sample in D_2O. Moreover, implementation of the CH_2-TROSY pulse scheme into three-dimensional spin state selective experiments allowed the expression of a number of coupling constants in glycine. Their measurements can offer much precise information about the structure and dynamic of the residue, with glycine acting as a local, information rich probe about the protein conformation.

To understand the choice of glycine as the model for CH_2-TROSY optimization, one has to review the particularities of this residue. Glycine is one of the most abundant amino acid in proteins with an occurrence of 7.76%[24,25]. It is the smallest amino acid and the only one that is achiral (figure 4.1). When involved in a peptide sequence, it has no charge, is nonpolar and

possesses two diastereotopic protons Hα2 (pro*R*) and Hα3 (pro*S*) directly linked to the peptide backbone.

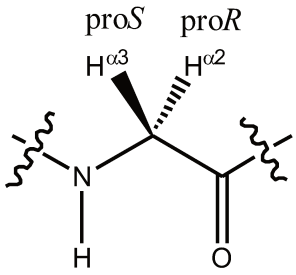


Figure 4.1: Chemical illustration of the Glycine residue.

The interatomic distances within the glycine residue are $d_{C'=O}$ = 1. 25 Å $d_{C\alpha-C'}$ = 1.52 Å ,$d_{C\alpha-N}$ = 1.39 Å , $d_{H\alpha2-H\alpha3}$ = 1.57 Å and $d_{C\alpha-H\alpha}$ = 0.97 Å . The angle between the Cα and the two protons is close to tetrahedral (109°28')[26], with a coupling constant $^1J_{C\alpha H\alpha}$ of usually around 140 Hz[27] for amino acids. As glycine contains a hydrogen in lieu of a side chain, its φ and ψ angles are less limited than for other residues in proteins, its φ angle capable of populating the positive part of the Ramachandran plot, in which are found left handed helices. Thus, it plays a unique role in the secondary structures of proteins such as loops[28] and turns where it is the most prevalent residue, in particular at the i+2 position of β-turn[47]. Thus, glycine often contributes to the rupture of secondary structure elements. It also frequently found in short regions rich in serine and glycine often corresponding to flexible linkers. It is much less frequent in alpha helixes and beta-sheets (0.56% and 0.92% respectively[29]) although it is present in several membrane spanning helices[30,31] participating in glycine-glycine packing interactions. It is the most abundant residue in collagen (33%). Interestingly, glycine residues are often found at ligand binding sites[32] and having direct contact with the ligand. Exploiting these properties, we show that glycines, when selectively labeled, can act as accurate and unique probes for the study of biological processes using the aforementioned CH_2-TROSY or specific 3D experiments taking advantage of the glycine specific topology.

IV.2. Challenges for the specific labeling of Gly residues in proteins.

IV.2.1. Limitations *in vivo*

There has been to our knowledge no expression protocol described to selectively label the two diastereotopic protons of glycine in a completely perdeuterated background. This is probably because of the difficulty of achieving such a labeling strategy *in vivo* due to the

numerous enzymes involved in amino acid metabolism present in cells[33], one would have to supplement a deuterated M9 medium with glycine (^{13}C-Gly, ^{13}C-^{15}N-Gly or unlabeled Glycine). However, the cell organism would also synthetize fresh glycine with undesired isotopic patterns. The use of *E.coli* auxotrophic strains for glycine[34] could be imagined, but there have been no reports to our knowledge of their use for labeling strategies as there have been for other amino acids[35]. The culture could be also performed using algal extracts as a source of deuterated amino acids and large excess of protonated glycine. However, the isotopes specifically introduced on this residue would most probably diffuse to other amino acids through diverse metabolic pathways. The medium would need to be D_2O based, otherwise protons from the solvent would be incorporated on H$\alpha 2^{proR}$ positions of the other residues due to the transaminase activity. But conversely, in D_2O the transaminase activity would lead to incorporation of deuterium on the H$\alpha 2^{proR}$ position of glycine.

IV.2.2. Limitations in cell free

Cell free expression gives the experimenter more control over the production environment than cell-based techniques. The amino acids have to be incorporated for expression to take place, and isotopically labeled amino acids can be added from algal extracts or individually for the desired labeling scheme. As developed in the previous chapter, there are a number of enzymes involved in metabolic pathways whose presence in the cell-free extract can lead to isotopic scrambling but the use of specific inhibitors can limit their activity. For the optimization of our labeling strategy, we started with a mix of individual unlabeled amino acids and an excess of glycine (^{1}H-^{15}N-^{13}C) labeled. Transaminase activity exchanges amide groups between keto-acids and other residues, *in fine* diluting the specific labeled nitrogen (upper panel figure 4.2). If deuterated algal extracts U-(^{2}H-^{15}N-^{12}C) are used as amino acid source, the transaminase activity would result in inserting a proton on the Cα pro*R* positions of all amino acids when the expression is conducted in H_2O (middle panel figure 4.2), or in producing CHD methylene groups in glycines when the solvent is deuterated. The figure 4.2 summarizes the label scrambling due to transaminases activity.

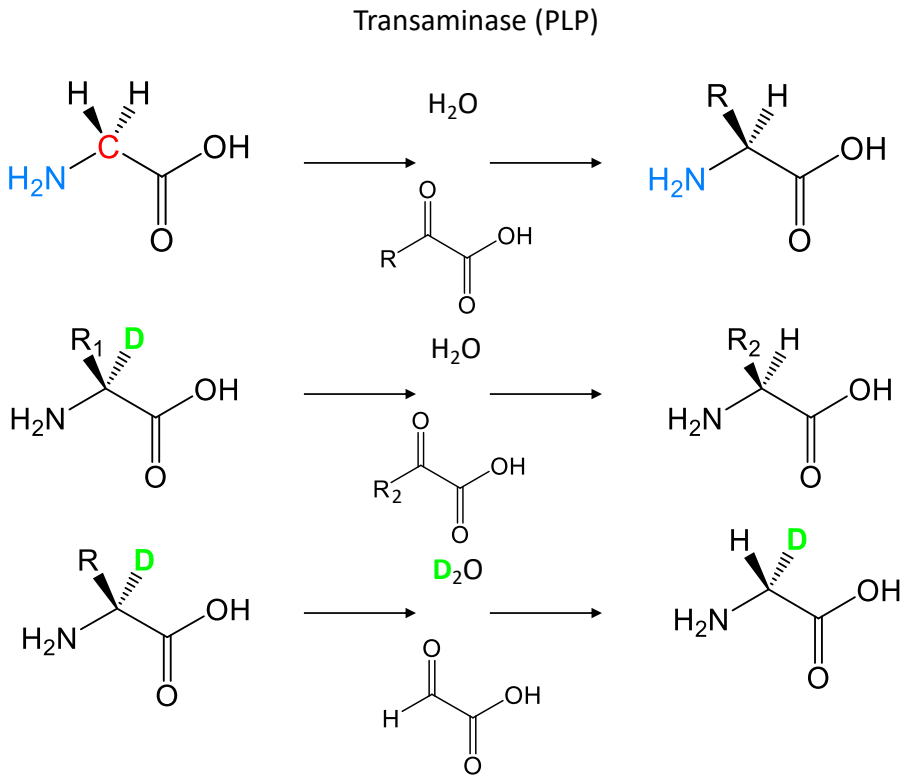


Figure 4.2: Schematic overview of the isotope scrambling resulting from the transaminase activity. ^{13}C, ^{15}N and ^{2}H isotopes are displayed in red, blue and green, respectively.

The glycine residue is involved in numerous other metabolic pathways[36], acting both as substrate and product for a number of enzymatic reactions. Threonine aldolase (ItaE) catalyzes the reversible reaction of L-threonine into glycine and acetaldehyde, by means of the cofactor PLP. Glycine, by the intermediary of serine, can be converted into pyruvate for supplying the TCA cycle. Of the enzymes listed previously (Chapter III), the activity of serinehydroxymethyl transferase (glyA) is looked at in more detail here. GlyA, also a PLP dependent enzyme, is responsible for the catalysis of two coupled reactions: the reversible conversion of L-serine to L-glycine and the reversible conversion of tetrahydrofolate to 5,10-methylene-tetrahydrofolate[37]. The cofactor, Pyridoxal-5'-phosphate, is originally bound to an amine of the serine by a covalent Schiff base linkage. The Glu 57 residue of the glyA enzyme removes a proton of the hydroxyl group of the serine residue which leads to liberation of a formaldehyde. In a second step, the carbon of the iminium group can catch a hydrogen on a water molecule, thanks to the nucleophilic assistance of the PLP moieties leading to a PLP-glycine complex (Figure 4.3). The proton (or deuterium) from the water becomes the Hα3 of the glycine.

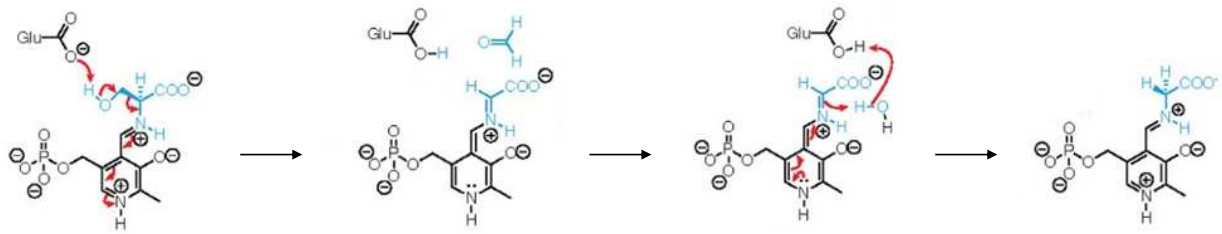


Figure 4.3: Main steps of the mechanism of action of serine hydroxy methyltransferase on the PLP-serine adduct, taken from Fernandes et al., 2018[37]

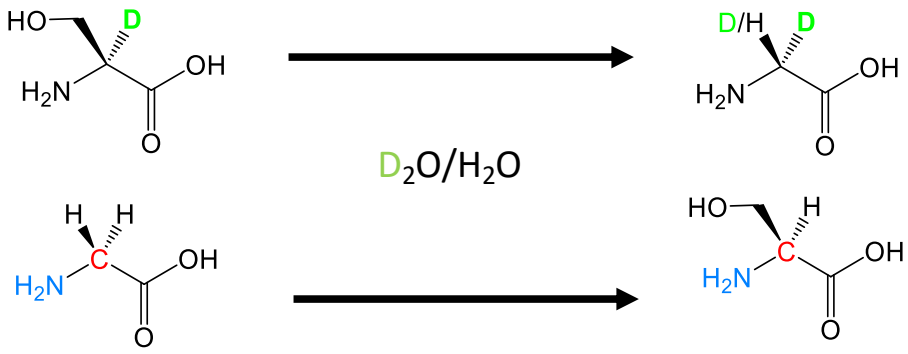


Figure 4.4: Schematic overview of the potential isotopic scrambling resulting from the serine hydroxy methyltransferase activity. ^{15}N, ^{13}C and ^{2}H isotopes are colored in blue, red and green, respectively.

Glycine is the only amino acid that possesses the ^{15}N and ^{13}C isotopes in the fresh expression solution, while other amino acids are deuterated. But isotopic scrambling may occur due to residual activity of glyA in the S30 extract. glyA activity results in the formation of glycine with undesired isotopes. When the hydroxymethyl of serine residue is removed, the obtained glycine lacks ^{15}N and ^{13}C isotopes. Moreover, the glycine produced do not possess diastereotopic protons but a $C^{\alpha}HD$ or a $C^{\alpha}D_2$ group, depending on the solvent used for the expression (Figure 4.4). The reverse glyA reaction produces ^{15}N-^{13}C- isotopically labeled serine, protonated on the Ca, thus undermining selective isotopic labeling efforts (Figure 4.4). This example demonstrates how residual enzymatic activity involved in glycine metabolism may lead to isotopic scrambling.

Many more enzymes could have such effects during the expression. One interesting characteristic of the two enzymes described previously, namely transaminases and serine hydroxymethyl transferase, is that they both require PLP as a cofactor. PLP dependent enzymatic activity can be selectively limited through the use of inhibitors. Thus, to evaluate the role of the transaminases and the serine hydroxymethyl transferase in the isotopic scrambling, we propose to specifically inactivate this class of enzyme and to compare the isotopic distributions when the expression is performed in the absence of inhibitors.

The H23 protein was expressed in a similar fashion than the previously used cell-free protocol (Chapter III) but using isotopically labeled glycine in a ten-fold excess compared to the other amino acids from individual powders or algal extracts. Optimization was undertaken varying the purification of the S30 extract or not, the use of the three inhibitors of PLP-dependent enzymes ($NaBH_4$ treatment of the S30 extract followed by incubation with 20 mM AOA and 20 mM DM before use), and using a H_2O or D_2O based buffers. The cell-free protocols are detailed in section IV.4.1. Materials and Methods. An analysis of the expressed protein for each different protocol was undertaken.

IV.2.3. Optimization of the Gly labeling protocol using the H23 protein

A critical problem faced was protonation of other amino acids than glycine. Protonation of Cα sites was quantified when cell-free expression takes place in H_2O in the previous Chapter (III) and represented less than 5% average protonation with the use of the three inhibitors, 20 mM AOA, 20 mM DM and $NaBH_4$ treatment.

H23 possesses 15 glycines (see Chapter II for the complete sequence) of which 4 are part of the His-tag. After cell-free expression of the H23 protein with ^{15}N-^{13}C-^{1}H labeled glycine, unlabeled amino acids added individually from pure powders and the three inhibitors AOA, DM and $NaBH_4$, ^{1}H-^{15}N HSQC spectra were recorded. Only a few correlations were observed that correspond to Glycine residues (assignment determined in chapter II). Eight glycines were absent on the spectra recorded at pH 9 but some of them were recovered at a lower pH, presumably by slowing down the amide proton exchange (Figure 4.5). G144 and G74 appeared, as well as an unresolved cluster of correlations that could correspond to the glycines belonging to the highly flexible His-tag. No other minor correlations were observed at lower contour levels indicating that a satisfactory specific labeling of the glycine residues can be achieved using these conditions.

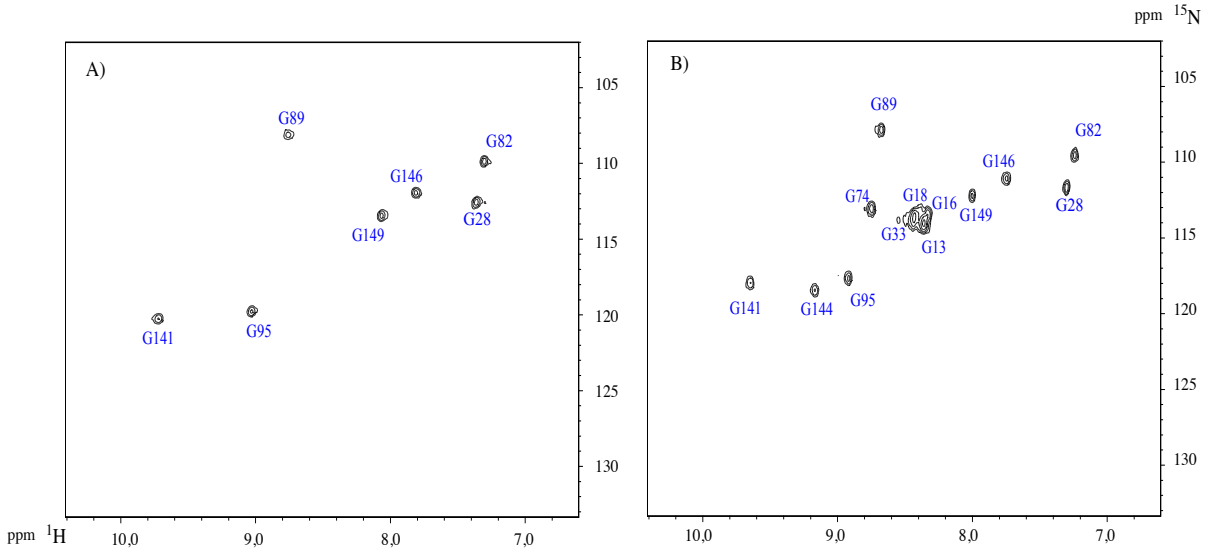


Figure 4.5: ^{1}H-^{15}N HSQC spectra of H23 protein using a specifically ^{15}N, ^{13}C, ^{1}H-Gly labeled sample and no isotopic label on the other amino acids (^{14}N, ^{12}C, ^{1}H). Protein expression made by cell-free. A) pH : 9. B) pH 4,5. Spectra recorded with a protein concentration of 200 µM in 5mm shigemi tubes, Tris HCl buffer 20 mM, NaCl 100 mM, 0,11 mM DSS, 10% D_2O, 40°C, , 500 MHz spectrometer equipped with a TCI cryoprobe. Recording conditions are detailed in section IV.4.2. Materials and Methods.

In a second assay, a deuterated algal extract was used as a source of amino acids. Again, the glycine was introduced in large excess with ^{1}H, ^{13}C, ^{15}N isotopes. Expressions were performed with/without inhibitors and ^{1}H-^{15}N HSQC spectra enabled us to directly observe the improvements to specific glycine labeling allowed by cell-free expression. The figure 4.6 compares three different expressions, highlighting the importance of inhibitor use in the expression medium. Absence of inhibitors led to unspecific labeling of other amino acids than glycine. Weak correlations visible on the spectrum suggest that these are other amino acids that received a ^{1}H-^{15}N label after isotopic scrambling. Assignment of H23 (see Chapter II) enabled us to ascertain that these weak correlations were seven serines, although there are 18 in the H23 sequence (see Chapter II), but nine of which are in extremely flexible structures (the His-tag and loops). Their labeling was most probably the result of glyA activity. However, a threonine had also been assigned, suggesting there is probably activity of other active enzymes, perhaps the threonine aldolase. Use of three inhibitors AOA, DM and $NaBH_4$ led to disappearance of these weak serine and threonine correlations. Importantly, the threonine aldolase that interconvert threonine and glycine is also PLP dependent and could be inhibited by the inhibitor cocktail that we used. However, complete and exclusive labeling of glycine residues could not be confirmed at more than 98% (peak intensities less than 2% of the highest glycine peak would not be distinguished from the noise). Addition of PLP inhibitors was deemed crucial for

preventing major undesired amino acid labeling, although a minor residual isotopic scrambling by non PLP dependent enzymes was still possible.

In order to quantify the amount of labeled serine, we measured the integral of the correlation of glycine 141 which is involved in the strand β5 and not experiencing fast proton exchange. Comparison with the integration of three weak correlations (serine 64, serine 118 and serine 128) gave ratios of 8 to 12% (see Graph 4.1). This suggests that there is about a 10% leak of labeled glycine to serine and that, conversely, there could be a leak of unlabeled serine into glycine. ^{15}N labeling of the glycine residues could be therefore decreased due to the glyA activity.

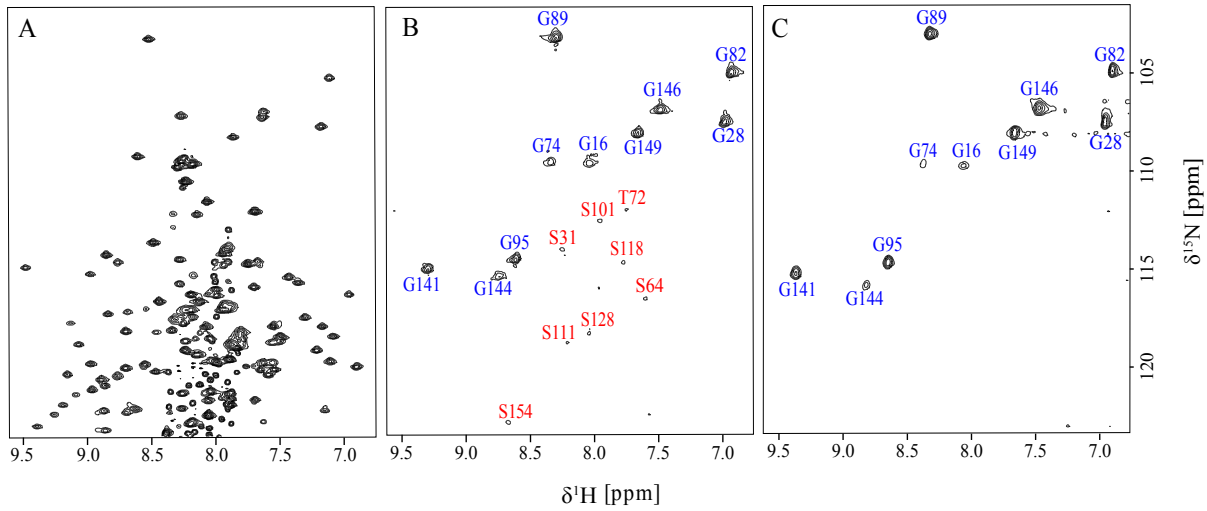
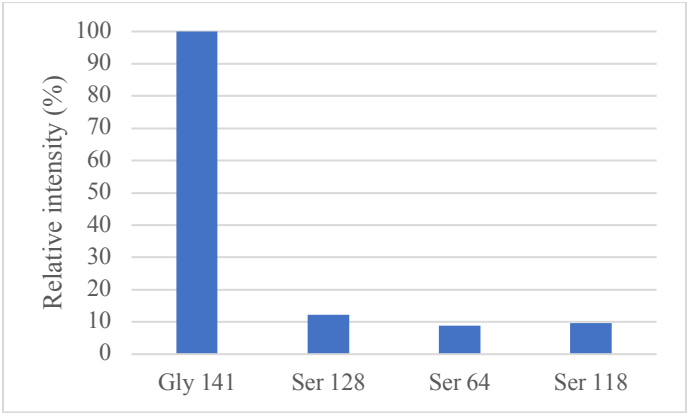


Figure 4.6 : ^{1}H-^{15}N HSQC spectra of H23 protein expressed by cell free. A) Uniformly ^{15}N labeled sample. Data acquired at 700 MHz, pH 4.5, 25°C B) ^{1}H, ^{15}N, ^{13}C-Gly and other amino acids bearing deuterium labels (^{2}H-^{14}N-^{12}C), expressed without inhibitors. C) ^{1}H, ^{15}N, ^{13}C-Gly labeled and other amino acids ^{2}H-^{14}N-^{12}C, expressed with 3 inhibitors (S30 extract 20 mM DM, 20 mM AOA and $NaBH_4$ treatment). Spectra B and C were recorded at a protein concentration of 300 µM in shigemi 3 mm, Tris buffer 20 mM pH 6.5, NaCl 100 mM, 0,11 mM DSS, 10% D_2O. 40°C. Data acquired at 500 MHz on a Bruker spectrometer equipped with a TCI cryoprobe,



Graph 4.1: Intensity of S64, S118 and S128 normalized to the intensity of G141 using the Topspin integral module.

Abundance of ^{14}N isotopes in Gly could be estimated from 1D proton spectra of different samples: ^{15}N-Gly should correspond to doublet in the amide region because of the $^{1}J_{NH}$ couplings, whereas singlets are obtained for ^{14}N-Gly. As a control, adding a ^{15}N decoupling sequence suppress the doublet due to the $^{1}J_{NH}$ couplings. The resonance of G141 appeared at 9,7 ppm (see Figure 4.6) and was well isolated from other resonances. For samples produced without inhibitor, the 1D proton spectrum consists of a singlet resonance that corresponds to ^{1}H-^{14}N pairs is observable in between the doublet of the ^{1}H-^{15}N Gly141 (Figure 4.7). The integral of this component is estimated to represent 15% of the total integral of the doublet. This suggests that about 15% of glycines of the protein are ^{14}N when no inhibitors are used in the reaction mixture. This incorporation of unlabeled nitrogen may come from serine and the glyA activity as specifically described in this section but also from any other amino acids through transaminase activity. For samples produced with the presence of three inhibitors, the integral of the component represents less than 5% of the total integral, threshold for which signal is undiscernible from noise. Again, this suggests that addition of the 3 inhibitors of PLP-dependent enzymes blocks the mechanisms responsible for decreasing the ^{15}N label level of glycine residues. This means that presence of amino acid metabolism enzymes is responsible for up to 15% of unlabeled glycine, and use of the three inhibitors blocked these mechanisms almost entirely. Interestingly, when only one inhibitor was used, 20 mM AOA, the component in between the doublet was still present but represented a lesser percentage of the total integral (9%) than when no inhibitor was used. Adding only one inhibitor was then insufficient to completely remove ^{14}N glycine presence.

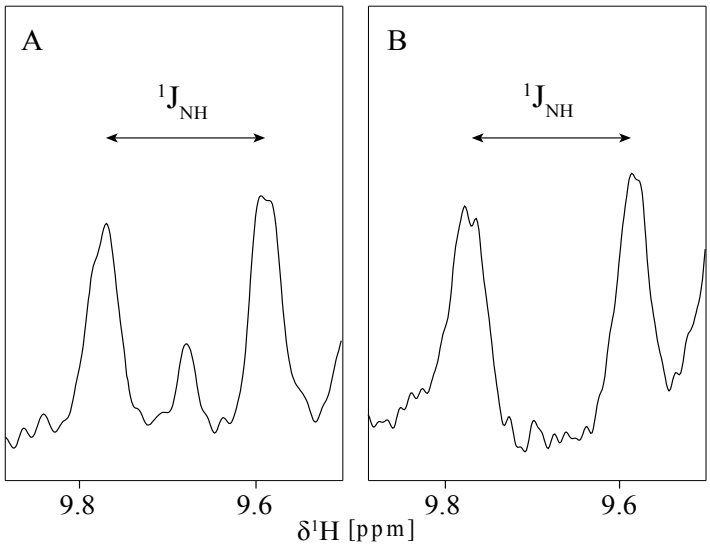


Figure 4.7: 1D spectra of H23 protein produced in H_2O with ^{1}H, ^{15}N, ^{13}C -Gly and unlabeled other amino acids (^{1}H-^{14}N-^{12}C) , A) Without inhibitors and B) with three inhibitors 20 mM AOA, 20 mM DM and $NaBH_4$ treatment. Protein concentration 200 µM in shigemi tubes 5mm, Tris HCl buffer 20 mM, pH 9, NaCl 100 mM, 0,11 mM DSS, 10% D_2O, 40°C, 500 MHz.

It cannot be excluded that some of the unlabeled glycine incorporated into the H23 protein came from the glycines already present in the S30 extract and not from the activity of scrambling enzymes. To explore this possibility, we would need to compare samples produced in a medium in which the S30 has been purified of its amino acids, using desalting columns. However, purifying the S30 extract on a desalting column affects greatly the yields obtained (see section IV.4.1. Material and Methods) which reverberates on the quality of spectra and signal/noise ratio. All samples produced with purified S30 were un-exploitable due to the limited sample concentration and this treatment was therefore abandoned.

With this data, we can therefore ascertain that sufficient glycines are ^{15}N labeled and that without inhibitor use, some other amino acids (mostly serine) can also be ^{15}N labeled. To pursue our analysis, we were interested in knowing whether some isotopic scrambling occurred on carbon nuclei. ^{15}N-^{12}C-Gly could be obtained with the combined action of transaminases and the serine hydroxymethyl transferase, or transaminases and the threonine aldolase. Overlay of the 1D ^{15}N traces of very well resolved ^{1}H-^{15}N HSQC spectra of samples produced with and without the three inhibitors (Figure 4.8) shows a doublet characterizing the ^{15}N-^{13}Ca coupling. There was no visible component in between the doublet that would account for the presence of a ^{15}N linked to a ^{12}C nucleus. However, the $^{1}J_{CN}$ coupling is weak (11 Hz) and the observation of a resolved central singlet in between the two components of the ^{15}N-^{13}Ca doublet is unlikely, but spectra differences have been performed (with and without inhibitor) which were undiscernible from the noise. This suggests that, at this signal to noise ratio, no detectable ^{15}N-^{12}C-Gly were present in the H23 sequence, even in the absence of inhibitor.

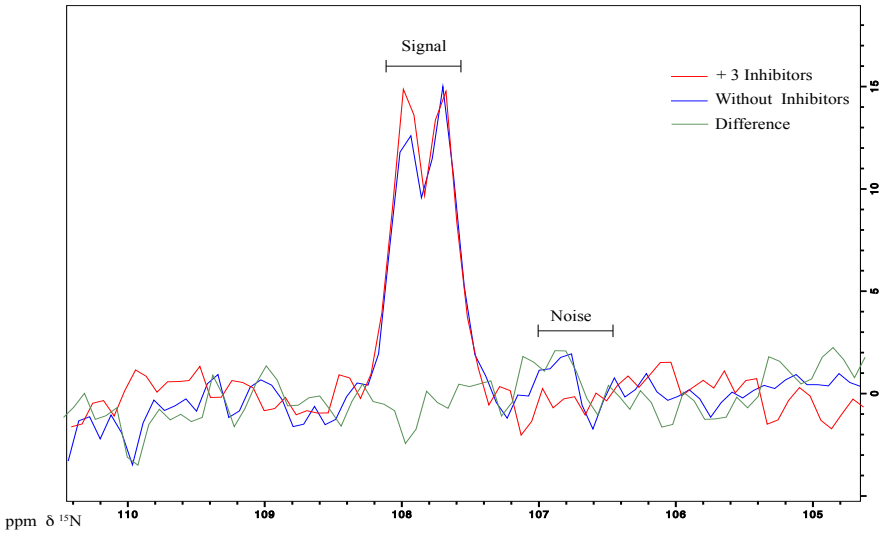


Figure 4.8: overlay of ^{15}N 1D traces of Gly 149 of H23 in ^{1}H-^{15}N HSQC spectra. Gly ^{1}H-^{15}N-^{13}C and other AA ^{1}H-^{14}N-^{12}C labeled produced in H_2O without inhibitors (blue) and with 3 inhibitors 20 mM AOA, 20 mM DM and $NaBH_4$ treatment (red). 200 µM, pH: 8.5, 40°C, 10% D_2O, 0,11 mM DSS, shigemi 5mm, 500 MHz

When ^{1}H-^{13}C HSQC was recorded on this sample (^{1}H, ^{15}N, ^{13}C-Gly and other amino acids unlabeled, *ie* ^{1}H, ^{14}N, ^{12}C), produced without inhibitors in H_2O, intense correlations were visible in the glycines region between 3.5 and 4.5 ppm in the ^{1}H dimension and 43 and 47 ppm in the ^{13}C dimension (see Figure 4.9). These are distinguishable from other carbon groups because the correlations are doublets of doublets, the greater coupling constant corresponding to the $^{1}J_{C\alpha\text{-}C'}$ (55 Hz) and the smaller one corresponding to the $^{1}J_{N\text{-}C\alpha}$ coupling (11 Hz). This spectrum does not allow us to assign all the glycine residues present in the protein sequence because of overlapping signals. The less intense correlations correspond to all the other ^{13}C present on the H23 protein at natural abundance, including the methyl groups that are well recognized in the shielded proton and carbon dimensions. None of these correlations displayed a hyperfine structure. Using these labeling conditions, ^{1}H-$^{13}C_{\alpha}$ correlations of nonGly amino acids were too weak to observe $^{1}J_{N\text{-}C\alpha}$ doublet that would result from the transaminase activity. Such a transaminase activity that takes place in the absence of inhibitors was already demonstrated in the previous chapter using HNCA experiments (see Chapter III.2.).

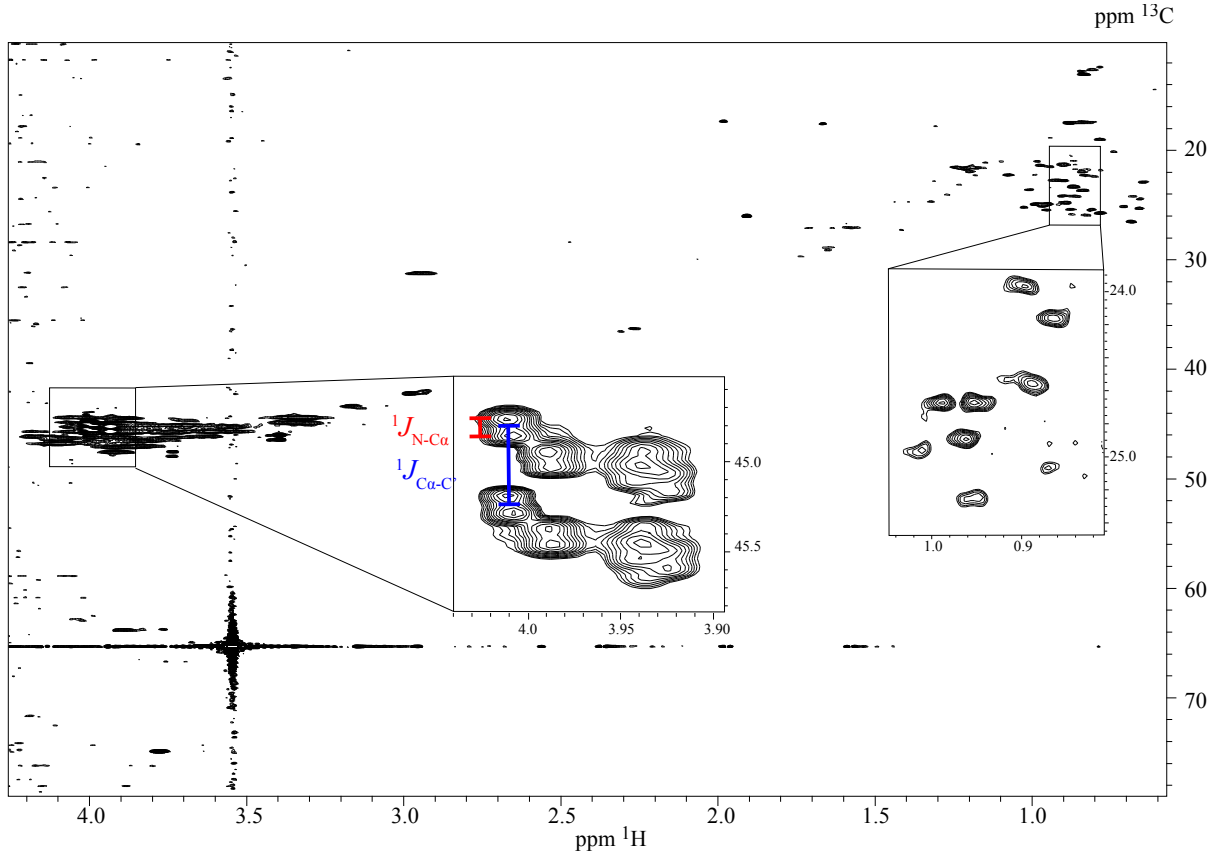


Figure 4.9 : ^{1}H-^{13}C HSQC spectrum of H23 protein Gly ^{1}H-^{15}N-^{13}C and other AA ^{1}H-^{14}N-^{12}C labeled produced in H_2O without inhibitors. Protein concentration 200 µM in a 5 mm shigemi tube, 25°C, Tris buffer 20 mM, pH : 8.5, NaCl 100 mM, 100% D_2O, 0,11 mM DSS, 500 MHz, TCI cryoprobe.

To confirm that the ^{1}H-^{13}C correlations observed for non-Gly amino acids in the previous spectrum were not due conversions of the glycine residue, a H23 sample was produced in D_2O using ^{1}H, ^{15}N, ^{13}C-Gly and ^{2}H, ^{14}N, ^{12}C isotopes for the other amino acids. In the corresponding ^{1}H-^{13}C HSQC spectrum, glycine correlations were very intense while most of the correlation of other amino acids disappeared, including the methyl groups (Figure 4.10). Despite a ^{15}N decoupling during the ^{13}C labeling, most of the 15 glycine residues of H23 were superimposed, revealing the need of a more resolved experiment for the study of these methylene groups. In addition, weak signals corresponding to $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ and possibly $^{1}H^{\beta}$-$^{13}C^{\beta}$ correlations of serine residues were visible when the sample was produced without any inhibitors. These correlations disappeared for the sample expressed in the presence of AOA only, suggesting that its use inhibited the glyA enzyme responsible for the production of $^{13}C^{\alpha}$ serine from the isotopically labeled glycine residues, although this does not explain the inhibition of the $^{13}C^{\beta}$ label.

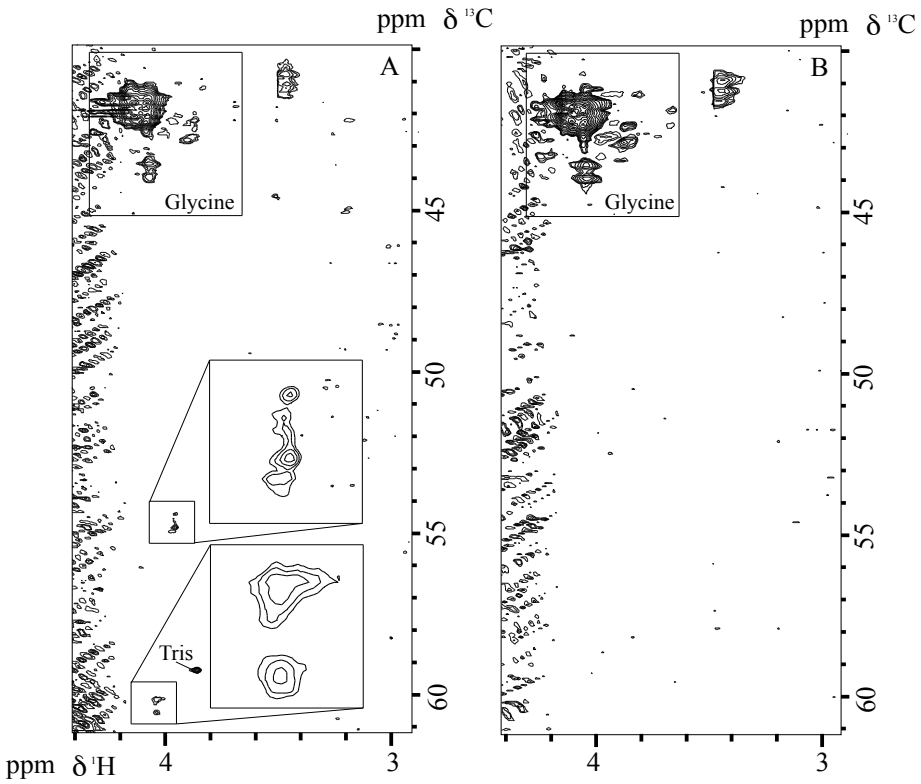


Figure 4.10: ^{1}H-^{13}C HSQC ^{15}N-decoupled of H23 produced with ^{1}H, ^{15}N, ^{13}C-Gly and ^{2}H, ^{14}N, ^{12}C-nonGly in D_2O without inhibitors (panel A) and with 20 mM AOA inhibitor (panel B). Protein concentrations: 170 µM (A) or 275 µM (B), pH 7, 40°C, 10% D_2O, 0.11 mM DSS, 3 mm tube, cryoprobe 500 MHz.

Regarding the transaminase activity, the inspection of the 1D traces in the ^{13}C dimension may reveal the existence of ^{14}N, ^{13}C-Glycine residues, if any. As mentioned before, the expected doublet of doublet was observed for the H23 glycine residues attesting the presence

of $^{1}J_{N\text{-}C\alpha}$ and $^{1}J_{C\alpha\text{-}C'}$ couplings (Figure 4.11). No component in the middle of the small doublets was detected that would correspond to ^{14}N, ^{13}C-Glycine residues. It is important to stress that, the transaminase activity would have severe impacts on the ^{1}H-^{13}C spectrum of the H23 protein since, in addition to replacing the ^{15}N by a ^{14}N isotope, the $H^{\alpha 2}$ proton would be exchanged by a ^{2}H for cultures performed in D_2O. It would lead to profound changes in the ^{1}H-^{13}C correlation of Gly residues as the disappearance of the $^{1}H^{\alpha 2}$-$^{13}C^{\alpha}$ correlations and an upfield shift for the $^{13}C^{\alpha}$ chemical shift due to the isotopic effect. The severe overlaps observed in the Glycine region did not allow us to detect the 25% of the glycines that should be processed by the transaminases in the CECF cell free expression mode (as displayed in figure Chapter III, 3.14).

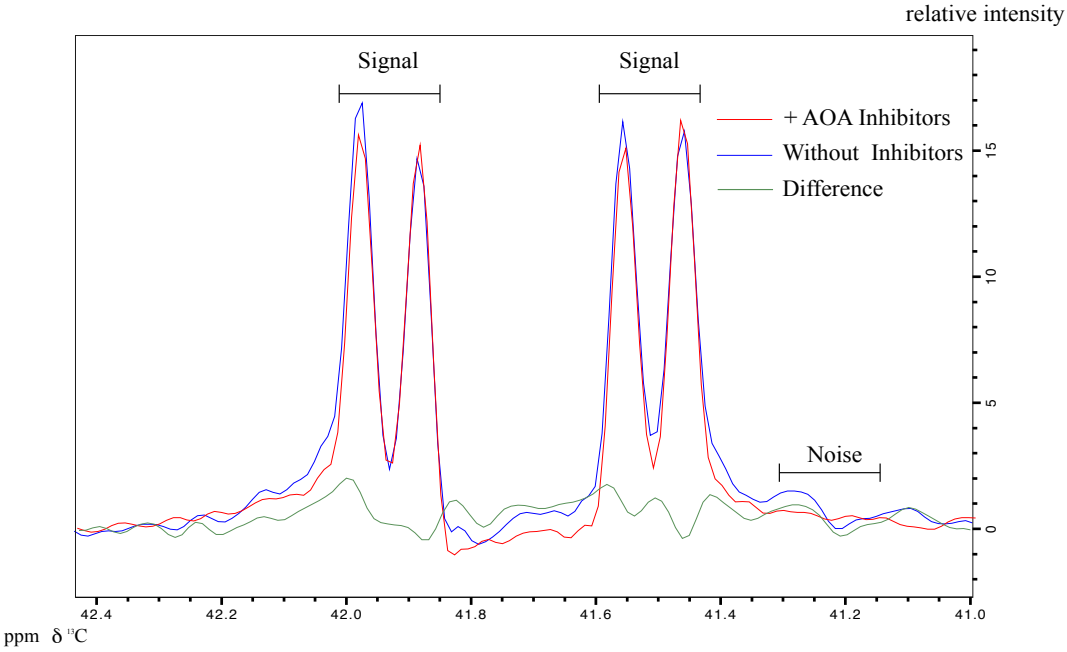


Figure 4.11 : overlay of 1D traces of G13 (δ 4.157 ppm) of ^{1}H-^{13}C HSQC spectra of H23 protein labeled with ^{1}H-^{15}N-^{13}C-Gly and ^{2}H, ^{14}N, ^{12}C-nonGly. Expression in D_2O. Blue: protein concentration 170 µM without inhibitors. Red protein concentration 270 µM with 20 mM AOA inhibitor. pH: 7, 40°C, 10% D_2O, 0,11 mM DSS, 3 mm tube, cryoprobe 500 MHz.

Analysis of the feeding mixes was also undertaken to have an additional overview of the isotope distribution (figure 4.12). Analysis gave clues about whether purification of the S30 cell-extract, adding of inhibitors were significant to maintain the desired isotope pattern on the glycine on the one hand and on the other amino acids on the other hand. We wanted to see if adding inhibitors of transaminases enabled us to reduce the ^{15}N and ^{13}C labeling of other amino acids, notably serine. Four feeding mixes were examined (BD1 to BD4). They correspond to the cell-free expression of H23 in H_2O labeled with ^{1}H, ^{15}N, ^{13}C-Gly and unlabeled for the other amino acids, with (BD2 and BD4) or without (BD1 and BD3) the three inhibitors, and using

crude S30 extracts (BD1 and BD2) or purified S30 extracts (BD3 and BD4). For these last two conditions, a desalting column was used to eventually remove any traces of unlabeled amino acids. Acquisition and processing parameters were identical for each analysis. (See IV.4.4 Material and Methods). For each sample, 1.5 mL of feeding mixes were lyophilized and resuspended in 0.55 mL of D_2O.

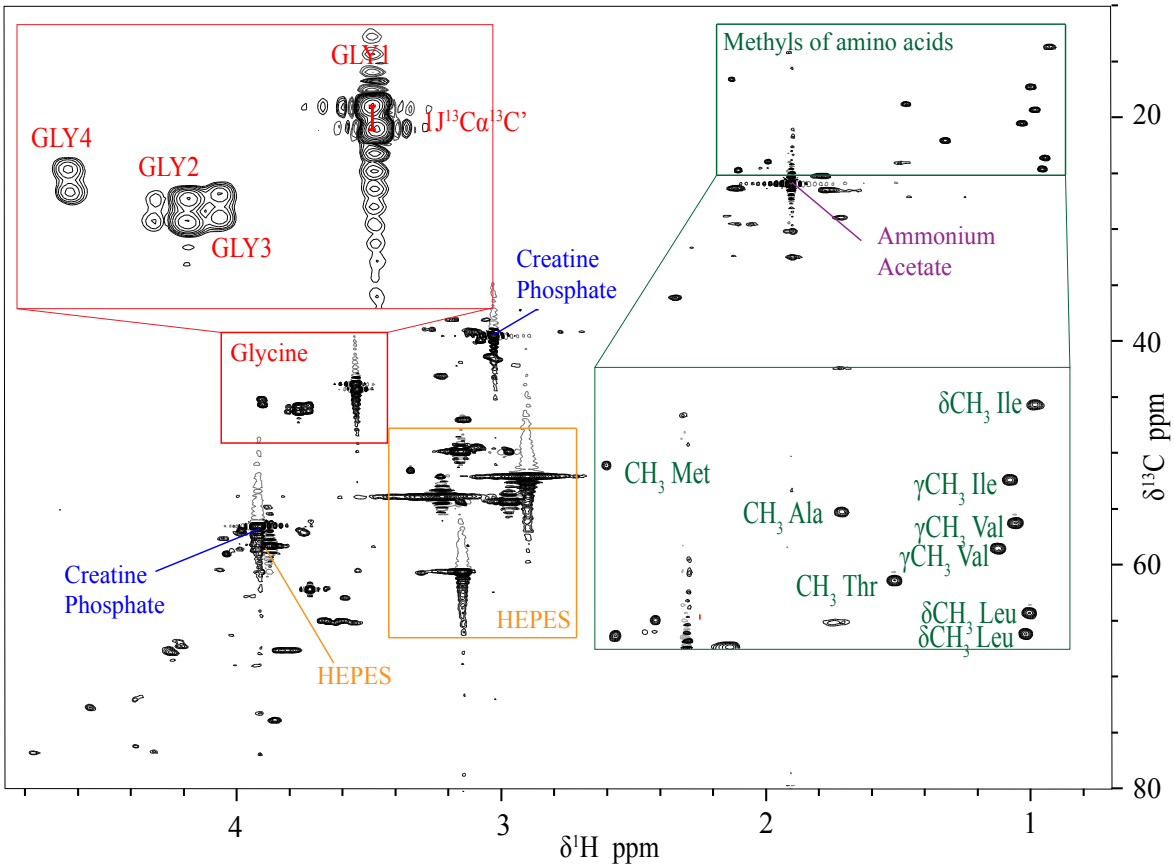
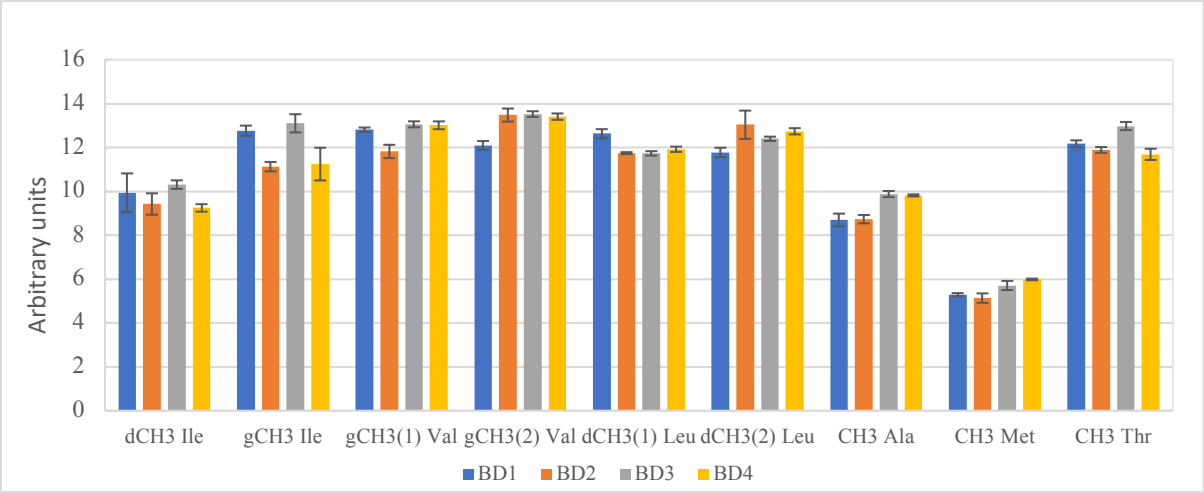


Figure 4.12: ^1H-^{13}C HSQC spectrum of feeding mix from cell-free expression Gly ^{15}N-^{13}C-^1H and other amino acids ^{14}N-^{12}C-^1H labeled produced in H_2O, without inhibitors, normal S30 extract. The boxed regions are of the methyls (green) and glycines (red).

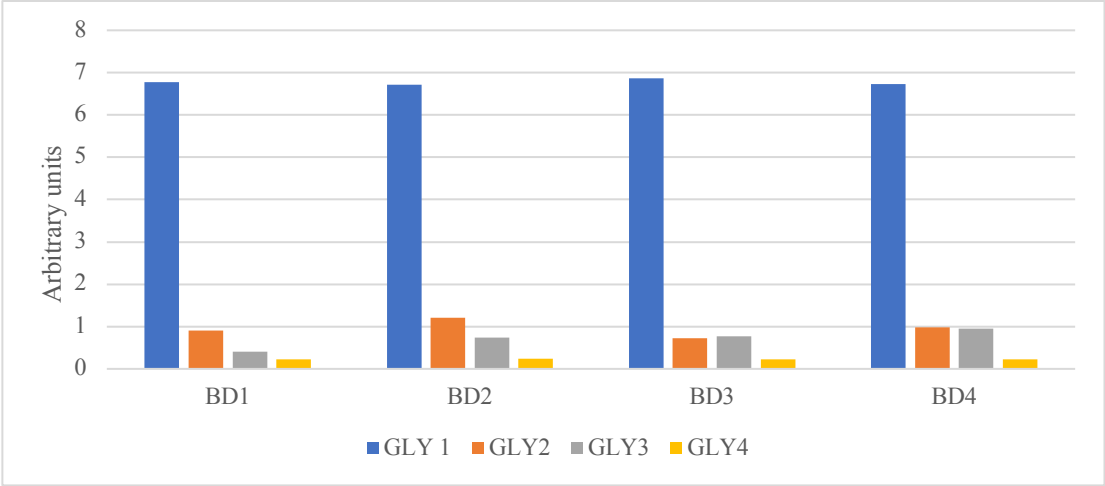
Very intense correlations are visible for HEPES, ammonium acetate, and creatine phosphate which is explained by their high concentrations, hence visible on the spectra even at natural abundance of ^{13}C. Less intense correlations are visible corresponding to natural ^{13}C abundance of methyl groups of aliphatic amino acids; isoleucine, leucine, valine, alanine and methionine. The integrals of these correlations were measured and compared between the feeding mixes to see if there were notable differences in concentration (Graph 4.2).



Graph 4.2 : Histogram of average calculated integrals (n=6) and root mean square deviations for correlations of different methyl groups corresponding to natural abundance of ^{13}C of methyls of amino acid side chains of certain amino acids for each feeding mixture (BD1 to BD4).

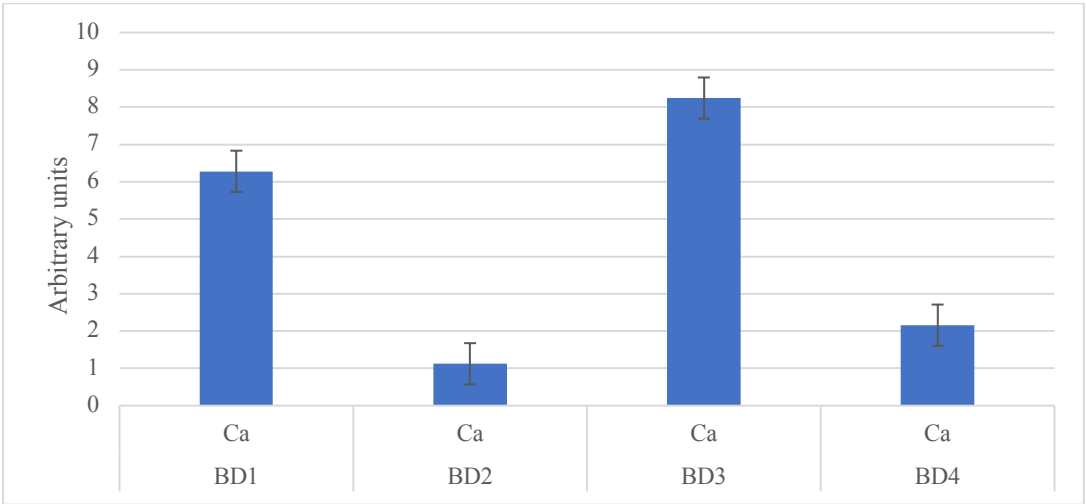
For a given methyl group, the difference of integral between the four feeding mixes is not significative. This indicates that the S30 extract doesn't need to be further purified and will not introduce unlabeled amino acids in the expression system. Small fluctuations are probably due either to sample preparation or to the errors in the integral measurements. The same concentrations were initially introduced in the feeding mixtures, as a single master mix was prepared beforehand. However, the quantity used for protein expression should decrease these concentrations.

We can observe on the ^{1}H-^{13}C HSQC spectrum of the feeding mix four correlations corresponding to glycines or glycine derivatives, of which one very intense and the others weaker correlations. The spectra are resolved enough that we can observe the $^{1}J_{C\alpha\text{-}CO}$ coupling indicating that these are U-^{13}C labeled compounds. These four glycines/glycine derivatives (Gly 1, 2, 3, 4) were systematic in the feeding mixes and they were integrated and compared (Graph 4.3). The intensity of Gly 2, 3 and 4 represents less than 15% of Gly 1. The nature of these glycines derivatives Gly 2, 3 and 4 has not been identified but their presence in BD2 and BD4 suggests that they are not produced by PLP dependent enzymes.



Graph 4.3 : Integrals of the 4 spin systems identified as glycines and glycine derivatives for each feeding mixes (BD1 to BD4).

One $^{1}H\alpha-^{13}C\alpha$ correlation was assigned to the serine residue, with chemical shifts of 3.83 ppm and 59.11 ppm in the proton and carbon dimension, respectively, in agreement with previous studies[38]. Although there was severe overlap observed with the HEPES correlation, we compared the intensity of this correlation depending on the feeding mix (Graph 4.4), to evaluate the ability of inhibitors to diminish the serine production from the ^{13}C-labeled glycine. We verified a significant decrease in intensity of the $C\alpha$ of serine (4 to 6 fold) residues in the feeding mixes when inhibitors are used in the expression mixture which is in agreement with the measurements on the H23 proteins. However, the slight increase in labeled serine observed when the S30 has been purified is not explained.



Graph 4.4: Histogram of intensity of $^{1}H-^{13}C\alpha$ correlations of serine from the feeding mixes of cell-free expressed H23, in H_2O, without (BD1 and BD3) or with (BD2 and BD4) the three inhibitors, S30 purified (BD3 and BD4) or not (BD1 and BD2) Gly $^{1}H-^{15}N-^{13}C$ and other amino acids $^{2}H-^{14}N-^{12}C$, expressed in D_2O. Values are a mean of 4 measurements with standard deviation.

We also analyzed our protein samples by mass spectrometry (MALDI TOF) to have additional data for assessing our labeling procedures, although the distribution of the isotopes will not be obtained at the residue scale. Samples were diluted in milliQ H_2O at 10 µM and mixed with matrix (HCCA) for acquisition (see IV.4.6. for complete protocol). The figure 4.13 corresponds to the cell-free expressed protein H23 expressed in different cell-free conditions:

- Specifically labeled with ^{1}H, ^{15}N, ^{13}C-Gly and unlabeled for the other amino acids (^{1}H, ^{14}N, ^{12}C) produced in H_2O without (panel A) or with inhibitors (panel B).

- H23 protein was also expressed in H_2O with ^{1}H-^{15}N-^{13}C-Gly and triply labeled (^{2}H-^{15}N-^{13}C) on the other amino acids without (panel C) and with inhibitors (panel D).

- H23 protein expressed in D_2O, with ^{1}H, ^{15}N, ^{13}C-Gly and deuterated amino acids for the other residues (^{2}H, ^{14}N, ^{12}C) without (panel E) and with (panel F) the three inhibitors.

Two different forms of H23 protein are observed on certain spectra, with one about 2 kDa lighter than the other, suggesting a truncation of the protein, which corresponds to an old sample in this case. It recalls the degradation that we observed in NMR when the data were not recorded on fresh samples (Chapter II).

There are a total of 20 negatively charged residues (Asp + Glu) and 18 positively charged residues. With an atomic composition of $C_{724}H_{1169}N_{211}O_{231}S_2$ hydrogens at pH 7, theoretically, the mass of unlabeled H23 is 16589.64 Da if the methionine is not present. When labeled ^{15}N-$^{13}C_2$ on the 15 glycines, the mass found by MASS spectrometry for H23 was 16615.2 Da, therefore the predicted mass when unlabeled should be 16570.2 Da (MassH23_1) which is <20 Da weaker than the theoretical mass.

To evaluate the percentages of labeling of the H23 U-(^{2}H-^{15}N-^{13}C) labeled and Gly (^{1}H-^{15}N-^{13}C) produced in H_2O, the percentage of labeling was calculated using the percentages given by the provider following the equation:

= (MassH23_real – MassH23_1) / (Δ_theoretical_labeling) x100 = (18183.96 – 16570.2/ (724*0.98 ^{13}C + 211*0.98 ^{15}N + 864*0.97 $^{2}H_{(non\text{-}exchangeable\_2H\ (not\ counting\ glycine))}$). This equals to 92% for the sample produced without inhibitors. The calculation amounts to 95.4% for the sample produced with the three inhibitors.

For the H23 U-(^{2}H) labeled and Gly (^{1}H-^{15}N-^{13}C) produced in D_2O, the percentage of labeling was calculated following the equation :

= (MassH23_real – MassH23_1) / (Δ_theoretical_labeling) x100 = (17388.21 – 16570.2/ (864*0.97 $^{2}H_{(non\text{-}exchangeable\ 2H)}$+ $45_{(Gly\ 15N\ 13C)}$). This equals to 92.7% of labeling for these samples expressed with or without inhibitors. This suggests that there is a global labeling efficiency of 92.7% for ^{2}H nuclei. Which is relatively similar to the labeling percentage calculated in Chapter III for the Cα positions of amino acids. Previously, protonation at the Cα sites was estimated at around 40% when no inhibitors were used, and fell to an average of 5% with the use of the three inhibitors. However, expression having been effected in a D_2O solvent, this does not explain the source of protons. It is possible that during hydrolysis of the Celtone mix, performed in H_2O, some protons could have stayed present.

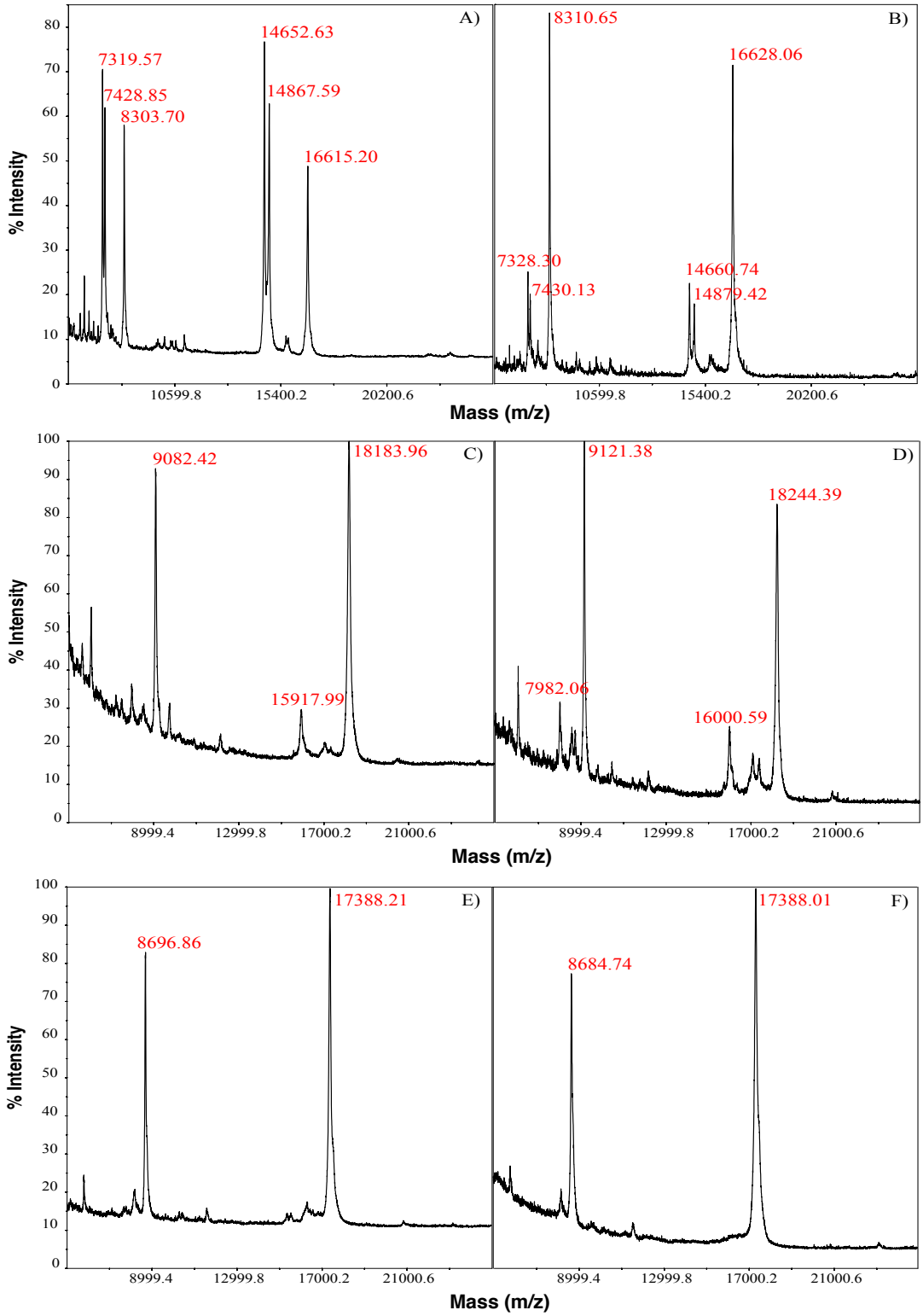


Figure 4.13 : MALDI-TOF spectra of cell-free expressed H23 Gly ^{1}H-^{15}N-^{13}C- labeled and other AA ^{1}H-^{14}N-^{12}C produced in H_2O (A and B), or other amino acids ^{2}H-^{15}N-^{13}C- labeled produced in H_2O (C and D) or other amino acids ^{2}H-^{14}N-^{12}C labeled produced in D_2O (E and F). The panels on the left are produced without inhibitors, on the right produced with the three inhibitors (20 mM AOA, 20 mM D-malate, $NaBH_4$ treatment).

IV.2.4. Application of the protocol to the Pin1 protein

Pin1 was expressed following the cell-free protocol tested on H23 for the specific labeling of glycines (^{1}H, ^{13}C, ^{15}N) with all the other amino acids deuterated (^{2}H, ^{12}C, ^{14}N). For Pin1 however, yields ranged from 3 to 30 fold lower. The purification protocol was adapted to allow the handling of very weak amounts of protein. When Pin1 was expressed in *E.coli*, the purification consists of a passage on a nickel affinity column, overnight dialysis coupled to incubation with the PreScission enzyme for tag cleaving, a second passage on a nickel affinity column, passage on an exclusion chromatography column and concentration. But this protocol is adapted for cell-based expressions, with several mg yields of protein, large volumes and significant presence of other proteins. However in cell-free expression, volumes range around 1 to 3 mL and there are much less quantities of the protein of interest as well as unspecific proteins. This can cause issues with purification protocols because the many steps involved increase the loss of protein of interest. Hence we have adapted the purification for cell-free expressed Pin1. The table below lists three cell-free protocols undertaken for expression of Pin1 and the purification protocols applied with yields and concentration measures. Total losses account for more than 80 % of initial yield with these purification steps although this does not take into account the contaminant proteins which represented a large fraction of the expressed proteins in these productions. As a side note, reaction mixtures for the second production were prepared in advance and frozen but temperature problems during transport resulted in almost no protein production. A detailed account of the third production is detailed on the next pages.

	Production Vol	Purification Step	Mg/mL Nanodrop	Mg total
1	15 mL (5x3 mL)	-Purification FPLC His-trap -Dialysis 3kDa tube + PreScission 16h, 4°C -2nd purification FPLC His-trap -Exclusion chromatography -Concentration pressure cell	0,35 mg/mL 0,175mg/ml <0,01mg/mL <0,01 mg/mL 500 µl at 76 µM	5,27 mg 2,59 mg 0,7 mg
2	6 mL (2 x 3ml)	-Purification nickel resin -Dialysis 3 kDa cassette + PreScission 16h, 4° -2nd purification nickel resin -Concentration pressure cell	0,063 mg/mL 0,004 mg/mL 110 ul at 38 uM	0,38 mg 0,024 mg 0,077 mg
3	15 mL (5x3 mL)	-Purification nickel resin -Dialysis 3 kDa cassette 30 minutes -PreScission 16h, 4°C -2nd Purification nickel resin -Concentration pressure cell	0,1 mg/mL 0,14 mg/ml 250 uL at 25 uM	1,54 mg 0,44 mg 0,23 mg

After the first passage of cell-free expressed Pin1 through a nickel resin, different fractions were recuperated and migrated on a SDS-PAGE electrophoresis acrylamide gel.

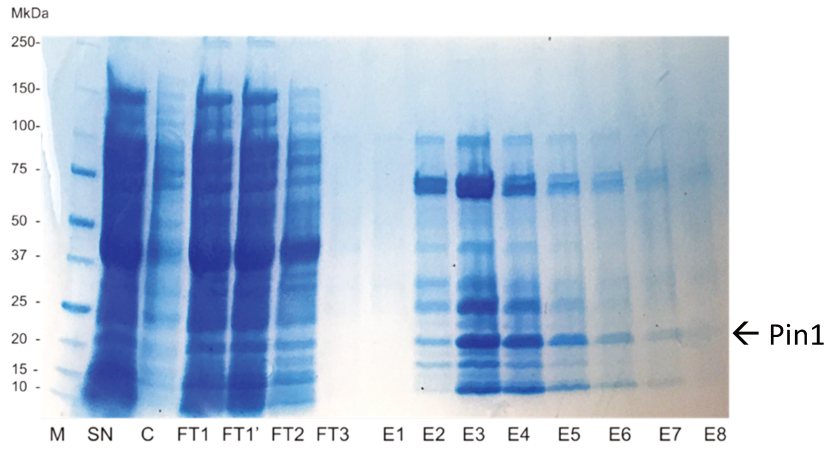


Figure 4.14**:** Photo of the elution fractions of Pin1 on a SDS-PAGE electrophoresis gel after a first passage on a nickel affinity resin (The deposits are all made with 15 µL of sample and 5 µL of 4x bromophenol blue). SN: Supernatant, C: deposit, FT1 : Flow through 1,FT1' : Flow through 1 after being passed a second time on the nickel resin, FT2 : Flow through 2 (washed with 30 ml phosphate buffer 20 mM, pH 7.5, NaCl 150 mM, 5mM imidazole), FT3 : Flow through 3 (washed with 30 ml phosphate buffer 20 mM, pH 7.5, NaCl 150 mM, imidazole 15mM). E1 to E8: elution fractions 1 to 8 (750 µl of phosphate buffer 20 mM, pH 7.5, NaCl 150 mM, 300 mM imidazole) Pin1 + His-tag = 20,11 kDa.

Pin1 is clearly visible in fractions E3 and E4 in which other proteins of different molecular mass are also visible. Notably there seems to be a high concentration of proteins with molecular masses of 25 kDa and 75 kDa. This purification technique using the nickel resin leaves the sample still very impure, considering the number of non-specific proteins visible. Elution fractions E3 et E4 (noted as $F_{3,4}$ hereafter) were pooled together for the future steps. The pooled fractions $F_{3,4}$ were dialyzed 30 minutes to lower the imidazole concentration against a phosphate buffer 20 mM, pH 7,5, NaCl 150 mM, Imidazole 15 mM, EDTA 1mM, DTT 1mM at 4°C, then recuperated and incubated with the PreScission enzyme, 100 µL at 1 mg/mL, overnight at 4°C. The next day, $F_{3,4\ was}$ passed through the nickel resin a second time to remove the cleavage enzyme. A weak band migrates to the corresponding molecular weight of Pin1 and the quantification of the $EF_{3,4}$ suggests an additional 60% protein loss during the dialysis and/or the second purification step. $EF_{3,4}$ was concentrated using an Amicon® pressure cell on a 3.5 kDa membrane, which resulted in a supplemental loss of 50% of protein (0.23 mg total).

Moreover, the electrophoresis SDS-PAGE acrylamide gel of this final fraction revealed that other proteins were still present (especially the proteins with molecular masses of 25 kDa et 75 kDa already observed in the first eluates). When purifying the Pin1 proteins was obtained from *E.coli* cultures, the losses observed at each step of this purification protocol is probably

negligible compared to the large amount of protein produced, but this is not true for the low protein quantities obtained by cell free.

To palliate the issue of purification, adaptation of the protocol was undertaken. The strategy employed to optimize each purification step was to use uniformly labeled ^{1}H-^{15}N Pin1 samples expressed in *E.coli* and place them in the same sample conditions of volume and yield than for the cell-free expressions and quantify losses during these different steps as well as find new techniques of purification. Pin1 ^{15}N labeled was produced following the protocol in Chapter II.

Initially, the biggest loss of sample was the dialysis step. To analyze in real time the dialysis process, we quantify protein concentration at different time points during dialysis. A range of phosphate buffers (30 mM pH 7, 50 mM NaCl) at different imidazole concentrations (0, 10, 50, 100, 300 and 400 mM) was made as a standard reference. The absorbance was measured by UV spectrometer (optical trajectory of 1 mm) at 230 nm and 280 nm and extinction coefficients were determined for imidazole at these two wavelengths.

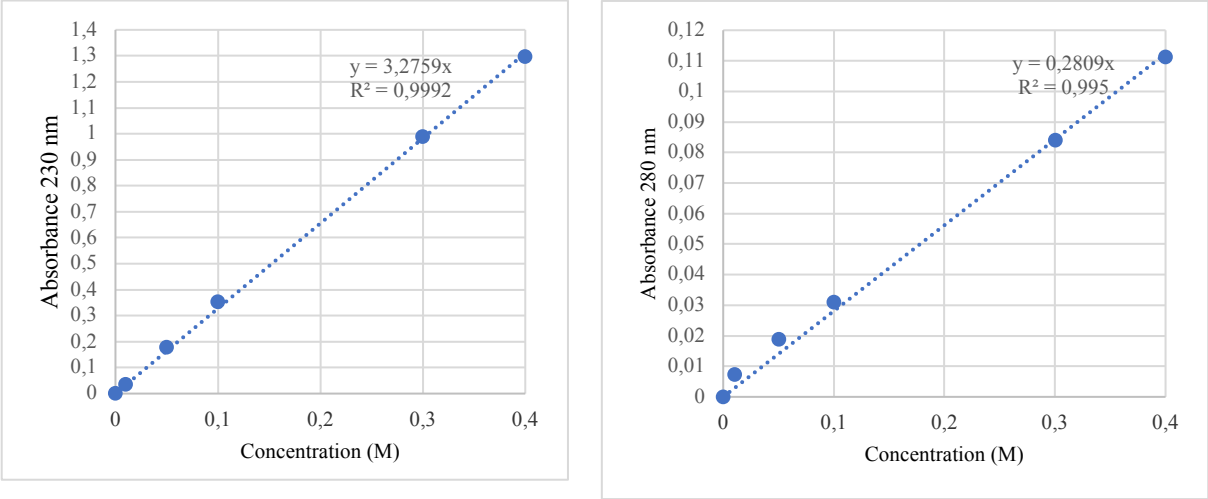


Figure 4.15 : The ε of imidazole were estimated at 230 nm : 3,276 $L.mol^{-1}.cm^{-1}$ and 280 nm : 0,28 $L.mol^{-1}.cm^{-1}$

An initial sample of Pin1 at 0.82 mg/mL of protein and 350 mM imidazole was prepared and dialysis was performed using a 3 kDa cut-off cassette membrane and OD_{280} and OD_{230} were regularly measured at 15 min, 30 min, 1h, 1h30, 3h and 4h. The UV spectra for Pin1 0.82 mg/mL at 350 mM imidazole is displayed in figure 4.16.

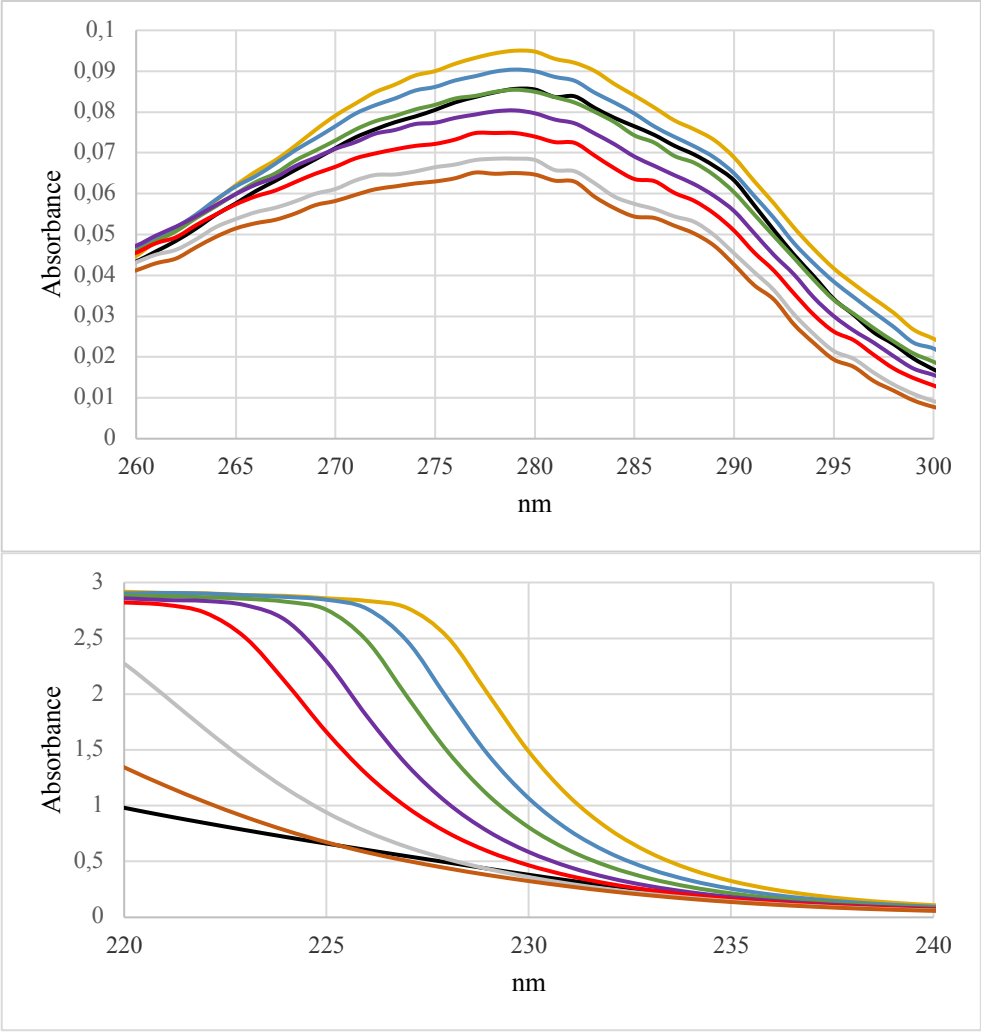


Figure 4.16 : UV absorbance spectra of the Pin1 sample initially containing 350 mM imidazole during dialysis in the 260-300 nm range (upper panel) or in the 220-240 nm range (lower panel). In black Pin1 0,82 mg/mL. In yellow Pin1 0,82 mg/mL at 350 mM imidazole. In blue after 15 minutes dialysis. In green after 30 minutes dialysis, In purple after 1 hour dialysis. In Red after 1h30 dialysis. In grey after 3h dialysis and in brown after 4 hours dialysis.

A slight increase in absorbance at 280 nm was visible (+ 0,01) after addition of imidazole to a concentration 350 mM to the initial Pin1 sample. This was expected because imidazole has a small (but not null) extinction coefficient at this wavelength. However, absorbance diminishes well under the initial value of Pin1 alone during dialysis (from 0.082 to 0.063), indicating a loss of protein as well as imidazole. When looking at absorbance at 230 nm, an increase was also visible (+ 1,1) between the sample containing the Pin1 protein only and the sample containing the protein and 350 mM imidazole. Absorbance values are additive, therefore theoretical A_{230} of Pin1 (0,4) added to A_{230} of Imidazole at 350 mM (1,15 expected from the standard calibration) equals 1.55, which corresponds well to the values found experimentally here. Using the Beer-lambert law, the concentration of Pin1 during dialysis can be estimated using a 2-equation system with two unknowns:

A(total) = A(Pin1) + A(Imidazole) giving :

$A_{230} = \varepsilon_{230}\_(Imidazole) \times c\_(Imidazole) + \varepsilon_{230}\_(Pin1) \times c\_(Pin1)$

$A_{280} = \varepsilon_{280}\_(Imidazole) \times c\_(Imidazole) + \varepsilon_{280}\_(Pin1) \times c\_(Pin1)$

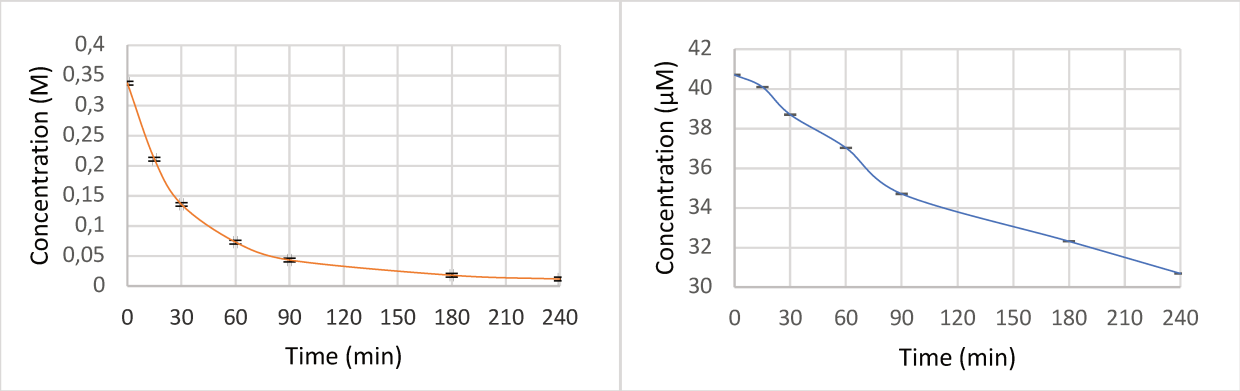


Figure 4.17 : Calculated concentrations of imidazole (left) and Pin1 (right) during dialysis.

The concentration of imidazole and Pin1 have been then determined at each time out of this system of two equations. Loss of protein is estimated at 10% after 1 hour of dialysis and 25% after 4 hours of dialysis. However, most of the imidazole (78%) is dialyzed out during the first hour suggesting dialysis could be limited to one hour, a reasonable compromise between loss of protein and dilution of imidazole. But to assess whether the dialysis step could be skipped, considering it is only useful to lower the concentration of imidazole for the cleavage of the His-Tag, a test was done to find out whether PreScission was still active at high imidazole concentrations. To achieve this, 2 batches of Pin1 (150 µL at 0,47 mg/mL) were incubated 16h at 4 °C in a phosphate buffer 30 mM, pH 7, NaCl 50 mM with either 0 mM imidazole or 220 mM imidazole and with increasing concentrations of PreScission enzyme: 0, 1 µg, 2 µg, 5 µg, 20 µg and 50 µg of (from aliquot of PreScission at 1mg/mL).

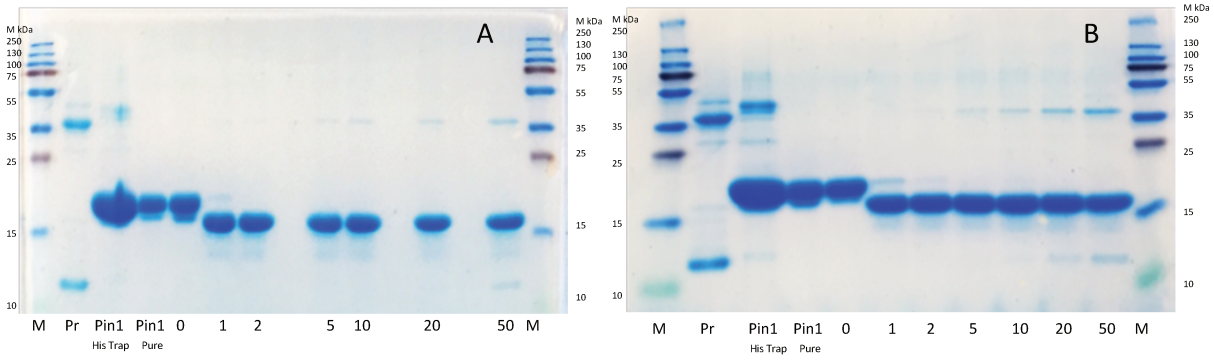


Figure 4.18 : Electrophoresis gels revealing the PreScission activity on Pin1. The left gel : PreScission was incubated without imidazole , the right gel : PreScission was incubated with 220 mM of imidazole. M : molecular ladder, Pr : PreScission, $Pin1_{His}$: Pin1 after a first passage of the His-Trap, Pin1 : Pin1 after exclusion chromatography, 0 : Pin1 + 0µg of PreScission, 1: Pin1 + 1µg of PreScission, 2 : Pin1 + 2µg of PreScission, 5 : Pin1 + 5 µg of PreScission, 10 : Pin1 + 10 µg of PreScission, 20 : Pin1 + 20 µg of PreScission, 50: Pin1 + 50 µg PreScission

Stocks of PreScission proteins (produced in the lab, see Annex 2 for expression) are not extensively purified and several bands are visible when concentrated in addition to the protease appearing at 36 kDa (see wells, figure 4.18). After a first passage on the His-trap, Pin1 (produced here in *E. coli*) is not so perfectly purified either although a thick band is visible at 20 kDa. A passage on the exclusion chromatography column is needed to properly eliminate impurities (Pin1 wells). For the samples incubated with at least 1 µg of PreScission enzyme, the thick bands have migrated to 18 kDa, suggesting the His-tag was properly cleaved. Except for the 1 µg addition of PreScission, where a thin yet visible band is still present at 20 kDa suggesting incomplete cleavage. A control experiment has been performed (no PreScission added but the Pin1 sample was incubated overnight) which proved that the observed cleavage was due to the presence of the PreScission protease. No difference was observed between the samples with or without imidazole suggesting high concentrations of this salt (up to at least 220 mM) does not affect the activity of PreScission. To validate these claims MALDI-TOF tests were undertaken of each sample after incubation (see section Chapter IV.4.6. for preparation details).

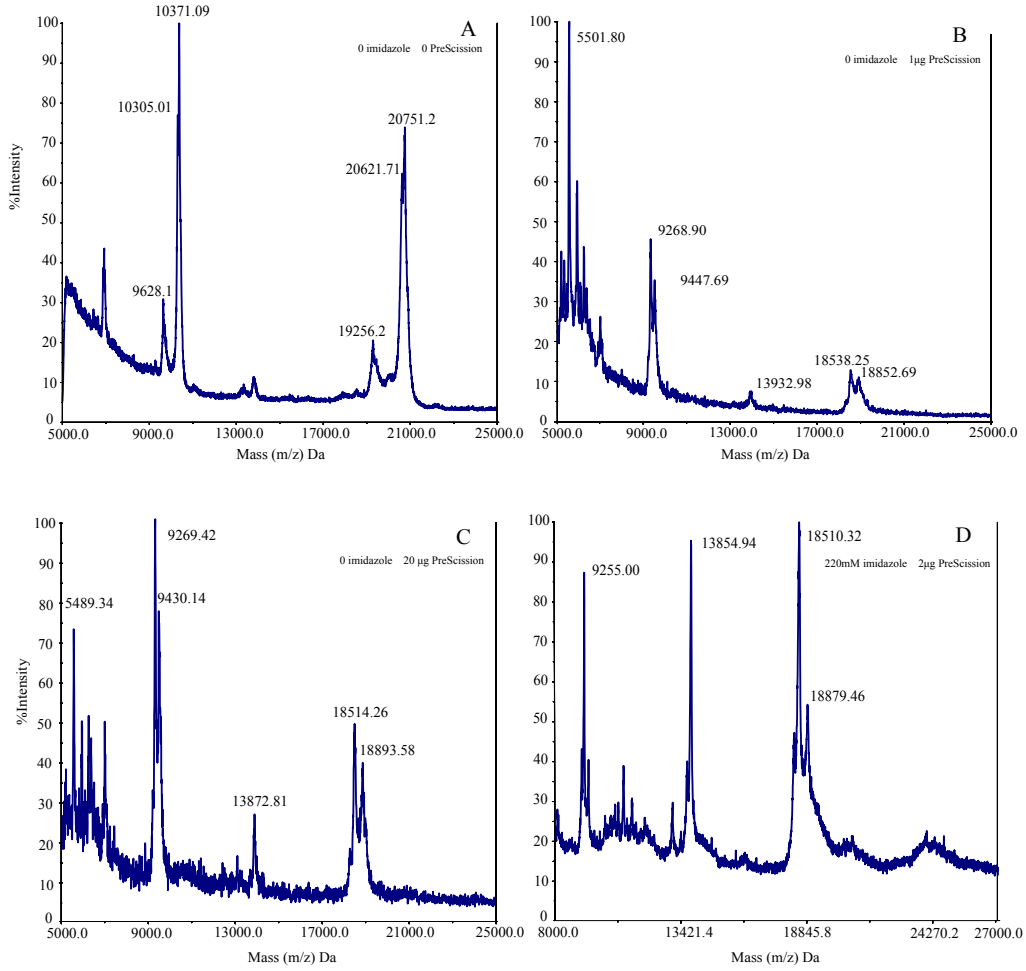


Figure 4.19: MALDI TOF spectra of Pin1 with or without different concentrations of PreScission enzyme

In all the MALDI spectra figure 4.19, a component was found around 13,8 kDa which corresponds to the isolated PPIase domain of Pin1. We frequently observed this WW/PPIase cleavage when the samples were standing a few days after purification, even when no PreScission protease were added. Full ^{15}N labeled Pin1 with the attached His-tag should have a mass of 20860,74 Da (Pin1-HisTag = 185 aa 20586,74 Da + 274 N). Cleavage of the His-Tag should result in a mass of 18814,6 Da (Pin1 = 167 aa 18571,6 Da + 243 N), corresponding to a mass difference of 2045,4 Da. Either way, several forms of Pin1 are visible after several steps of purification with mass differences of c.a. 300 Da. The imidazole had no real influence on the spectra profiles and when the PreScission protease was added, the two peaks standing for the uncleaved Pin1 (at 20621 and 20751 Da) disappeared and peaks at ~18800 Da appeared, close to the expected mass of the cleaved Pin1 (about 18.5 kDa). The His-tag was then correctly cleaved even with high concentrations of imidazole.

In light of these results, it was decided to skip the dialysis step altogether and just dilute the sample before using in a second His-trap purification step so to avoid any loss of protein of interest and because the PreScission enzymes activity is unaltered by the strong imidazole concentrations.

Concentrating the protein was a step which caused much loss of protein materials, whether Centricon tubes or pressure cells were used. Pin1 was often found at relatively high concentrations in the flow through, for unclear reasons. Cut-off membranes as low as 3.5 kDa were used to limit the flow of Pin1 through the membrane without any success. Several membrane types and several buffer conditions were tested with limited success. Finally, the Amicon® pressure cells were preferred over the Centricon tubes and the Phosphate buffer 30 mM, pH 7.8, 50 mM NaCl over Tris or HEPES Buffer, results obtained by DSF (Differential Scanning Fluorimetry) but not displayed in this thesis.

High-performance liquid chromatography (HPLC) was also tested for the purification of the full length Pin1 but was not giving satisfactory results. Most of the lyophilized powder obtained after the purification was not soluble anymore, suggesting a loss of the Pin1 tridimensional structure and aggregation (see Chapter II). In contrast, HPLC purifications were very efficient on the isolated WW domain.

IV.3. Benefits of Gly specific labeling for NMR experiments

IV.3.1. CH_2-TROSY experiments

Gains in signal to noise ratio (S/N) and resolution in the proton dimension were measured between the CH_2-HSQC and CH_2-TROSY spectra (see section IV.4.6 for recording conditions). The CH_2-HSQC experiment refers to the sensitivity enhanced HSQC[39] where the double INEPT transfer back was optimized for CH_2 groups. Two samples of H23 protein were used for this purpose, obtained by cell-free expression with the optimized protocol described previously. One was uniformly ^{13}C-^{15}N labeled and the other was specifically Gly ^{1}H-^{15}N-^{13}C labeled with the other amino acids ^{2}H-^{15}N-^{12}C labeled. Spectra were recorded in two different solvents: H_2O or, after lyophilization, in D_2O and at three different temperatures (40°C, 25°C and 10°C). Figure 4.20 presents both a CH_2-HSQC spectrum and a CH_2-TROSY spectrum, recorded at 25°C for the Gly specifically labeled sample in 100% D_2O, enabling the appreciation of the difference in spectral quality achieved through use of the CH_2-TROSY. Quantification of signal to noise ratio and proton linewidth were undertaken on 9 resolved pairs of diastereotopic protons of glycine residues. G13, G16, G18 and G74, which are situated either in the linker or flexible loops and G82 in a β-strand, do not present significant chemical shift differences for their diastereotopic protons and were therefore not analyzed. In some spectra, correlations were overlapping each other or were close from the trace of water, making quantification difficult or impossible (Figure 4.21).

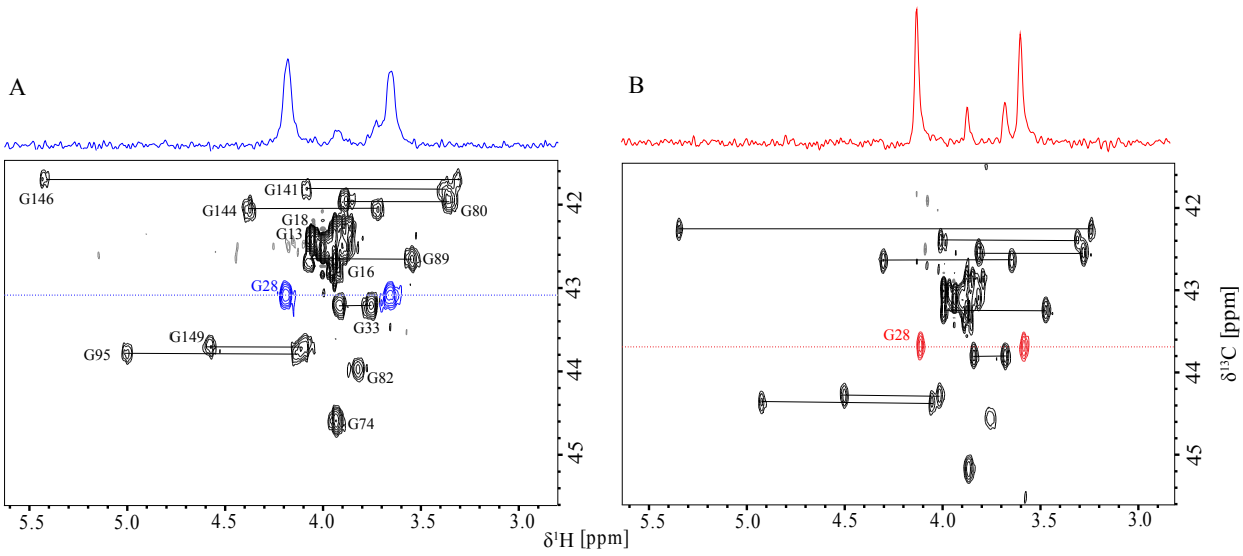


Figure 4.20: CH_2-HSQC and CH_2-TROSY spectra of H23 protein, Gly specifically labeled ^{15}N-^{13}C and other amino acids ^{15}N-^{2}H expressed by cell-free with the optimized protocol. A) CH_2-HSQC spectrum with a 1D extracted trace above. B) CH_2-TROSY spectrum with a 1D trace extracted above. Protein concentration: 200µM in 3 mm Shigemi tube, phosphate 20 mM, pH 7.0, NaCl 100 mM, 1mM DTT, 1% DSS, 0,05% NaN3, 100% D_2O. Spectra recorded at 25°C, on a 950MHz spectrometer equipped with a cryoprobe .

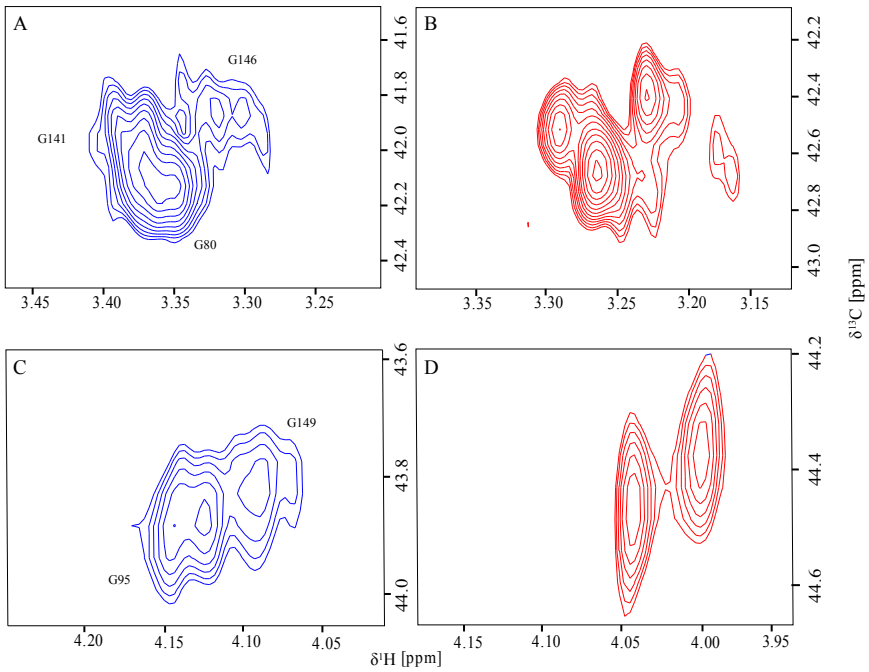


Figure 4.21: CH_2-HSQC and CH_2-TROSY spectra of H23 protein, Gly specifically labeled ^{15}N-^{13}C and other amino acids ^{15}N-^{2}H expressed by cell-free with the optimized protocol. A) and C) CH_2-HSQC spectra, B) and D) CH_2-TROSY spectra. Protein concentration: 200µM in 3 mm Shigemi tube, phosphate 20 mM, pH 7.0, NaCl 100 mM, 1mM DTT, 1% DSS, 0,05% NaN3, 10% D_2O, 90% H_2O. Spectra recorded at 25°C, on a 950MHz spectrometer equipped with a cryoprobe.

Visually, the TROSY experiment enhances glycine resolution considerably in the proton dimension compared to the HSQC. Even in H_2O based solvent (Figure 4.21), the specifically labeled sample enables the experimenter to gain precious information when recorded with a TROSY spectra. G149 and G95 correlations are undiscernible in the HSQC spectra but almost completely separated in the TROSY spectra. An important part of the resolution gain is due to the $^{2}J_{H\alpha 2\text{-}H\alpha 3}$ homodecoupling achieved in the TROSY spectra.

For each correlation, the 1D trace in the proton dimension was extracted and signal to noise ratio as well as resolution were measured using respectfully the Topspin SINO command and by measuring the width of the peak at half-height. Note that we do not look at the gain in resolution in the carbon dimension in this analysis, only the resolution in the proton dimension was measured. The values obtained for each correlation in both spectra and for all condition are reported below (Figure 4.22 and 4.23). Error bars were calculated for S/N as the root mean square deviation of all the S/N values of the protons of a given spectrum. For resolution, error bars were calculated as the width of the correlation divided by S/N value. From these values, the apparent trend is that the width of correlations diminishes in the proton dimension and the intensity increases when using the TROSY experiment compared to the HSQC, with some

exceptions, notably for the uniformly labeled sample at low temperature. These conditions favorized the relaxation as they correspond to the highest proton density and the highest correlation time. The signal to noise ratio measured for the specifically labeled glycines seem overall slightly reduced compared to the uniformly labeled glycines (about 20% less), even though there is technically the same concentration of protein in each. This could be explained because the specifically labeled sample was made using deuterated algal mixtures in which there is already glycine. Thus, even though the reaction mixture is supplemented with 10-fold excess of labeled glycine compared to the supposed amount in the algal mix, there may be some unlabeled ^{12}C algal glycine incorporated into the protein. Hence, absolute values cannot be directly compared and enhancement factors had to be determined. The enhancement factor for signal to noise ratio $\varepsilon_{S/N}$ was calculated as the ratio of the value of intensity found in the TROSY spectrum by the value found in the HSQC spectrum. The enhancement factor for resolution ε_R was calculated as the ratio of the value of the width at half height of the correlation found in the HSQC spectrum by the value of the width at half height of the correlation in the TROSY spectrum. The theoretical expression of ε_R is given in the chapter 1.
The ratios are given below:

$$\varepsilon_{S/N} = \frac{S/N_{TROSY}}{S/N_{HSQC}} \text{ and } \varepsilon_R = \frac{R_{HSQC}}{R_{TROSY}}$$

For each diastereotopic proton of glycine, the enhancement factors were determined. In order to correlate these enhancements with the proton density surrounding each glycine proton, we introduced the d_A parameter. This parameter defines a distance between a given glycine proton and a virtual proton that would be solely responsible for all the dipolar interactions imparted by the surrounding protons. d_A parameters are then calculated for each glycine protons, taking into account the local structure and more precisely the proton in the immediate vicinity, when not be replaced by ^2H nuclei. Details are given in section IV.4.8 for d_A calculations. Enhancements factors were plotted against this apparent distance d_A (Figure 4.24). In this way, comparing values between different conditions was possible and we can understand the influence of neighboring protons on the gain in sensitivity and resolution, which is equivalent to the gains in relaxation obtained. Indeed, from this figure we can observe that the more the environment of a glycine is deuterated, the more the enhancements in resolution and sensitivity are important, regardless of the temperature.

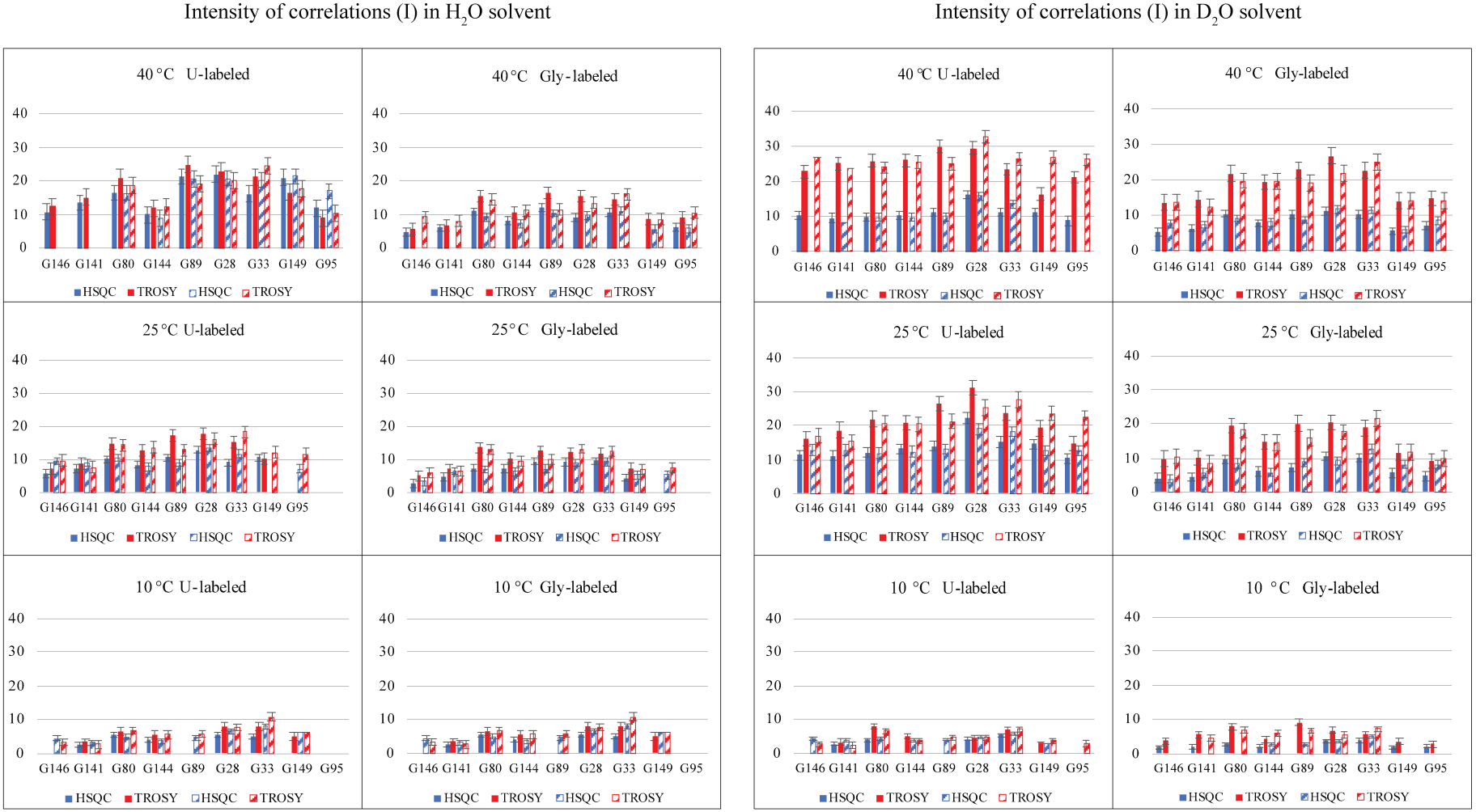


Figure 4.22: Intensities of correlation stand for signal to noise ratio. For each glycine residue, up to two measurements are performed in each experiment corresponding to the two diastereotopic protons H 2 and H 3. Values measured in the HSQC (blue) and TROSY(red) spectra for the different conditions. In H_2O or D_2O solvent, at 40°C, 25°C and 10°C .

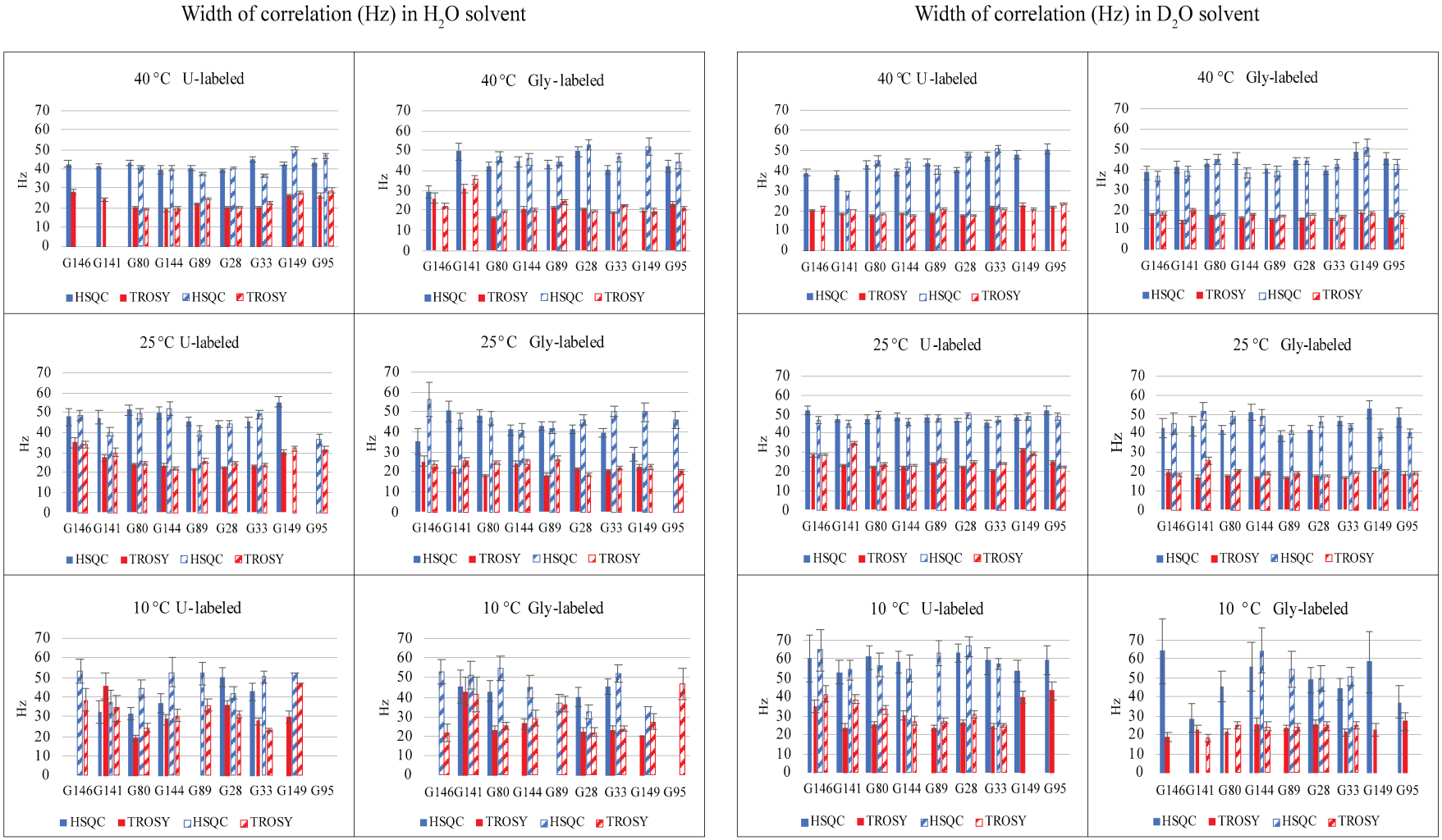


Figure 4.23: Widths of correlations measured at half height. For each glycine residue, up to two measurements are performed in each experiment corresponding to the two diastereotopic protons H 2 and H 3. Values measured in the HSQC (blue) and TROSY(red) spectra for the different conditions. In H_2O or D_2O solvent, at 40°C,

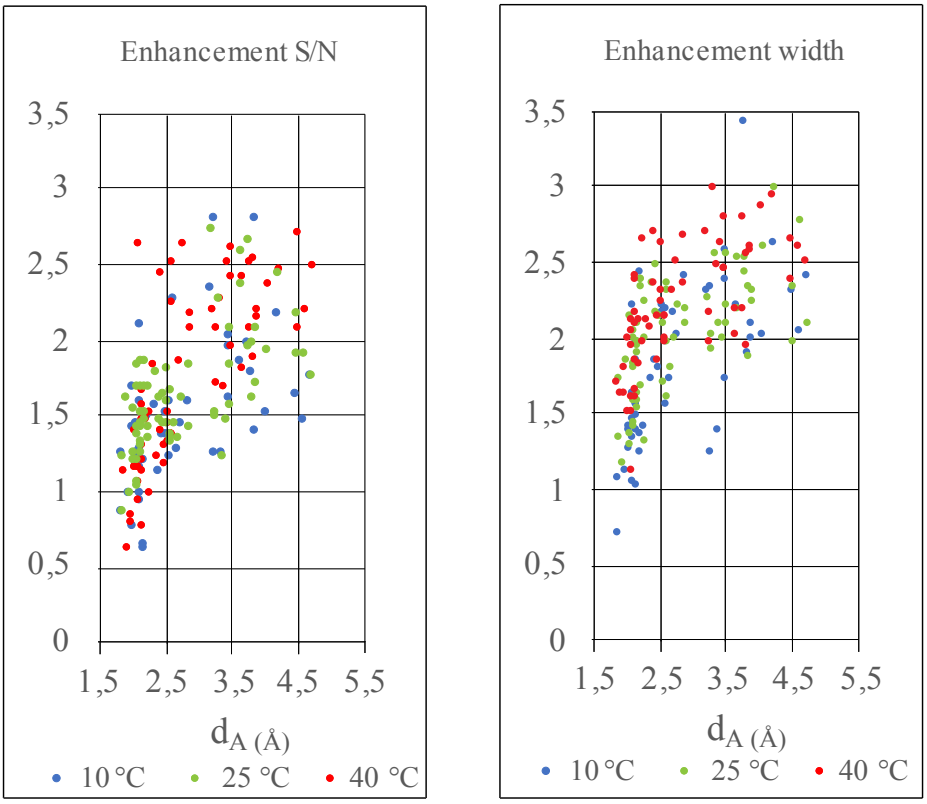


Figure 4.24 : Plot of enhancement S/N factor ($\varepsilon_{S/N}$, left panel) and of enhancement linewidth factor (ε_R, right panel) for each diastereotopic proton of glycine against the apparent distance d_A (Å)found for that proton at different temperatures, 40°C (red), 25°C (green), and 10°C (blue).

In Figure 4.25, seven ranges of d_A have been defined, from 1.5 Å to 5 Å, separated by 0.5 Å steps. This allows for the calculation of averaged values merging all the enhancements measured in a same d_A range. Averages of enhancements in S/N and resolution of the protein are then reported as a function of the d_A range at 10°C, 25°C and 40°C. This graph clearly shows that these enhancements increase as the environment is more deuterated, which was expected. The glycine specifically labeled sample in D_2O is characterized by a large occurrence of high d_A values with a mean of ~3.75 Å (blue curve on the left panel). Whereas the greatest enhancements were obtained for this sample, the removal of HN protons using a deuterated buffer allows, by itself, a significant improvement of the spectral parameters in uniformly labeled samples. At 10°C and for uniformly labeled samples in H_2O, enhancements for resolution and S/N are of 1, signifying that there is little to no difference in signal between the TROSY and HSQC experiments whereas for specifically labeled samples in D_2O, enhancement means are of 2,2 for resolution and 1,8 for S/N. When correlation time τ_C decreases, enhancements increase (mean of 2.3 for resolution and 2.0 for sensitivity at 25°C, and mean of 2.5 for resolution and 2.3 for sensitivity at 40°C for specifically labeled samples in D_2O).

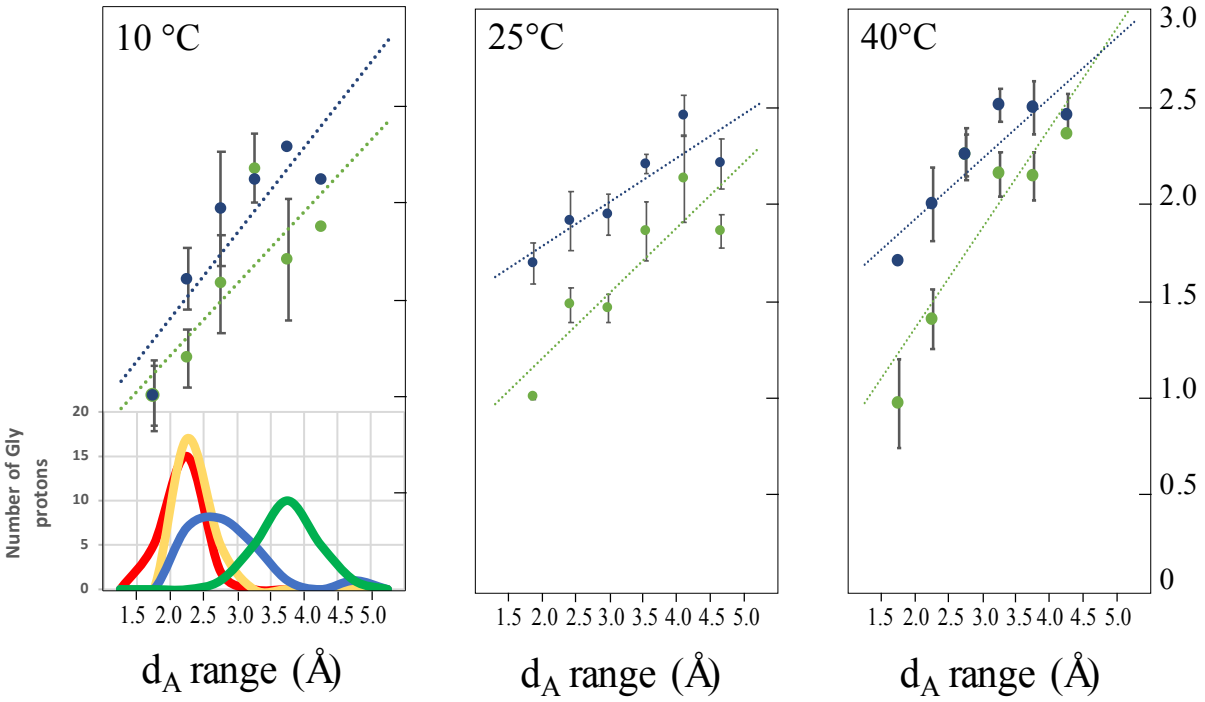


Figure 4.25: Graph of S/N and resolution enhancements for diastereotopic protons of glycines displayed in function of the d_A. Underneath the left panel is represented the distribution of the d_A for each type of labeling. The green dots are S/N enhancements, the dark blue dots the resolution enhancements. The curves count the number of protons of methylenes of glycines in function of their d_A. In red, the number of protons for uniform labeling in H_2O, in yellow the number of protons glycine specifically labeled in H_2O, in blue the number of protons for uniform labeling in D_2O and in green the number of protons for glycine specific labeling in D_2O.

In light of these results, the CH_2-TROSY experiment demonstrates clear enhancements both to resolution in the proton dimension and sensitivity when applied to a highly deuterated sample. Glycine residues are often implicated in more dynamic regions of the protein and the possibility to record more resolved spectra is valuable for a precise analysis of these amino acids. In the following, we provide an application of the CH_2-TROSY using a partly deuterated sample. As demonstrated, one of the benefits of these enhancements can be an advantage for aliphatic titration experiments.

Monitoring protein interactions in the ^{1}H and ^{13}C dimensions has certain advantages, notably because of the sensitive nature of Hα and Cα to secondary structure changes. Surface mapping using CSP of these groups give much more information about structural changes than the HN groups, which tend to over-estimate interactions with long-chain residues and under-estimate interactions with smaller residues, such as glycine[40]. It has also been suggested that CSP guided docking gives best results with aliphatic CSP[41]. Moreover, it enables to perform titrations at high pH, high temperatures, and in D_2O, which is difficult or impossible using HN groups. Based on previous works,[42] we have developed with our collaborators a new ligand,

CZ181 (figure 4.26), possessing a C-N bond that mimics the transition state of the pS-P motif. This ligand possesses an affinity in the tens of micromolar range (Kd = 14.4 ± 2.6 µM) for the catalytic domain PPIase as measured by ^{1}H-^{15}N titration. To display the interest of the CH_2-TROSY experiment, even without a specifically glycine labeled sample, the PPIase domain of Pin1 was produced in *E.coli* (see Chapter II for expression), uniformly ^{15}N-^{13}C labeled. Titration was monitored with increasing concentrations of CZ181 ligand by CH_2-TROSY spectra of the methylene of glycines in D_2O. Titration was also undertaken on ^{1}H-^{15}N HSQC spectra for the amide groups in H_2O. The CH_2-TROSY titration spectrum displaying chemical shift perturbation (CSP) for 0.2, 0.4, 0.8, 1.2 , 2 and 4 equivalents of ligand is shown in Figure 4.26. The CSP observed are of a maximum of 0,1 ppm shifts in the ^{1}H dimension. With such small shifts, titration using a traditional CH_2-HSQC experiment would be greatly hampered by the width of correlations. As illustrated in the box in figure 4.27, the CH_2-TROSY experiment enables the experimenter to gain significant resolution in the proton dimension compared to a regular CH_2-HSQC spectrum, hence facilitating peak picking and the very fine tracking of the correlations (see IV.4.10 for recording conditions).

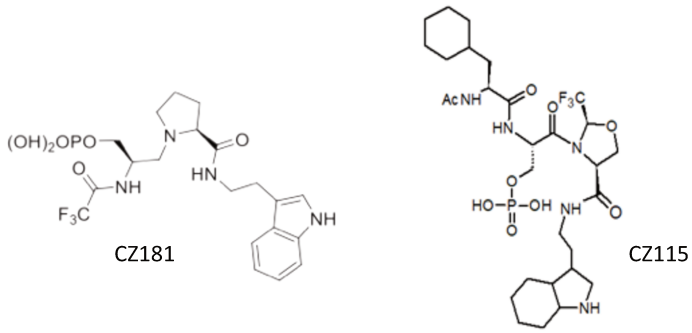


Figure 4.26 : Illustration of the two peptide ligands used for the titration experiments on the PPIase domain.

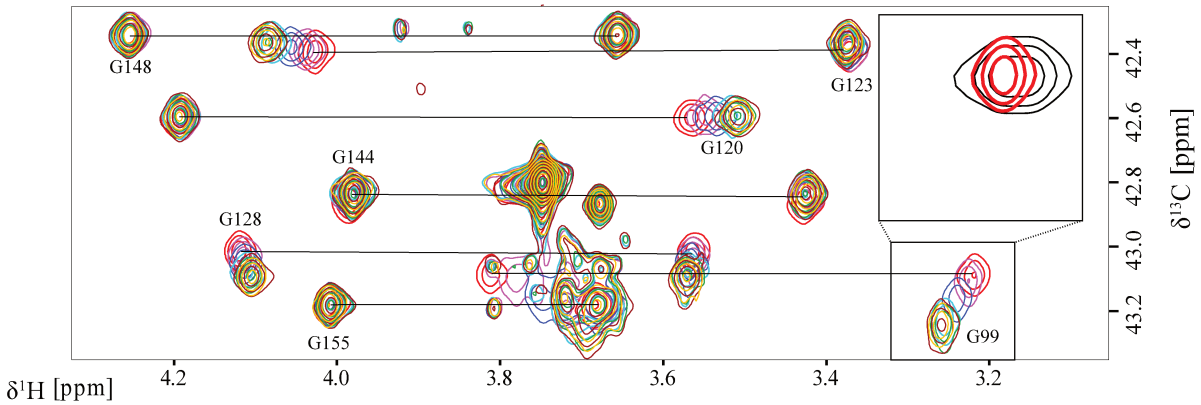


Figure 4.27 : CH_2-TROSY spectra of PPIase Pin1 domain uniformly ^{13}C-^{15}N labeled, 1mM, 30 mM Tris pH 7, 50 mM NaCl, 2 mM DTT, 100% D_2O, 25°C, 950 MHz cryoprobe with increasing concentrations of CZ181; 0 mM (red) 0.2 mM, (pink), 0.4 mM (blue), 0.8 mM (cyan), 1.2 mM (green), 2 mM (yellow) and 4 mM (brown). The framed correlations represent the overlapping upfield diastereotopic proton of glycine 99, with the CH_2-TROSY experiment in red and CH_2-HSQC experiment in black. The processing parameters were the same for both spectra.

CSP values for the diastereotopic protons of glycine residues were measured by measuring in both the ^{1}H and ^{13}C dimensions the difference in chemical shift between the conditions without ligand and with the maximum equivalent of ligand. The values were weighted by a factor of 1/3.6 in the carbon dimension to take into account the dynamic range variability of chemical shifts compared to the proton. Values are displayed in figure 4.28. There are 2 values for each glycine, corresponding to each of the diastereotopic protons. It is interesting to note that diastereotopic protons of one glycine do not necessarily present similar behavior, hence do not display comparable CSP. G120, for example, displays a high CSP for its upfield diastereotopic proton but not for the downfield proton, which hardly moves. For G123, the upfield proton presents a high CSP whereas the downfield one does not. This suggests that only one proton is perturbed by the presence of the ligand. G128 and G99 present a relatively high CSP for both protons, with the upfield proton of G99 particularly perturbed by the presence of the ligand. G144, G148 and G155 do not present significant CSP in the Hα or Cα dimensions, suggesting they are not perturbed by the presence of the ligand. G91 does not have diastereotopic protons with a sufficient chemical shift difference and is not represented in the spectrum as its chemical is much farther from the other glycines, at 45 ppm in the ^{13}C dimension. Adding 4 equivalents of CZ181 ligand induced only a slight chemical shift perturbation, with 0.015 ppm shift of its downfield proton.

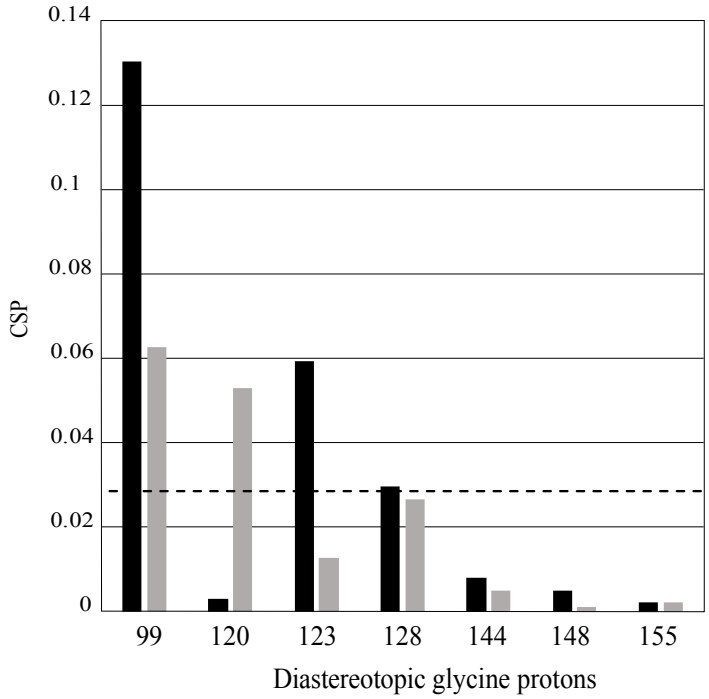


Figure 4.28 : ^{1}H-^{13}C CSP for residues of the PPIase domain of Pin1 titrated with 4 mM of CZ181 ligand. The black bars represent downfield diastereotopic protons whereas the grey bars are the upfield diastereotopic protons. The dotted line, measured as the mean of all the CSP values found, is the determined limit for which a CSP is significant, equal to 0.029.

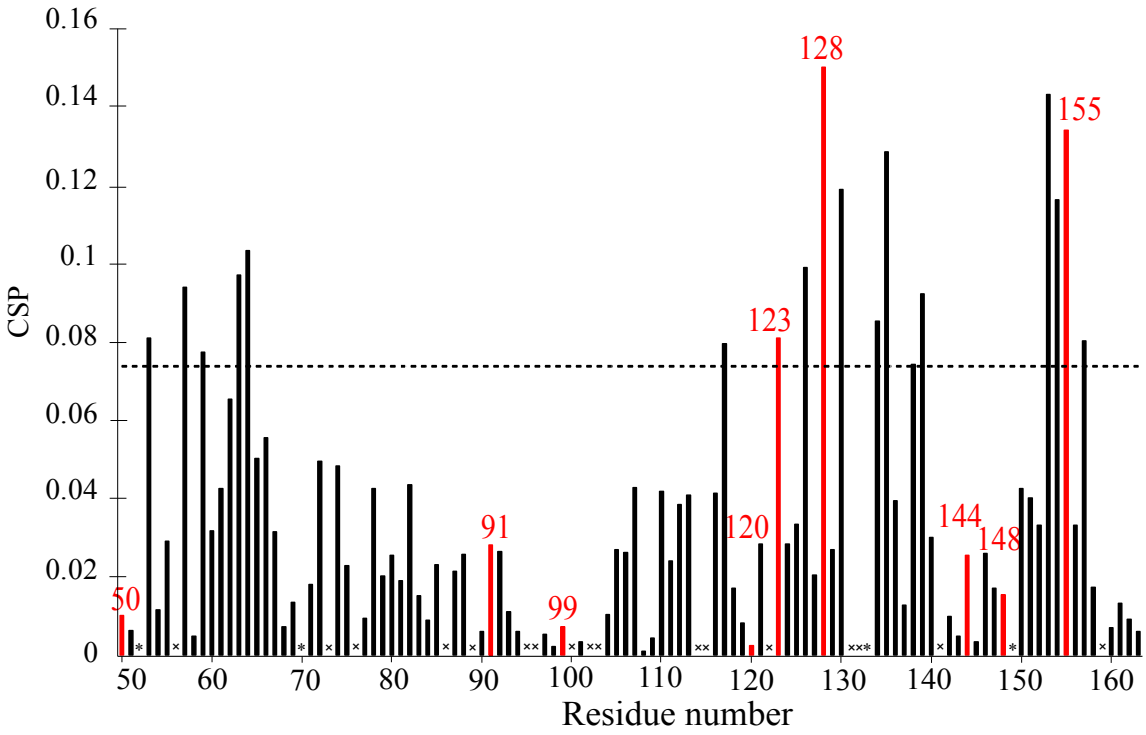


Figure 4.29: ^{1}H-^{15}N CSP for residues of the PPIase domain of Pin1 titrated with 4 mM of CZ181 ligand. The red bars represent glycine residues, the stars are proline residues for which there are no correlations and the crosses represent residues for which titration was not possible either because of overlap or the correlation was not available. CSP considered significant are above 0.073, calculated as the mean of all CSP + 1 standard deviation.

From this data, we can observe 4 glycines that have an important CSP (>0,03) for at least one of the diastereotopic protons when increasing concentrations of CZ181 are added: G99, G120, G123, G128. These values can then be compared to the CSP of a HN titration experiment, to see if there is any correspondence. A histogram of the CSP observed for all the correlations of ^{1}H-^{15}N HSQC spectra of the PPIase domain titrated with the same equivalents of ligand is displayed in figure 4.29. The values were weighted by a factor of 1/10 in the nitrogen dimension to take into account the dynamic range variability of chemical shifts compared to the proton. Three regions are clearly influenced by the presence of the ligand, as they display significantly higher CSP, from residues 53 to 64, 117 to 138 and 153 to 157. Notably, 3 amides of glycine residues are highly affected by the presence of the ligand according to this data, G123, G128 and G155. G50, G91, G99, G120, G144 and G148 display low to very low CSP in the HN titration. Residues of the catalytic site of Pin1 are generally considered to be D112, C113, K117, G123, M130, K132, E135, S154, G155 (Chapter II), most of these residues display high perturbation when in the presence of a ligand because of direct contact with it. In addition, the perturbations observed for residues located in the WW domain

arise from a reorganization of the PPIase/WW interface upon binding, such an allosteric effect being responsible for a fine and reciprocal regulation of the two Pin1 domains (Chapter II).

Overall, the CSP measured on the methylene groups of glycines present some large difference with the titration based on amide correlations. Indeed, G99 and G120, which displayed at least one proton with high CSP in the CH_2-TROSY spectra show almost no perturbation by the ligand in the HN TROSY spectra. This is intriguing, meaning that either one of the spectra is over or under-evaluating the effect of the ligand on this residue. Also, this could suggest that the residue is partly affected by the presence of the ligand, that the amides are not affected by its presence but the methylene is. Contrariwise, G155, which displayed almost no perturbation in the CH_2-TROSY spectra, shows an extremely high CSP in the HN spectra. This, again, could suggest that the HN group of this residue which belongs to the catalytic site directly faces the ligand whereas the methylene group is oriented in another direction. However, not all data are diverging, G123 and G128 both display perturbation of at least one proton in the CH_2-TROSY spectra and perturbation of the amide group. This could suggest that these residues are particularly affected by the presence of this ligand. G144 and G148 show no CSP in either spectra, suggesting they are not affected by the ligand.

To further compare these two different titration experiments, Kds for the amino acids displaying high CSP were measured for the ^{1}H-^{15}N and ^{1}H-^{13}C spectra. The curve fits for CH_2-titration are displayed on the next page (Figure 4.30) as well as for four residues of the HN titration. The other curve fits for ^{1}H-^{15}N spectra titration are available in the Annex 8. The Kds vary significantly between different curve fits, including within the two protons of a same methylene group. In that regard, such titrations may benefit from additional data points to constrain the fits, especially between 0 and 1 equivalent. The mean Kd calculated for CH_2-TROSY titration is 18.54 = μM if the values of G128 which diverge considerably are discarded. From the HN titration, a mean Kd of 17.13 μM when all the CSP values of the PPIase domain are taken into account.

Although titration using only the methylene of glycine residues does not present as many values as for titration using the amides of all residues, they can give some insight into binding affinity and dynamics of the backbone, providing supplementary structural information on protein interactions.

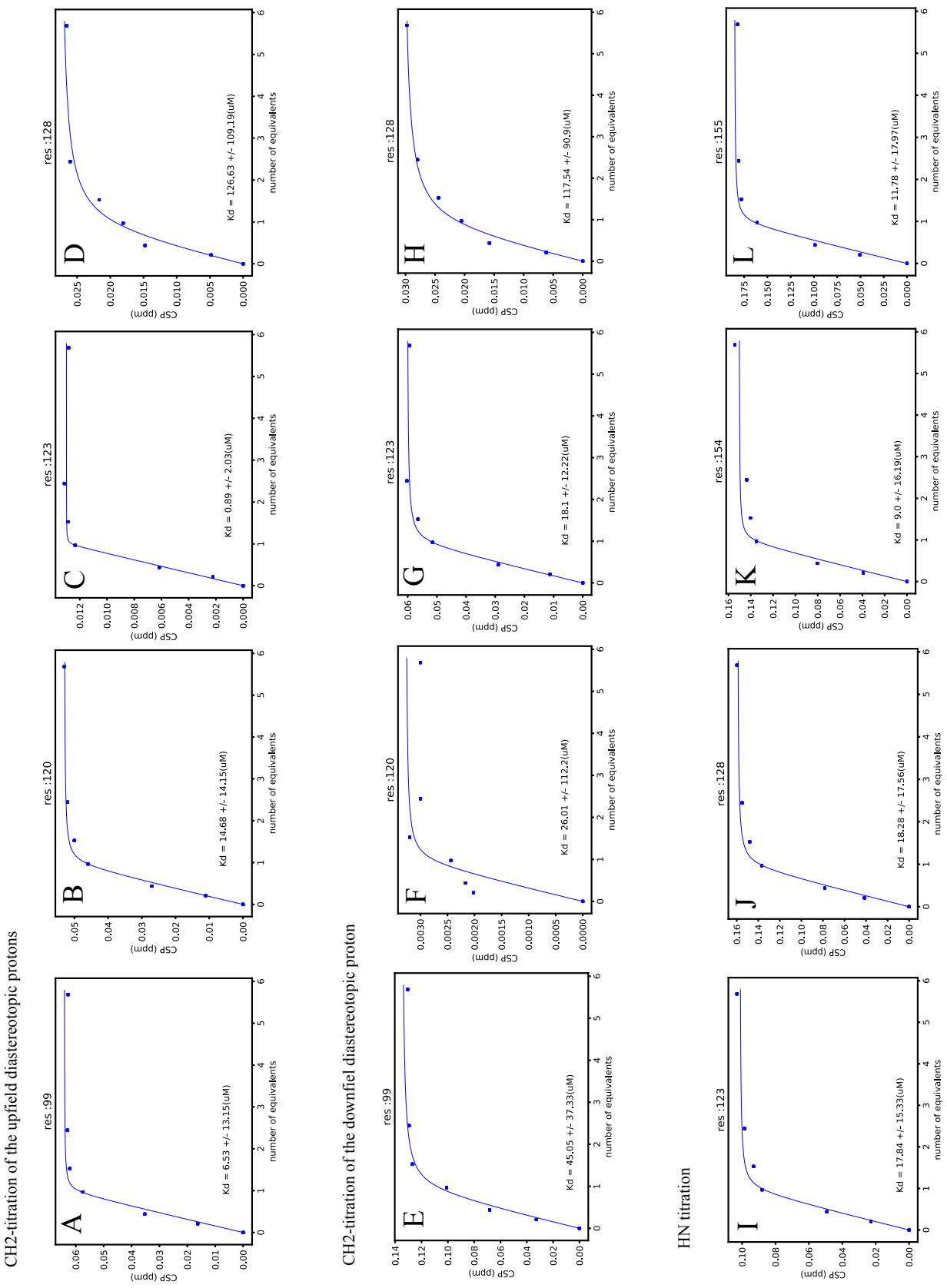


Figure 4.30: Curve fits for the glycine residues titrated against CZ181. A-H: Fits of methylene groups derived from CSP measured in CH_2-TROSY spectra. I-L: Fits of amide groups derived from CSP measured in ^{1}H-^{15}N TROSY spectra.

When looking at the molecular structure of the PPIase domain, and the position of the glycine residues within the structure, it is possible to understand how G123 and G128 would be perturbed by the presence of the ligand as they are part of the catalytic pocket (figure 4.31). But G120 and G155 are also in close proximity to the catalytic site, with G120 having high CSP of one of the diastereotopic protons of the methylene after ligand binding and G155 having high CSP of the amide.

The docking pose can elucidate this difference, by simulating low energy complexes between the ligand and the protein, using Autodock Vina[48] (figure 4.32) see Material and Methods IV.4.11. It is interesting to notice that an Hα of the methylene of G120 points directly towards the indole of the tryptamine of the ligand which creates strong anisotropy, perhaps explaining why one of the diastereotopic protons is considerably influenced by the presence of the ligand and not the other. On the other hand, the amide of G120 makes a hydrogen bond with the CO of the residue -3 explaining why it isn't perturbed by the presence of the ligand. The amide is strongly perturbed for G155 as it points directly towards the ligand, whereas its methylene points downwards, away from the ligand, thus providing an explanation for the difference in CSP observed between techniques.

G144, G148 and G99 are on the opposite side of the catalytic pocket, and whereas 144 and 148 don't display high CSP in any technique, it is curious that G99 would display such high CSP, compared to the other glycines, notably in the Cα dimension (figure 4.27) and going downfield. This suggests perhaps that there are some long range structural changes that occurred with the ligand binding. Moreover, G99 is at the end of an α helix, the high CSP suggest that perhaps what we observe is a rigidification of this residue into the α helix when the ligand binds the catalytic site.

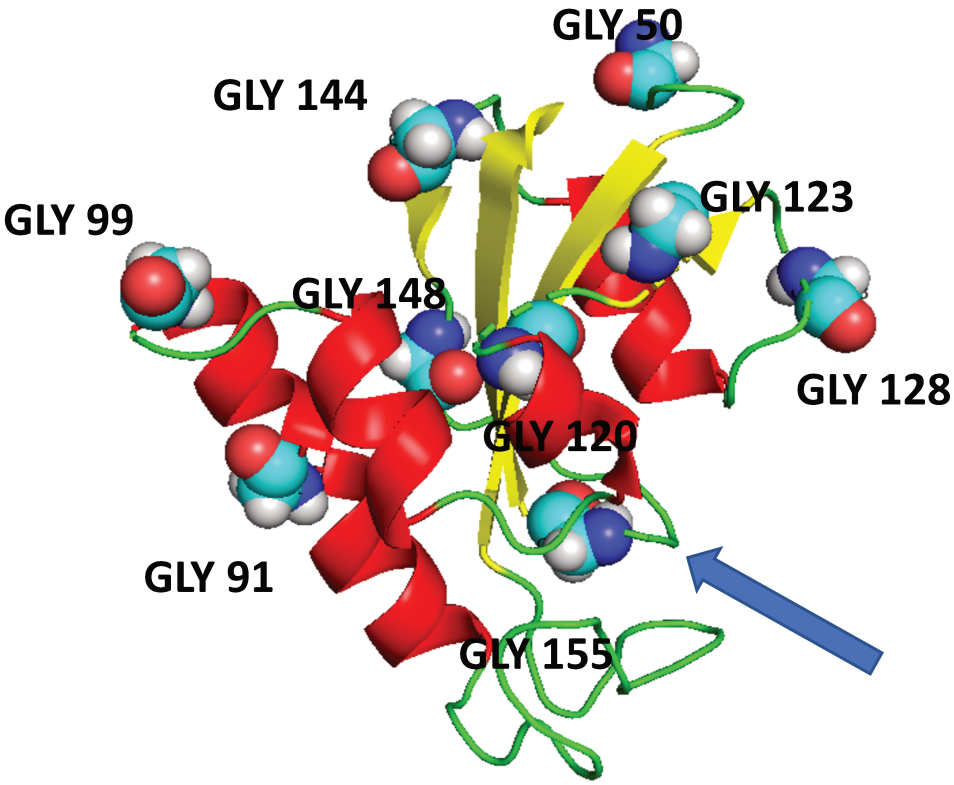


Figure 4.31: Crystallographic structure of the PPIase domain of Pin1 (PDB Entry: 1PIN, resolution of 1.35 Å)[44]. The glycines of the PPIase domain are tagged and displayed in the CPK mode. The catalytic pocket is made up, amongst other residues, of G123, G128 (see blue arrow).

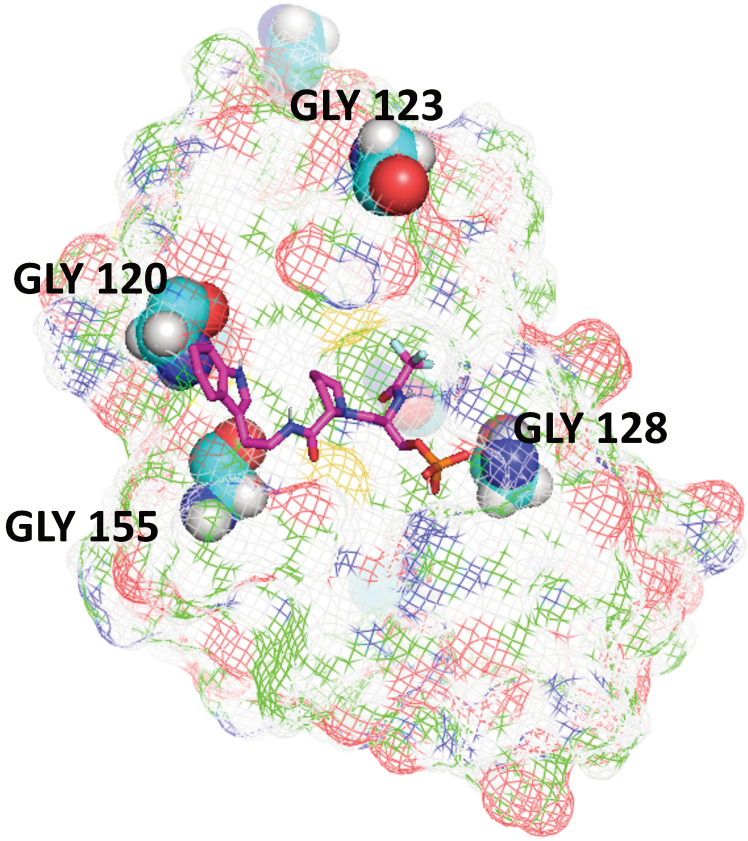


Figure 4.32: Docking pose for PPIase domain of Pin1 (PDB Entry: 1PIN, resolution of 1.35 Å)[44] and the CZ181 ligand generated using AutodockVina[48].

IV.3.2. S3-Gly experiments

It is possible to extract a number of coupling constants of glycine residues from spin state selective experiments to obtain structural and dynamic information. A spin state selective-gly-HACACO experiment (S3Gly-HACACO) has been designed based on the CH_2-TROSY experiment (figure 4.33). This 3D experiment is an "out and back" pulse scheme, starting from the Hα protons of the glycine residues. By the mean of two INEPT blocs, the magnetization is transferred from the Hα to the Cα with a selective inversion pulse, and from the Cα to the carbonyl carbon C'. The magnetization is then allowed to evolve during t_1, which allows to record chemical shifts of the C' carbons, without any decoupling pulse, neither for Cα nor for Hα. The magnetization is then transferred back to the Cα which are labeled in the same time using a $1/(2\times{}^1J_{C\alpha C'})$ constant time evolution period. Finally, the double INEPT transfer, using optimized delays as in the CH_2-TROSY experiment, allows the magnetization to come back on the Hα. Then, the acquisition is started without any decoupling.

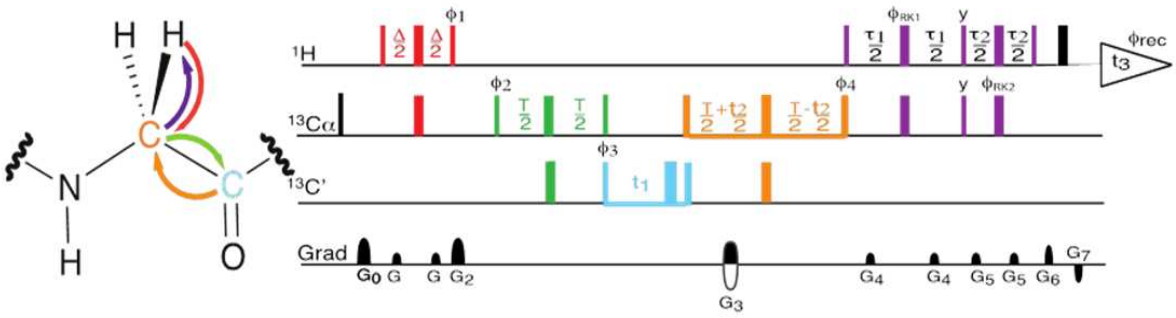


Figure 4.33 : « spin states selective » pulse sequence (S3-HACACO). A color code is used for which each period of the sequence is associated to a magnetization transfer on the glycine residue. The thin and thick bars correspond to 90 and 180° pulses respectively. When not specified, the pulse phase is x. Delays: $\Delta = 1/(2\times J_{C\alpha H\alpha})$; $T = 1/(2\times J_{C\alpha C'})$; $\tau 1 = 0.34/J_{C\alpha H\alpha}$ et $\tau 2 = 0.23/J_{C\alpha H\alpha}$. $\Phi_1 = y : \Phi_2 = y, -y ; \Phi_3 = x, x, -x, -x ; \Phi_4 = x, y, -x, -y$ $\Phi_{RK1} = x ; \Phi_{RK2} = x$. The carbon pulses are rectangular pulses specific to either C^α or C'.

From this experiment, one can obtain a number of precise ^1J and ^2J coupling constant measures between the TROSY and anti-TROSY correlations, and stereospecific to the diastereotopic protons. Spectra were recorded at 40°C on U-^{15}N-^{13}C-^{1}H labeled samples of PPIase domain in 30 mM Phosphate, 50 mM NaCl, 1 mM DTT, 0.5% DSS, 0,05% NaN_3, 1 mM EDTA, 100% D_2O at 950 MHz. The PPIase domain of the Pin1 protein (500 µM) was introduced in a 5 mm Shigemi tube (see Chapter IV.4.11. for recording conditions). The figure 4.34 represents the projections observed for the Gly 148 system. We can see that the coupling constants appear several times in the different projections. The numerous possible measures of

these coupling constants enable us to increase considerably the precision with which they are measured. The Table 4.1 indicates the average coupling constants found for the seven diastereotopic glycines of the PPIase domain. By default, the Hα2 proton is defined as the upfield resonance and Hα3 the downfield, although the stereospecific assignment has not be performed. It is striking to notice the difference in intensity of the correlation the most downfield in the ^{13}C dimension and upfield in the ^{1}H dimension, compared to the correlation the most upfield in the ^{13}C dimension and downfield in the ^{1}H dimension. This was however expected as the first correlation retained by the spin state selective experiment corresponds to the TROSY component, whereas the second is not optimized in term of relaxation. The position of each correlation was obtained using the NMRPipe[43] peak picking module. The coupling constants correspond to the differences in frequency between the peaks. Taking into account the linewidth and the signal to noise ratio, an error can be deduced for each coupling measurement.

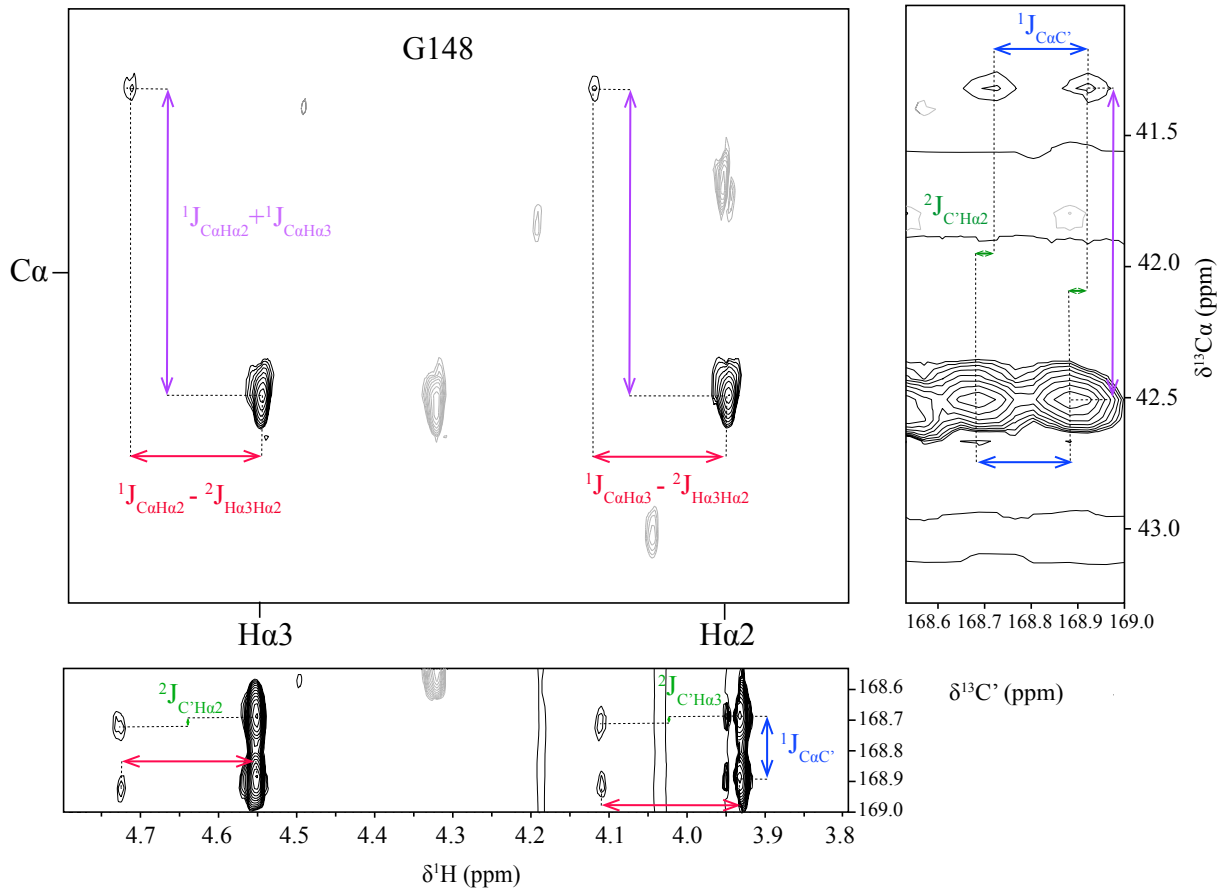


Figure 4.34: Projections in three planes of the s3-HACACO recorded on U-^{15}N-^{13}C-^{1}H labeled samples of PPIase domain 500 µM in 30 mM Phosphate, pH 7.4, 50 mM NaCl, 1 mM DTT, 0.5% DSS, 0,05% NaN_3, 1 mM EDTA, 100% D_2O, 950 MHz, Shigemi 5 mm tube.

	G148	G123	G99	G128	G155	G144	G120
$^{1}J_{C\alpha H\alpha 2} + {}^{1}J_{C\alpha H\alpha 3}$ (4)	**279,22**	**278,82**	**283,29**	**288,01**	**285,72**	**281,81**	**283,37**
Error	0,22	0,26	0,23	0,18	0,15	0,14	0,15
$^{1}J_{C\alpha H\alpha 2} - {}^{2}J_{H\alpha 3 H\alpha 2}$ (2)	**158,80**	**160,35**	**159,32**	**158,49**	**163,49**	**162,28**	**161,60**
Error	0,19	0,17	0,21	0,17	0,22	0,14	0,16
$^{1}J_{C\alpha H\alpha 3} - {}^{2}J_{H\alpha 3 H\alpha 2}$ (2)	**163,24**	**156,83**	**158,56**	**164,26**	**163,03**	**163,36**	**161,97**
Error	0,20	0,19	0,24	0,16	0,15	0,14	0,18
$^{2}J_{C'H\alpha 2}$ (2)	**8,35**	**9,26**	**7,65**	**6,75**	**12,48**	**12,79**	**12,82**
Error	0,19	0,17	0,21	0,17	0,22	0,14	0,16
$^{2}J_{C'H\alpha 3}$ (2)	**8,00**	**10,92**	**6,01**	**9,55**	**10,66**	**10,65**	**9,39**
Error	0,43	0,50	0,53	0,38	0,35	0,34	0,37
$^{1}J_{C\alpha C'}$ (4)	**47,68**	**50,67**	**51,72**	**52,34**	**52,83**	**51,75**	**53,54**
Error	0,26	0,27	0,34	0,23	0,21	0,20	0,20
$^{1}J_{C\alpha H\alpha 2}$	**137,25**	**141,39**	**141,94**	**140,95**	**143,13**	**140,49**	**141,36**
Error	0,61	0,62	0,68	0,51	0,51	0,42	0,49
$^{1}J_{C\alpha H\alpha 3}$	**141,66**	**137,68**	**141,18**	**146,67**	**142,64**	**141,58**	**142,18**
Error	0,61	0,62	0,68	0,51	0,51	0,42	0,49
$^{2}J_{H\alpha 3 H\alpha 2}$	**21,57**	**19,08**	**17,39**	**17,59**	**20,46**	**21,78**	**20,33**
Error	0,61	0,62	0,68	0,51	0,51	0,42	0,49

Table 4.1: Table of weighted average of measurements and errors for the coupling constants of seven diastereotopic glycines of the PPIase domain. Error bars represent the square root of the sum of the width and height of correlations divided by the noise level. The $^{1}J_{C\alpha H\alpha 2}$, $^{1}J_{C\alpha H\alpha 3}$ and $^{2}J_{H\alpha 2 H\alpha 3}$ couplings are deducted from the previously measured couplings constants, their error bars are the sum of the errors of each coupling used. The numbers in parenthesis next to the coupling constant measured is the number of measures that could be done from the spin state selective spectrum.

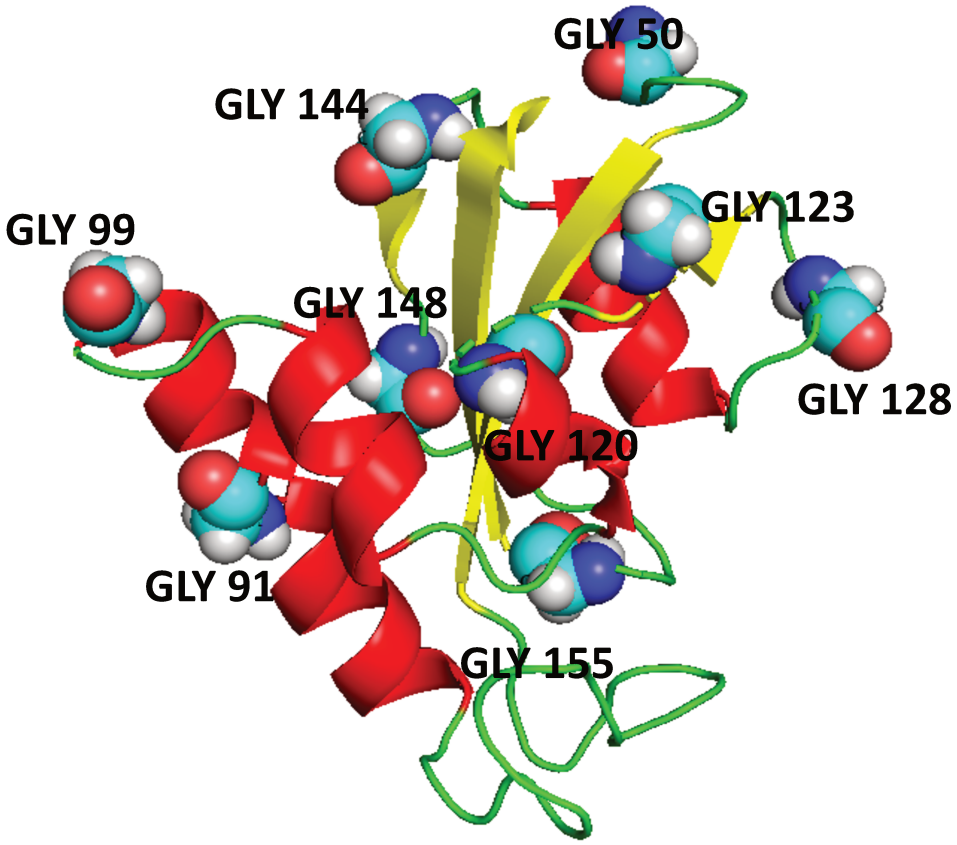


Figure 4.35: Crystallographic structure of the PPIase domain of Pin1 (PDB Entry: 1PIN, resolution of 1.35 Å)[44]. The glycines of the PPIase domain are tagged and displayed in the CPK mode.

For each coupling constant, there is notable variability found for the different glycines, and for certain couplings there is intra-variability within the residue. The constants are measured with low error rates (maximum of 0,68 Hz). For residues G99, G144, G120 and G155, found at the extremities of structural elements (beginning of β strands, except for G99 found at the end of an α-helix, see crystallographic structure figure 4.35), the $^{1}J_{C\alpha H\alpha}$ couplings are similar: 141.94 and 141.18 Hz, 141.36 and 142.18 Hz, 140.49 and 141.58 Hz, and 143.13 and 142.64 Hz respectively. Whereas differing values would suggest a glycine in a rigid structural element as the magnetic environment of the two protons are singular, similar values suggest the protons have a comparable magnetic environment and that molecular movements may average the coupling constant values. For G120, G144 and G155, the $^{2}J_{C'H\alpha}$ coupling constants are near 12,5 and 10 Hz whereas for G99 this coupling constant is much lower: 7,65 and 6,01 Hz. The location of this residue on the top of a helix could singularize these ^{2}J couplings that experience a two-fold variations. Taken together, this suggests that coupling constant values could be characteristic of the structural element in which the glycine is found. Noteworthy, G123 and G148 have $^{1}J_{C\alpha H\alpha}$ coupling constants that are very dissimilar: 141.39 and 137.68 Hz for G123 and 137.25 and 141.66 for G148. G123 is found in the β2-strand and G148 in the middle of β3.

These measurements then reflect that these glycines are in much more structured and rigid elements, the two protons have singular environments. Interestingly, such a $^{1}J_{CH}$ dissymmetry was already used to perform the stereospecific assignment of diastereotopic protons in α-helices of peptidomimetics, with values of (135.72±0.76) and (137.83 ± 0.87) Hz for the (pro-*R*) and the (pro-*S*) methylene protons, respectively[45] or on RNA helix, where $C^{5'}H_2$ methylene displayed $^{1}J_{CH5''}$ values (pro-*R*) systematically lower than $^{1}J_{CH5'}$ (pro-S), with average values of (147.17 ± 0.53) Hz and (149.80 ± 0.83) Hz, respectively[46]. The $^{2}J_{C'H\alpha}$ coupling constants of G123 and G148 were however are rather alike, 9,26 Hz and 10.92 Hz for G123 and 8.35 Hz and 8.00 Hz for G148. Moreover, their $^{1}J_{C\alpha C'}$ are lower than of the other glycines (47.68 Hz and 50.67 Hz respectively compared to more than 51 Hz for the previously mentioned glycines). The residue G91 is also in a secondary structure element, α1, however it does not present two diastereotopic protons with a significant chemical shift difference. Hence, it precludes coupling measurements. This low chemical shift difference for $H^{\alpha 2}$ and $H^{\alpha 3}$ of G91 was not expected as this residue is located on the external face of the helix, therefore in a supposed anisotropic environment created by this secondary structure element. This suggests that the chemical shifts of the glycine protons are also influenced by the anisotropy created by the whole 3D structure, and are, in that respect, very sensitive probes of the structural events. Finally, G128 is situated in a loop between β3 and α4. Its $^{1}J_{C\alpha H\alpha}$ coupling constants displayed the biggest variability with almost 6 Hz of difference. This is an indication of restricted motions for this part of the loop, reinforced by a chemical shift difference of about 0,6 ppm for the two diastereotopic protons. It should be highlighted that this residue is next to the active site of the PPIase enzyme and was the most affected by the ligand binding in the ^{1}H-^{15}N titration.

A second s3-Gly HACACO experiment was undertaken with 4 equivalents of CZ115 ligand (an analogue ligand of CZ181, see Annex 9 for synthesis). Coupling constants have also been measured and are presented on the next page (Table 4.2). With the ligand, the errors are more elevated than before as a consequence of a decreased signal to noise ratio. The Table 4.3 shows the absolute difference in coupling constants between the experiment without and with 4 equivalents of CZ115 ligand for each glycine. The differences were categorized into three scenarios:

- Glycines where coupling constants changed a lot (>3Hz) G123 G128 G99 and G148
- Glycines where some coupling constants change significantly (between 2 and 3 Hz) G120 and G144

- Glycines where coupling constants change very little (<2Hz) G155

4 eq CZ115	G148	G123	G99	G128	G155	G144	G120
$^{1}J_{C\alpha H\alpha 2} + {}^{1}J_{C\alpha H\alpha 3}$ (4)	**282,54**	**277,07**	**281,63**	**284,92**	**285,35**	**281,71**	**285,97**
Error	0,29	0,23	0,34	0,30	0,22	0,20	0,24
$^{1}J_{C\alpha H\alpha 2} - {}^{2}J_{H\alpha 3H\alpha 2}$ (2)	**165,79**	**162,09**	**156,64**	**163,34**	**162,04**	**160,19**	**160,24**
Error	1,01	0,62	1,35	1,06	0,84	1,01	0,79
$^{1}J_{C\alpha H\alpha 3} - {}^{2}J_{H\alpha 3H\alpha 2}$ (2)	**161,52**	**165,55**	**161,18**	**179,50**	**165,88**	**168,64**	**165,61**
Error	1,14	0,70	1,33	1,05	0,74	0,79	0,74
$^{2}J_{C'H\alpha 2}$ (2)	**4,47**	**10,92**	**11,29**	**10,54**	**12,44**	**13,41**	**9,90**
Error	1,01	0,62	1,35	1,06	0,84	1,01	0,79
$^{2}J_{C'H\alpha 3}$ (2)	**6,44**	**6,79**	**11,65**	**16,97**	**9,49**	**12,01**	**10,42**
Error	0,61	0,50	0,77	0,59	0,54	0,47	0,58
$^{1}J_{C\alpha C'}$ (4)	**48,32**	**48,60**	**48,95**	**50,71**	**51,02**	**50,12**	**51,10**
Error	0,40	0,28	0,52	0,39	0,34	0,30	0,33
$^{1}J_{C\alpha H\alpha 2}$	**143,45**	**135,79**	**138,60**	**135,54**	**141,10**	**138,78**	**140,13**
Error	2,44	1,54	3,01	2,41	1,80	2,00	1,77
$^{1}J_{C\alpha H\alpha 3}$	**139,57**	**140,40**	**142,71**	**149,26**	**144,05**	**143,18**	**145,80**
Error	2,44	1,54	3,01	2,41	1,80	2,00	1,77
$^{2}J_{H\alpha 3H\alpha 2}$	**21,99**	**25,67**	**17,88**	**27,49**	**21,14**	**25,97**	**19,83**
Error	2,44	1,54	3,01	2,41	1,80	2,00	1,77

Table 4.2 : Table of weighted average measurements and errors for the coupling constants of seven diastereotopic glycines of the PPIase domain with 4 equivalents of CZ115 ligand. Error bars represent the square root of the sum of the width and height of correlations divided by the noise level. The $^{1}J_{C\alpha H\alpha 2}$, $^{1}J_{C\alpha H\alpha 3}$ and $^{2}J_{H\alpha 2H\alpha 3}$ couplings are deducted from the measured couplings constants, their error bars are the sum of the errors of each coupling used.

4 eq CZ115	G148	G123	G99	G128	G155	G144	G120
$^{2}J_{C'H\alpha 2}$	3,88	1,66	3,64	3,79	0,04	0,62	2,92
$^{2}J_{C'H\alpha 3}$	1,56	4,13	5,65	7,42	1,17	1,36	1,04
$^{1}J_{C\alpha C'}$	0,65	2,07	2,77	1,63	1,82	1,62	2,44
$^{1}J_{C\alpha H\alpha 2}$	6,19	5,60	3,34	5,41	2,03	1,71	1,23
$^{1}J_{C\alpha H\alpha 3}$	2,09	2,72	1,53	2,59	1,42	1,59	3,62
$^{2}J_{H\alpha 3H\alpha 2}$	0,41	6,59	0,48	9,90	0,68	4,18	0,49

Table 4.3: Absolute difference in coupling constant measurements between PPIase with and without 4 equivalents of CZ115. In orange are highlighted the values for which >3 Hz are seen, and in light yellow between 2 and 3 Hz difference.

As ^{1}J and ^{2}J couplings evolve in function of φ and ϕ angles, these differences in coupling constants were correlated to the CSP found with the CZ181 ligand to see if there were structural changes that could be correlated to ligand binding. And we can draw some parallels between them, with strong CSP corresponding to increased differences in ^{1}J and ^{2}J coupling, as is the case for G123 and G128, in the catalytic site. Glycine 155 is at the C-terminal end of the protein, in β4, and its coupling constants are not affected by the ligand CZ115 even though D153, S154 and H157 are known as the residues surrounding the peptidyl prolyl bond of the ligand. The CSP were also very low for the methylene group of this glycine residue whereas the ^{1}H-^{15}N titration indicated significant chemical shift perturbations. This suggests that the amide proton of G155 may be involved directly in the ligand binding without creating structural and dynamic change at the residue level. G144 is not part of a binding region and has only one affected constant $^{2}J_{H\alpha 2H\alpha 3}$. G99 at the end of α helix 1, is also far from the ligand binding site, but its coupling constants $^{1}J_{C\alpha H\alpha 2}$, $^{2}J_{C'H\alpha 2}$ and $^{2}J_{C'H\alpha 3}$ change significantly after ligand addition. The $^{2}J_{H\alpha 2H\alpha 3}$ and $^{1}J_{C\alpha H\alpha 3}$ change very little (<2Hz) and the $^{1}J_{C'C\alpha}$ diminishes slightly (2.77 Hz). These strong differences in coupling constants measures, notably for $^{1}J_{C\alpha H\alpha 2}$, $^{2}J_{C'H\alpha 2}$ of one of the protons shows strong dissymmetry, a hallmark of helical structures, and could be further evidence that ligand binding induces long range local structural changes. G120 coupling constants are not affected by ligand binding although it is close to the proline binding site (G123, Q129, M130). The $^{1}J_{C\alpha H\alpha 3}$ was affected (>3Hz) but that is the only coupling that displays important changes with the $^{2}J_{C'H\alpha 2}$ and $^{1}J_{C'C\alpha}$ showing a slight change. G120 and G99 are interesting because both are not directly involved in ligand biding but they have certain coupling constants that are clearly affected by its presence. In comparison, G128 is also close but not directly implicated in the binding to the ligand proline, however it has important changes in coupling constants. Indeed, both the $^{1}J_{C\alpha H\alpha 2}$ and $^{1}J_{C\alpha H\alpha 3}$ are affected, although not to the same degree (>5 Hz and 2.59 Hz). Both $^{2}J_{C'H\alpha 2}$ and $^{2}J_{C'H\alpha 3}$ show important change (>3 Hz) and the $^{2}J_{H\alpha 2H\alpha 3}$ displays the most important change of all the measurements. This is a strong indication of major changes in the loop conformation upon binding. G123 is directly implicated in the pocket that binds the proline and displays significant changes in coupling constants (>4 Hz for the $^{2}J_{C'H\alpha 3}$, $^{1}J_{C\alpha H\alpha 2}$ and $^{2}J_{H\alpha 2H\alpha 3}$). Whether these changes in coupling constants are significant in the general structure of the protein remains to be explored, the determination of the complex structure will be of prime value in that respect. However, these data clearly show how ligand binding can induce long range structural changes through extensive measurements performed on each glycine residues, acting as local probes.

However, there are still some issues left with the experiment itself. Notably, when the noise level is enhanced, one can observe phase artifacts which cannot be removed by phasing the spectra (figure 4.36), the resonance taking on a twisted shape. Such a behavior could bias the coupling measurements and efforts have to be made to get rid of these artifacts.

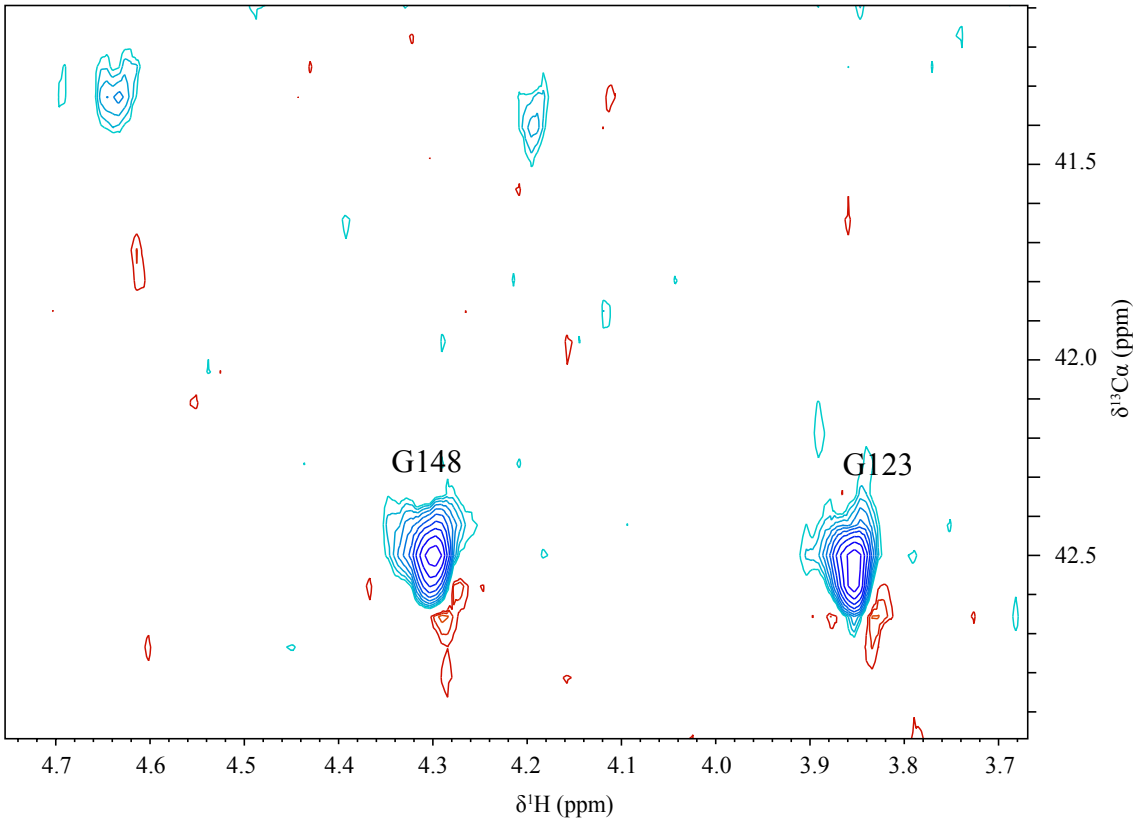


Figure 4.36 : 2D projection of the s3-HACACO recorded on U- ^{1}H ^{15}N-^{13}C labeled sample of PPIase domain 500 µM in 30 mM Phosphate, pH 7.4, 50 mM NaCl, 1 mM DTT, 0.5% DSS, 0,05% NaN_3, 1 mM EDTA, 100% D_2O, 950 MHz, Shigemi 5 mm tube.

IV.3.3. Conclusion and perspectives

Although the CH_2-TROSY experiment does not display the same gains in sensitivity and resolution as other TROSY experiments (notably HN and methyl TROSY), it can provide notable gains in spectral quality compared to regular HSQC experiments, especially using a robust and specific labeling strategy. Labeling methylene groups of glycines in a perdeuterated background is impossible or extremely difficult due to isotopic scrambling using *in vivo* protein expression techniques. Here we have demonstrated that using the cell-free protocol, we can achieve this labeling strategy with relative ease and in satisfactory quantities for these specific NMR experiments. Specific to the methylene of glycine residues, these pulse sequences were not tested for very large τ_c values, and there could be a size limit for proteins analyzed, as a

decrease in gains was observed at lower temperatures here. A perspective that could be envisaged would be to use bigger proteins, enabling a better understanding of the effect of the τ_c on spectral gains and could also provide more glycine residues for supplementary measurements. However, if there are few to no glycines, or they do not present diastereotopic protons, the number of measurements are reduced.

These CH_2-TROSY experiments enabled us to do a titration experiment using in the Cα dimension. Again, if there are few glycines or they do not present diastereotopic protons, the number of possible measurements are reduced but for those that are, they enable analysis on two sets of data, providing extra insight into structure and interaction with other molecules. Moreover, as aliphatic carbons are better predictors of interaction that HN groups, these can be a valuable source of information. In addition, this technique enables to study proteins at high pH, high temperatures and in D_2O. This approach would therefore be a complementary technique to the usual NH titration experiments undertaken, requiring no extra preparation steps than specific ^{13}C labeling.

The CH_2-TROSY pulse scheme can be integrated into 3-dimensional pulse sequences and, by selecting the components of interest, enables the measurement of several coupling constants. Their values can enable finer structural and dynamics studies of proteins, using glycine as a local probe. The preliminary results found with the S3-HACACO experiment on the PPIase protein showed that glycines have distinct one bond ^1J and two bonds ^2J couplings constants that are the reflect of the asymmetry between $H^{\alpha 2}$ and $H^{\alpha 3}$ methylene protons and that can evolve in different manners upon ligand binding.

IV.4. Materials and Methods

IV.4. 1. Cell free protocols for Gly specific H23 labeling

The different cell-free expressions for the H23 samples are given in the table 4.4 with the yields for a 3 mL reaction mixture. Purified S30 extracts were obtained by exclusion chromatography, compounds of less than 100 Da were eliminated.

Buffer	Labeling	S30 p+/p-	+/- inhibitors	Yields (mg per ml of RM)
55mM HEPES in H_2O	Gly $^{15}N/^{13}C/^{1}H$ AA unlabeled	P-	-	1.1
			+	1.3
		P+	-	0.49
			+	0.42
	Gly $^{15}N/^{13}C$, AA $^{15}N/^{13}C/^{2}H$ Celtone ®	P-	-	0.45
			+	0.39
55mM HEPES In D_2O	Gly $^{15}N/^{13}C$, AA $^{15}N/^{13}C/^{2}H$ Celtone ®	P-	-	5,5
			+	1.83
	Gly $^{15}N/^{13}C$, AA ^{2}H Celtone ®	P-	-	1.4
			+ AOA	1.57
		P+	-	0.9
			+ AOA	1.17
55 mM HEPES in H_2O	AA $^{15}N/^{13}C$ Celtone ®	P-	-	0.49
55 mM HEPES in H_2O	Gly $^{15}N/^{13}C$, AA $^{15}N/^{2}H$ Celtone ®	P-	+	0.32

Table 4.4 : Table of H23 cell-free expressions : P- : crude S30 extract, P+ : purified S30 extract, - : no inhibitors used, +: use of the three inhibitors 20 mM AOA, 20 mM DM and $NaBH_4$ treatment. + AOA : addition of only 20 mM AOA. Yields are given as mg of protein per mL of reaction mixture.

Cell-free expression was undertaken following protocols in Chapter III with the specified labeling. After expression, purification oh H23 was undertaken using a nickel affinity resin.

The purification protocol for 3ml of reaction mixture is as follows:

	Buffer 1 pH 7,8	**Buffer 2 pH 7,8**	**Buffer 3 pH 7,8**
Phosphate	20 mM	20 mM	20 mM
NaCl	150 mM	150 mM	150 mM
Imidazole	5 mM	15 mM	350 mM

1ml of Qiagen Nickel Affinity resin was dropped into a column (Empty Poly-Prep® gravity-flow column, Bio-Rad, 731-1550, 0,8 x 4 cm), and left to sediment by gravity flow of surplus liquid. Equilibration with 6ml of Buffer 1 without ever letting the resin dry. Then, flow is stopped by means of a plug and the supernatant is added (3ml) with 3ml of Buffer 1. It is resuspended well for 30 minutes by placing the column on an agitator. The flow-through (FT1) is recuperated and the column washed with 6 ml of Buffer 1 or until OD_{280} is <0.1. This step is repeated with Buffer 2. Fractions are eluted by adding 500 µl of Buffer 3 until OD_{280} is <0.1. The fractions containing the protein of interest H23 are kept and sample concentrations are measured using a Nanodrop with ξ= 9970 L.mol-1.cm-1 and MM = 16,7 kDa.

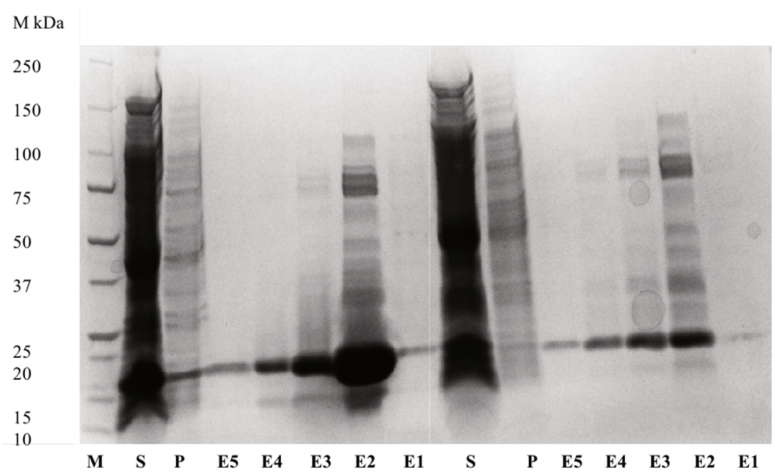


Figure 4.35 : Photo of polyacrylamide gel after H23 ^{1}H-^{15}N-^{13}C-Gly, unlabeled AA cell-free expression and purification on nickel affinity resin. Each deposit was done with 15 ul of sample + 5 ul of 4x bromophenol blue. On the left side of the panel are bands corresponding to H23 expressed in crude S30 extracts, without inhibitors. On the right side of the panel are bands corresponding to H23 expressed in purified S30 lysate, without inhibitors. M : Molecular marker, S: supernatant, P : pellet, E1 → E5 : elution fractions.

An intense band between 20 kDa and 15 kDa is observed in all columns which we can attribute to H23. This suggests that the protein was correctly expressed by cell free and purified although there are other non-specific proteins sill present in the fraction. Indeed, several bands at 100 kDa, 75 kDa and 25 kDa are visible and correspond to other proteins of the reaction mix which have high affinity with the Nickel resin. The protein yields are visually and much lower for the samples expressed with the purified S30 extract.

Concentration of the samples was undertaken simultaneously with the exchange for NMR buffer in D_2O or H_2O, 20 mM Phosphate, 150 mM NaCl. A concentrator with 3kDa cut off membrane was hydrated with distilled water and then buffer. The elution fractions

containing H23 were pooled and concentrated by centrifugation until desired concentration. Buffer change was undertaken with 3 sequential centrifugation steps.

IV.4. 2. Recording conditions for ^{1}H-^{15}N HSQC spectra of H23

For ^{1}H-^{15}N HSQC spectra recordings of H23 protein using a specifically ^{15}N, ^{13}C, ^{1}H-Gly labeled sample and no isotopic label on the other amino acids (^{14}N, ^{12}C, ^{1}H). pH 4,5. Spectra recorded with a protein concentration of 200 µM in shigemi 5mm, Tris HCl buffer 20 mM, NaCl 100 mM, 0,11 mM DSS, 10% D2O, 40°C, 500 MHz spectrometer equipped with a TCI cryoprobe.

H23 ^{1}H, ^{15}N, ^{13}C-Gly and other amino acids bearing deuterium labels (^{2}H-^{14}N-^{12}C), expressed without inhibitors or with 3 inhibitors 20 mM DM, 20 mM AOA and $NaBH_4$ extract treatment. Concentration of 300 µM in shigemi 3 mm, Tris buffer 20 mM pH 6.5, NaCl 100 mM, 0,11 mM DSS, 10% D_2O, 40°C. 500 MHz spectrometer equipped with a TCI cryoprobe.

- ^{1}H-^{15}N HSQC sequence parameters : TD F2 2048, F1 220, NS 64, D1 1 sec, AQ 0.128 sec 0.054, SW 16ppm 40ppm, O1P : 4.7 117, RG 362

- processing parameters : SI 4096 512, sine bell apodisation in F2 and F1, SSB 2 4, TDeff 0

IV.4. 3. Recording conditions for 1D spectra of H23

H23 protein produced in H_2O with ^{1}H, ^{15}N, ^{13}C -Gly and unlabeled other amino acids (^{1}H-^{14}N-^{12}C) , without or with three inhibitors 20 mM AOA, 20 mM DM and $NaBH_4$ treatment. Protein concentration 200 µM in shigemi tubes 5mm, Tris HCl buffer 20 mM, pH 9, NaCl 100 mM, 0,11 mM DSS, 10% D_2O, 40°C, 500 MHz.

- ^{1}H 1D sequence parameters : TD 2048, , NS 2048, D1 1 sec, AQ 0.170 sec, SW 12ppm, O1P : 4.7, RG 161

- processing parameters : SI 4096, sine bell apodisation, SSB 2, TDeff 2048

IV.4. 4. Recording conditions for ^{1}H-^{13}C HSQC spectra of H23

^{1}H-^{13}C HSQC spectrum of H23 protein Gly ^{1}H-^{15}N-^{13}C and other AA ^{1}H-^{14}N-^{12}C labeled produced in H_2O without inhibitors. Protein concentration 200 µM in a 5 mm shigemi tube, 25°C, Tris buffer 20 mM, pH : 8.5, NaCl 100 mM, 100% D_2O, 0,11 mM DSS, 500 MHz, TCI cryoprobe.

^{1}H-^{13}C HSQC ^{15}N-decoupled of H23 produced with ^{1}H, ^{15}N, ^{13}C-Gly and ^{2}H, ^{14}N, ^{12}C_nonGly in D_2O without or with 20 mM AOA inhibitor. Protein concentrations: 170 µM or 275 µM respectively, pH 7, 40°C, 10% D_2O, 0.11 mM DSS, 3 mm tube, TCI cryoprobe 500 MHz.

- ^{1}H-^{13}C HSQC sequence parameters: TD F2 1024, F1 2048, NS 8, D1 1 sec, D16 0.0001 sec, AQ 0.102 sec 0.1809 sec , SW 10ppm 100ppm, O1P : 4.7 45, RG 322.

- processing parameters: SI 2048 1024, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 0

IV.4. 5. Recording conditions for ^{1}H-^{13}C HSQC spectra of feeding mixes

All feeding mixes were analyzed with the following characteristics : 40°C, 5mm tube, 100% D_2O, 0,11mM DSS, 500 MHz Bruker TCI cryoprobe spectrometer,

- ^{1}H-^{13}C HSQC sequence parameters : TD F2 1024, F1 512, NS 4, D1 1 sec, AQ 0.102 sec 0.02, SW 10ppm 100ppm, O1P : 4.7 45, RG 1290

- processing parameters : SI 2048 1024, sine bell apodisation in F2 and F1, SSB 2 4, TDeff 0

IV.4. 6. Recording conditions for MASS spectrometry

Mass spectrometry was performed in a 4700 MALDI TOF/TOF proteomics analyzer (Applied Biosystems). The whole proteins were studied in linear mode and analyzed in reflector mode (MALDI TOF), Laser Yag 355 nm, 5 nsec impulses of 200 Hz. The matrix was a-cyano-4-hydroxycinnamic (HCCA). Samples of a concentration of about 100 µM. The samples were

purified using custom-made tips containing Poros R2 (C8) resin. These small columns were washed with acetonitrile and equilibrated with TFA/water 0.1%. Samples (2 µL) diluted in 10 µL TFA/water 0.1% were applied to each column and washed in with TFA/water 0.1% and bound peptides and protein were eluted with 0.5 µL of HCCA matrix. The purified material was spotted onto steel plates and analyzed in positive ion mode. Calibration was done using classical protein standards (Protein mix 1 or 2 from LaserBio Lab).

IV.4. 7. CH_2-TROSY and CH_2-HSQC recording conditions

On the Bruker Z gradient equipped Cryoprobe 950 MHz spectrometer were carried out the CH_2-TROSY and CH_2-HSQC experiments:

- For samples of H23 protein, either uniformly labeled ^{15}N-^{13}C or glycine ^{15}N-^{13}C specifically labeled in 10% D2O 90% H_2O or, after lyophilization 100% D_2O, at 10°C, 25°C and 40°C phosphate buffer 30 mM pH 7 100 mM NaCl 1 mM DTT, 1% DSS, Shigemi tube 3 mm

- CH_2 optimized HSQC sequence parameters: TD F2 2048, F1 64, NS 32, D1 1 sec, AQ 0.083 sec 0.019, SW 13ppm 7ppm, O1P : 4.711 43.5, RG 128

- CH_2 optimized HSQC processing parameters: SI 4096 128, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 0

- CH_2 optimized TROSY sequence parameters: TD F2 4096 F1 64, NS 32, D1 1 sec, D16 0.0002 sec, AQ 0.166 sec 0.019, SW 13 7, O1P : 4.711 43.5, RG 128

- CH_2 optimized TROSY processing parameters: SI 4096 128, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 2048

- For samples of PPIase protein, uniformly labeled, 100% D2O, 30 mM Tris pH 7, 50 mM NaCl, 1 mM DTT, 1% DSS at 25°C, Shigemi tube 4mm

- CH_2 optimized TROSY sequence parameters: TD F2 4096 F1 256, NS 2, D1 1.5 sec, AQ 0.166 sec 0.0766 sec, SW 13 7, O1P : 4.711 43.5, RG 2050

- CH_2 optimized TROSY processing parameters: SI 4096 1024, sine bell apodisation in F2 and Qsine bell apodisation in F1, SSB 2 and 2, TDeff 2048

IV.4. 8. Calculation of the d_A proton apparent distance

As relaxation of the spin system is directly influenced by the neighboring spins, we determine d_A an apparent distance that considers the remoteness and gyromagnetic ratio of all

the neighboring proton isotopes, depending on the type of labeling, less that 5 angstroms away from the spin system of interest, and evaluate it as the virtual distance of a single proton. The d_A was calculated as follows, with r the distance of the isotope from the spin system of interest, and γ the gyromagnetic ratio of that isotope:

$$d_A = \frac{1}{(\sum(\frac{1}{r^6} \times \%^1H \times {\gamma_{1_H}}^2 + \frac{1}{r^6} \times \%^2H \times {\gamma_{2_H}}^2)/ {\gamma_{1_H}}^2)^{\frac{1}{6}}}$$

The d_A was calculated for each proton of the spin system : the protons of the methylene of glycines of the H23 protein available in Annex 1 using VMD, pdb file 5blk . Depending on the type of labeling (uniformly or specific, in H_2O or D_2O based solvent) the d_A found for each methylene proton of glycine was reported into a certain range. The highest d_A values correspond to the most deuterated forms of the protein : the more the environment is deuterated, the further is the apparent distance of a fictive proton and thus the less influence it has on the relaxation of the spin system. There is little difference of the d_A value between the specifically labeled sample and uniformly labeled sample in H_2O solvent (maximum between 2 and 2.5 Å). This suggests that the proton of the amide has significant influence on relaxation. When the sample is uniformly labeled in D_2O, the d_A value increases slightly (maximum between 2.5 and 3 Å) but the protons of side chains and the Cα still have an important impact on spin system relaxation. It is when using the specifically labeled sample in D_2O where the d_A value increases significantly (maximum between 3.5 and 4 Å).

IV.4. 9. PPIase titration recording conditions

On the Bruker Z gradient equipped Cryoprobe 950 MHz spectrometer were carried out the CH_2-TROSY and CH_2-HSQC experiments: PPIase Pin1 domain uniformly ^{13}C-^{15}N labeled, 1mM, 30 mM Tris pH 7, 50 mM NaCl, 2 mM DTT, 100% D_2O, 25°C, 950 MHz cryoprobe with increasing concentrations of CZ181; 0 mM , 0.2 mM, 0.4 mM, 0.8 mM, 1.2 mM, 2 mM and 4 mM

- CH_2 optimized HSQC sequence parameters : TD F2 2048, F1 256, NS 4, D1 1.5 sec, D16 0.0002 sec, AQ 0.083 sec 0.077, SW 13ppm 7ppm, O1P : 4.711 43.5, RG 128

- CH_2 optimized HSQC processing parameters : SI 4096 1024, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 1024

- CH_2 optimized TROSY sequence parameters : TD F2 4096 F1 256, NS 2, D1 1.5 sec, D16 0.0002 sec, AQ 0.166 sec 0.077, SW 13 7, O1P : 4.711 43.5, RG 161

- CH_2 optimized TROSY processing parameters : SI 4096 1024, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 0

- ^{1}H-^{15}N TROSY sequence parameters : TD F2 4096, F1 128, NS 4, D1 1 sec, D16 0.0002 sec, AQ 0.154 sec 0.019, SW 14ppm 35ppm, O1P : 4.7 117, RG 912

- processing parameters : SI 8192 512, sine bell apodisation in F2 and F1, SSB 2 2, TDeff 0

IV.4. 10. S3-HACACO recording conditions

On the Bruker Z gradient equipped Cryoprobe 950 MHz spectrometer were carried out the s3-HACACO experiments: U- ^{1}H ^{15}N-^{13}C labeled sample of PPIase domain 500 µM in 30 mM Phosphate, pH 7.4, 50 mM NaCl, 1 mM DTT, 0.5% DSS, 0,05% NaN_3, 1 mM EDTA, 100% D_2O, 40°C, Shigemi 5 mm tube.

- s3-HACACO sequence parameters : TD F3 4096 F2 64 F1 160, NS 4, D1 1.5 sec, D16 0.0002 sec, AQ 0.166 0.026 0.034 sec, SW 13 5 10, O1P : 4.711 43.5 169 ppm, RG 128

- s3-HACACO processing parameters : SI 8192 128 340, sine bell apodisation in F3 F2 and F1, offset 0.45 end 0.98 power 2 c 0.5 ; 0.5 0.98 1 0.5 ; 0.5 0.98 1 0.5, TDeff 0

IV.4. 11. Docking simulations

The CZ181 ligand pdb file was generated using MolView in two N-configurations : with the proline N in “up” or “down” (tetrahèdre 121° or -120°). Then, the conformations were minimized so as to get the ligand conformation lowest in energy, using GROMACS 2018.6 (Ref 1) by steepest descent method. Once the minimalization was done for the two conformations, we used CHARMM GUI webserver (Ref 2) to generate a system (System composition: 1 peptide molecule, 1430 water molecules, 1 Na+ ion (neutralize system). MD simulation conditions: 300K, 1 bar, 150 ns duration). Using GROMACS again, we carried out molecular dynamic simulations with a previous phase of equilibration using CHARMM36 (Ref 3) and simulated for quantifying the two populations for Cγ exo and endo angles. For the “up”

this gave 121° 36° and 121° -42° but for the “down” this only gave one in exo as for the endo conformation steric hindrance made it much less plausible (maybe<5%).

The Protein Data Bank was searched for crystallographic structures of the PPIase domain or Pin1 with no mutations and with/without ligand interaction. The best structure was deemed to be 1PIN as it had the best resolution (1.35 Å) in complex with dipeptide ALA-PRO and, compared to another structure (4U86, resolution 1.6 Å) free/apo published by different authors, the catalytic site conformation remained unchanged, visible when structures were superimposed by VMD (rmsd site actif proche).

Using Autodock tools, 1PIN had the water molecules removed and the hydrogen atoms added. A grid box for the PPIase domain was determined which encompasses the entire domain. The ligands were also processed by Autodock tools so as to determine the rotatable and nonrotatable bonds. Amide bonds were unrotatable and an added nonrotatable bond was made between the N of the proline and the reduced C of the serine. Docking was undertaken only for the “up” conformations exo and endo with exhaustiveness 8 using VINA.

1. Gromacs version 2018.6: Abraham, M.J., T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E.Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1–2:19–25.
2. http://www.charmm-gui.org
3. CHARMM36: Klauda, J.B., R.M. Venable, J.A. Freites, J.W. O’Connor, D.J. Tobias, C.Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, and R.W. Pastor. 2010. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B. 114:7830–7843.

1. Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Current Opinion in Structural Biology* **9,** 594–601 (1999).
2. Riek, R., Wider, G., Pervushin, K. & Wuthrich, K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences* **96,** 4918–4923 (1999).
3. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in ^{13} C-Labeled Proteins. *J. Am. Chem. Soc.* **120,** 6394–6400 (1998).
4. Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. & Simorre, J.-P. Improved Sensitivity and Resolution in ^{1} H– ^{13} C NMR Experiments of RNA. *J. Am. Chem. Soc.* **120,** 11845–11851 (1998).
5. Fiala, R., Czernek, J. & Sklenář, V. Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *Journal of Biomolecular NMR* **16,** 291–302 (2000).
6. Miclet, E., Williams, D. C., Clore, G. M., Bryce, D. L., Boisbouvier, J. & Bax, A. Relaxation-Optimized NMR Spectroscopy of Methylene Groups in Proteins and Nucleic Acids. *J. Am. Chem. Soc.* **126,** 10560–10570 (2004).
7. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced ^{1} H– ^{13} C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society* **125,** 10420–10428 (2003).
8. Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L. E. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *J. Am. Chem. Soc.* **126,** 3964–3973 (2004).
9. Rashid, S., Lee, B. L., Wajda, B. & Spyracopoulos, L. Side-Chain Dynamics of the Trifluoroacetone Cysteine Derivative Characterized by ^{19} F NMR Relaxation and Molecular Dynamics Simulations. *J. Phys. Chem. B* **123,** 3665–3671 (2019).
10. Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wüthrich, K. Uniform and Residue-specific ^{15} N-labeling of Proteins on a Highly Deuterated Background. *J Biomol NMR* **29,** 289–297 (2004).
11. Clark, L., Zahm, J. A., Ali, R., Kukula, M., Bian, L., Patrie, S. M., Gardner, K. H., Rosen, M. K. & Rosenbaum, D. M. Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris. *J Biomol NMR* **62,** 239–245 (2015).
12. Proudfoot, A., Frank, A. O., Frommlet, A. & Lingel, A. in *Methods in Enzymology* **614,** 1–36 (Elsevier, 2019).
13. Miyanoiri, Y., Takeda, M., Terauchi, T. & Kainosho, M. Recent developments in isotope-aided NMR methods for supramolecular protein complexes –SAIL aromatic TROSY. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1864,** 129439 (2020).
14. Imiołek, M., Karunanithy, G., Ng, W.-L., Baldwin, A. J., Gouverneur, V. & Davis, B. G. Selective Radical Trifluoromethylation of Native Residues in Proteins. *J. Am. Chem. Soc.* **140,** 1568–1571 (2018).
15. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR †. *Biochemistry* **36,** 1389–1401 (1997).
16. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. 6
17. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression. *J. Am. Chem. Soc.* **136,** 11308–11310 (2014).
18. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. & Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angewandte Chemie International Edition* **49,** 1958–1962 (2010).
19. Religa, T. L., Ruschak, A. M., Rosenzweig, R. & Kay, L. E. Site-Directed Methyl Group Labeling as an NMR Probe of Structure and Dynamics in Supramolecular Protein Systems: Applications to the Proteasome and to the ClpP Protease. *J. Am. Chem. Soc.* **133,** 9063–9068 (2011).
20. Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P. & Boisbouvier, J. An optimized isotopic labelling strategy of isoleucine-γ_2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **48,** 1434–1436 (2012).
21. Ruschak, A. M., Velyvis, A. & Kay, L. E. A simple strategy for 13C,1H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *J Biomol NMR* **48,** 129–135 (2010).
22. Rosenzweig, R. & Kay, L. E. Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR. *Annu. Rev. Biochem.* **83,** 291–315 (2014).
23. Amero, C., Schanda, P., Durá, M. A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B. & Boisbouvier, J. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *Journal of the American Chemical Society* **131,** 3448–3449 (2009).

24. Klapper, M. H. The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochemical and Biophysical Research Communications* **78,** 1018–1024 (1977).
25. Smith, M. H. The Amino Acid Composition of Proteins. *J. Theoret. Biol.* **13,** 261–282 (1966).
26. Albrecht, Gustav. & Corey, R. B. The Crystal Structure of Glycine. *J. Am. Chem. Soc.* **61,** 1087–1103 (1939).
27. Vuister, G. W., Delaglio, F. & Bax, A. An empirical correlation between 1JC.alpha.H.alpha. and protein backbone conformation. *J. Am. Chem. Soc.* **114,** 9674–9675 (1992).
28. Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S.-Q., Haynes, S. R. & Steven, A. C. Glycine loops in proteins: their occurence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *International Journal of Biological Macromolecules* **13,** 130–139 (1991).
29. Creighton. *Proteins: Structures and molecular properties.* (W.H Freeman and Company, 1992).
30. Javadpour, M. M., Eilers, M., Groesbeek, M. & Smith, S. O. Helix Packing in Polytopic Membrane Proteins: Role of Glycine in Transmembrane Helix Association. *Biophysical Journal* **77,** 1609–1618 (1999).
31. Dong, H., Sharma, M., Zhou, H.-X. & Cross, T. A. Glycines: Role in α-Helical Membrane Protein Structures and a Potential Indicator of Native Conformation. *Biochemistry* **51,** 4779–4789 (2012).
32. Khazanov, N. A. & Carlson, H. A. Exploring the Composition of Protein-Ligand Binding Sites on a Large Scale. *PLoS Comput Biol* **9,** e1003321 (2013).
33. Hoogstraten, C. G. & Johnson, J. E. Metabolic labeling: Taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concepts Magn. Reson.* **32A,** 34–55 (2008).
34. Levine, E. M. & Simmonds, S. Metabolite uptake by serineglycine auxotrophs of Escherichia coli. *J. Biol. Chem.* **235,** 2902–2909 (1960).
35. Monneau, Y. R., Ishida, Y., Rossi, P., Saio, T., Tzeng, S.-R., Inouye, M. & Kalodimos, C. G. Exploiting E. coli auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *Journal of Biomolecular NMR* **65,** 99–108 (2016).
36. Cheng, Z., Guo, C., Chen, Z., Yang, T., Zhang, J., Wang, J., Zhu, J., Li, D., Zhang, T., Li, H., Peng, B. & Peng, X. Glycine, serine and threonine metabolism confounds efficacy of complement-mediated killing. *Nat Commun* **10,** 3325 (2019).
37. Fernandes, H. S., Ramos, M. J. & Cerqueira, N. M. F. S. A. Catalytic Mechanism of the Serine Hydroxymethyltransferase: A Computational ONIOM QM/MM Study. *ACS Catal.* **8,** 10096–10110 (2018).
38. Prabhu, V., Chatson, B., Abrams, G. & King, J. 13C Chemical shifts of 20 free amino acids and their use in detection by NMR of free amino acids in intact plants of Arabidopsis. *Journal of Plant Physiology* **149,** 246–250 (1996).
39. Kay, L., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114,** 10663–10665 (1992).
40. McCoy, M. A. & Wyss, D. F. Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations. *Journal of Biomolecular NMR* **18,** 189–198 (2000).
41. Aguirre, C. Analyse quantitative des perturbations de déplacements chimique pour la détérmination de structures tridimensionnelles de complexes protéine-ligand. (2014).
42. Namanja, A. T., Wang, X. J., Xu, B., Mercedes-Camacho, A. Y., Wilson, K. A., Etzkorn, F. A. & Peng, J. W. Stereospecific gating of functional motions in Pin1. *Proceedings of the National Academy of Sciences* **108,** 12289–12294 (2011).
43. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6,** 277–293 (1995).
44. Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition Is Phosphorylation Dependent. *Cell* **89,** 875–886 (1997).
45. Guichard, G., Violette, A., Chassaing, G. & Miclet, E. Solution structure determination of oligoureas using methylene spin state selective NMR at ^{13}C natural abundance. *Magn. Reson. Chem.* **46,** 918–924 (2008).
46. Ippel, J. H., Wijmenga, S. S., de Jong, R., Heus, H. A., Hilbers, C. W., de Vroom, E., van der Marel, G. A. & van Boom, J. H. Heteronuclear Scalar Couplings in the Bases and Sugar Rings of Nucleic Acids: Their Determination and Application in Assignment and Conformational Analysis. *Magnetic Resonance in Chemistry* **34,** S156–S176 (1996).
47. Singh, H., Singh, S. & Raghava, G. P. S. In silico platform for predicting and initiating β-turns in a protein at desired locations: Prediction of Beta Turn in Proteins. Proteins **83,** 910–921 (2015).
48. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

Chapter V: Towards protein D-DNP NMR

This part of my work was undertaken at the NMR lab of the Ecole Normale Supérieure de Paris, which also belongs to the Laboratoire des Biomolécules. I worked in close collaboration with Daniel Abergel and David Guarin (PhD student) who have a long experience in D-DNP. Developing D-DNP approaches requires the optimization of DNP juice, the suitable parametrization of the polarizer, and a deep know-how for launching the dissolution. Their help was invaluable to exploring novel techniques for protein D-DNP. My work here consisted of preparing and characterizing the samples for D-DNP experiments, and analyzing the first NMR spectra that we obtained at the ENS. I am also thankful to Gilles Clodic for the MASS spectrometry analysis, and Lotfi Habchi (M2 student) who started this project.

V.1. Introduction

One of the biggest drawbacks of solution NMR is the low sensitivity of nuclear signals which requires high concentrations for chemical and biological studies. The NMR sensitivity is closely linked to the spin polarization, i.e. the population unbalance between Zeeman levels. For nuclear spins, this polarization remains very weak, even at high magnetic fields. To compensate this low sensitivity, signal averaging is required, despite considerably increasing experimental duration. Signal averaging is also problematic when the biological sample evolves or is unstable. Dynamic Nuclear Polarization is an emerging technique in solution NMR that greatly enhances sensitivity by hyperpolarization of nuclear spins through electron polarization transfer, capable of increasing signal by several orders of magnitude and thus diminish experiment time significantly. The concept was originally theorized in the 1950's by Overhauser[1], and demonstrated on metal lithium by Carver and Slitcher[2] but, as cryogenic temperatures are needed to obtain maximum polarization, DNP remained a technique reserved for solid state NMR.

It regained particular interest in 2003 for solution NMR, when Jan H. Ardenkjaer-Larsen et al.,[3] demonstrated that a sample could conserve part of its hyperpolarization in solution. They achieved this feat by polarization of electrons from a radical at cryogenic temperatures (T = 1.2 K) in the presence of a constant magnetic field. Then, using a series of microwave irradiations enabling electron spin flip and, through dipolar interactions and spin diffusion, the polarization

was transferred to nuclei of their sample. They then injected the sample from the polarizer to the spectrometer by dissolution through a magnetic tunnel. This technique enabled increases of signal to noise ratio by several orders of magnitude for the ^{13}C and ^{15}N atoms of urea compared to a non-hyperpolarized sample. After this methodological advancement, many developments were undertaken for biological NMR. Notably, Bowen and Hilty[4] demonstrated the potential of using hyperpolarized compounds to monitor enzyme kinetics in real time. By hyperpolarizing a substrate (benzoyl-L-arginine ethyl ester BAEE) and rapidly injecting it into a spectrometer ready with the enzyme trypsin, they observed ^{13}C signals of substrate disappearance and formation of the product without needing isotopic enrichment, obtaining a signal to noise ratio of 62:1 in 1 scan. At the Laboratoire des Biomolecules, the pentose phosphate pathway is studied using real time D-DNP, using hyperpolarized glucose derivatives[5,6].

In 2014, Harris et al.,[7] applied hyperpolarization techniques on larger biomolecular systems through hyperpolarized water. The hyperpolarized water was dissolved in a biomolecular sample containing either an amino acid (alanine and arginine) or polypetides (peptides of about 3 kDa from aldehyde reductase). The exchangeable protons of amide and amine groups of these samples underwent rapid exchange with the dissolved hyperpolarized water, leading to signal enhancements of these ^{1}H resonances. Moreover, through heteronuclear Overhauser effect, the ^{15}N nuclei of these chemical groups also underwent spontaneous polarization, leading to signal to noise ratios hundreds of times more intense than for the non-hyperpolarized sample and lasting a significant amount of time. They obtained sensitivity enhancements of more than 180 and lasting up to 40 seconds after water injection for the NH of alanine. For arginine, these sensitivity enhancements went up to 280 and 360 for the $^{15}NH_2$ (guanidinium) and $^{15}NH_3$ (amine) sites and lasted up to 20 seconds. For the polypeptides, signal enhancement of backbone amide groups lasted more or less 20 seconds, much longer than the overall T_1 relaxation time of these backbone amides (1.8 seconds), reflecting the long T_1 of water suggesting repolarization of ^{15}N during the consecutive scans and leading to signal enhancements of more than 500 compared to a non-hyperpolarized sample. The signal enhancements achieved for exchangeable protons and the bonded nitrogen could significantly improve ^{1}H-^{15}N two-dimensional NMR experiments by use of hyperpolarized water.

In 2015, Chappuis et al.,[8] used hyperpolarized water coupled to Water-LOGSY experiments to study protein-ligand interactions. To detect ligand binding, the Water-LOGSY, experiment rests on the transfer of polarization via the nuclear Overhauser effect (NOE) from

water to ligand. This transfer is either directly from the solvent to the ligand or indirectly via the bound protein. The ligand, having a small τ_C , its signals are negative when in its free form. However, the Dot1L protein, having a large τ_C, displays a positive signal in such an experiment. In the presence of protein, the bound ligand takes the τ_C of the protein and thus also has positive signals. In 2017, Kurzbach et al.,[9] used hyperpolarized HDO to enhance two-dimensional signals of amide sites of osteopontin, an intrinsically disordered protein. IDPs are difficult to observe under physiological conditions and temperature, often giving rise to overlapping peaks – a characteristic of their very dynamic nature and lack of secondary structure. After injection of hyperpolarized HDO into the sample waiting in the spectrometer, the labile protons exchanged with the protons of the HDO and a ^{15}N-^{1}H SOFAST spectrum was recorded. Several correlations could be assigned with signal enhancements of about 40 but that went up to 58 for some very intense peaks compared to the thermal equilibrium (non-hyperpolarized).

Beginning of 2020, Olsen et al.,[10] applied a similar method for ^{13}C-detected two-dimensional recordings and for three-dimensional experiments of proteins. They used two models, ubiquitin, a globular protein and osteopontin. In their H^N-CON experiments, after dissolution of hyperpolarized HDO in their sample and proton exchange of protein backbone amides, the polarization was transferred to the ^{15}N nucleus by a selective INEPT pulse sequence and after an evolution period transferred to the ^{13}CO nucleus for detection. This way, polarized amide protons are constantly replenished for a certain amount of time, enabling successive scans. Moreover, detecting on the ^{13}C offers a bigger spectral width than for ^{1}H detection, thus less overlap. To further reduce spectral crowding, they undertook a 3D ^{1}H-^{13}C-^{15}N experiment (BEST-HNCO) where 96 scans were recorded in 65 (osteopontin) and 75 seconds (ubiquitin) demonstrating the utility of polarization for enhanced sensitivity for structural studies of proteins.

Using hyperpolarized water has the advantage of avoiding dissolution of the protein itself, often reaching boiling temperatures, which can irreversibly denature the protein. Moreover, residual protons in deuterated water acts as a reservoir for hyperpolarization, having a T_1 much greater than proteins. However, this method relies on exchangeable proton transfer, therefore much precious structural and dynamics information can be lost, especially in the hydrophobic cores of the proteins.

In 2011, Ragavan et al.,[11] hyperpolarized directly a 14.9 kDa protein in the polarizer, with consecutive dissolution into the spectrometer. Fractionally deuterated samples were directly polarized and ^{13}C signals were detected after the dissolution in the NMR spectrometer. Compared to thermal polarization, averaged signal enhancements of 700 or 500 were obtained for carbonyl groups and aliphatic carbons, respectively, but it was shown that with the protein used, the protein did not recover its initial structure after the dissolution. Enhancement were found very dependent on the T_1 relaxation time, and up to 2000 for the methyl Cδ of the methyl and isoleucine residues characterized by T_1 of ~5 seconds at 400 MHz.

Very recent works have demonstrated that D-DNP could be used to probe the real-time folding kinetics of an intrinsically disordered protein when interacting with a binding partner. For these purposes, the ^{13}C signals of the protein were observed after direct hyperpolarization in a deuterated background[12]. In addition, hyperpolarization was used to follow the refolding of RNAs based on proton exchange of the imino groups of nucleic acids with hyperpolarized water[13].

Ragavan et al.,[11] have also achieved direct ^{1}H polarization on peptide samples. Enhancements of 30 and 45 were obtained for amide and aliphatic proton resonances of a bacitracin sample, respectively, much smaller than for the carbon spectra obtained on the protein sample. However, for aromatic protons, these enhancements reached up to 180. In their paper, Ragavan et al., theorize that sensitivity gains differ because of the difference in T_1 of these protons which were all relatively low because the bacitracin sample they used was not partially deuterated. Widespread proton presence creates dipolar interactions which will severely affect T_1 relaxation times and thus the potency of the DNP polarization. The conclusion of this paper is extremely interesting in that they suggest to use selective isotopic labeling strategies to further reduce spectral overlap, but they don't go into more detail as how it could help increase T_1 relaxation times and the gains that could be obtained with such samples. As far as the biomolecular interactions are concerned, the Hilty group has developed ligand screening methods[14] whereby a ligand is hyperpolarized and injected into a spectrometer containing the protein sample for interaction studies. Indeed, they observed selective polarization transfer between the ^{1}H protons of folic acid and of DHFR protein by 2D SOFAST-HMQC which provides invaluable information about the binding mode.

Therefore, although D-DNP studies of peptides are emerging, these studies, taken together, revealed two significant barriers to signal recording of hyperpolarized proteins: the harsh dissolution step which affects the protein structure and the short T_1 relaxation times of the protein protons. Therefore, there has been no work yet published to our knowledge displaying proton NMR spectra of globular proteins in their native state, directly and thus homogeneously hyperpolarized by DNP at low temperatures. However, hyperpolarizing protein samples could be very valuable for several applications including ligand screening tests.

The WW domain of Pin1, for which a wide library of ligands has been described is characterized by its high thermodynamic stability. It possesses remarkably efficient refolding properties after thermic denaturation. Dissolution-DNP requires advanced instruments, for which few are available commercially, thus stays limited to very few laboratories in the world. GE SpinLab ™ have an available device, although its applications are focused on imagery. We are lucky to have instruments for NMR polarization at the Laboratoire des Biomolécules. In this last project, we thus used the Pin1 WW domain to explore the feasibility of using D-DNP on globular proteins. As a first step, DNP juice was adapted to a protein sample and the efficiency of the polarization was monitored. Secondly, we produced WW samples with different ^2H contents to evaluate the influence of the deuteration on the proton T_1 relaxation time.

The WW subunit of the Pin1 protein is 39 amino acids long and forms a stable antiparallel β-sheet composed of three strands (see Chapter II). With a molecular mass of 7.14 kDa for our construct counting the residues of the his-tag and linker, and a τ_C of 3,5 ns at 25°C, $\omega_0 . \tau_C \approx 1$ at 400 MHz which places it in the dip of the T_1 equation curve (Figure 6.1).

As shown previously, relaxation of the spins is influenced by their molecular environment. Isotopic labeling with deuterium of a protein sample can considerably extend T_1 and T_2 relaxation times, notably by removal of dipolar interactions. Developing an efficient labeling strategy, which would ensure sparse, random proton presence in a deuterated environment so as to minimize dipolar interactions and thus conserve long proton T_1, could offer the sensitivity gains needed in D-DNP experiments. Thus, a first objective of this last part of my thesis is to use the WW domain, expressed with different proton densities sparsely and randomly placed in a deuterated environment and to measure T_1 relaxation times. Indeed, for hyperpolarization to subsist the dissolution and transfer process from the polarizer to the

spectrometer, the T_1 of individual protons needs to be superior to ca. 3-4 seconds which isn't the case for small to medium sized proteins, when fully protonated.

V.2. Importance of T_1 relaxation in D-DNP experiments

At cryogenic temperatures, the T_1 of protein nuclei can be extremely long (several hours at 1.2K) as the molecular dynamics are extremely limited (τ_C very large). This means that, in these conditions, longitudinal magnetization takes a long time to return to thermal equilibrium after a perturbation. Once at room temperature though, the T_1 will be much shorter (a few seconds at best for carbons and less than one second for protons in small to medium sized proteins) as the τ_C diminishes. This is an issue for D-DNP as hyperpolarization longevity is conditioned by T_1 length (see equation 5.1). The precious hyperpolarized ^1H signal is therefore lost in a few seconds at the liquid state for proteins.

$$\frac{1}{T1} = \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma^4 h^2}{r^6}\left[\frac{2\tau_c}{1+\omega_0^2\tau_c^2} + \frac{8\tau_c}{1+4\omega_0^2\tau_c^2}\right] \quad (5.1)$$

with ω_0 the transition frequency of the spin, *r* the distance between the spins, γ the gyromagnetic ratio of the nuclei, *h* the Plank constant and μ_0 the magnetic moment of the spin. This relation corresponds to the T_1 of a nucleus belonging to an isolated pair standing in a magnetic field B_0 ($\omega_0 = \gamma B_0$). The expression of $1/T_1$ as a function of τ_C/τ_C^2 means that, depending on the value of τ_c, a term may take the upper hand . When the $\tau_C \rightarrow \infty$, then $1/T_1$, which contains the terms $\frac{\tau_C}{(1+\alpha(\omega_0\times\tau_C)^2}$, tends to $\frac{1}{(1+\alpha(\omega_0\times\tau_C)}$ which tends to zero and therefore T_1 will be very long. This will be true except for some values of τ_C where, because of the constant in front of the dominant term, the $1/T_1$ will increase with a maximum for $\omega_0\tau_c \approx 1$ and the T_1 will be at its minimum (Figure 5.1).

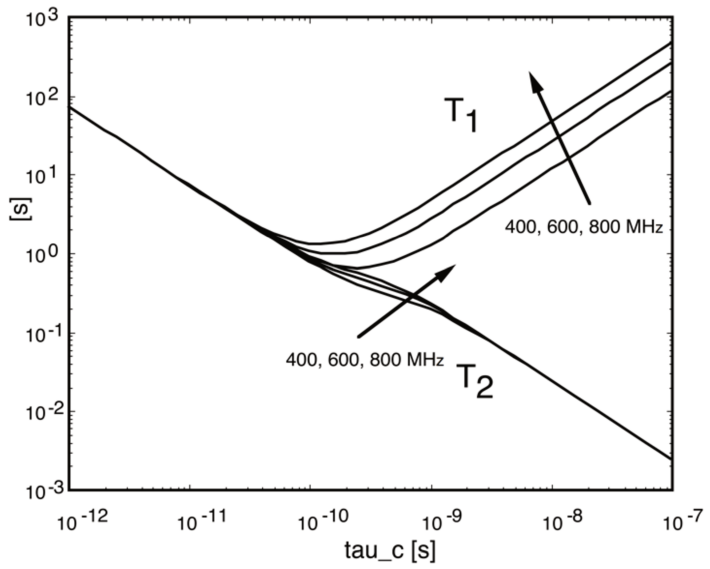


Figure 5.1 : T_1 and T_2 as a function of the correlation time for a two spin system consisting of two protons with identical Larmor frequencies at 400, 600 and 800 MHz, standing at a distance of 2Å.

The hyperpolarized signal can be observed by two means in 1D experiments, either by using one hard pulse to detect the whole magnetization in one scan or by repeating several small flip angle pulses to observe phenomena occurring in real times during several seconds. When a 90° pulse is applied to the system the longitudinal magnetization vanishes and the hyperpolarization is lost but very intense signals are expected. The signal is very intense with sensitivity proportional to hyperpolarized magnetization. When smaller angle pulses are applied, this brings only a portion of the spins to population balance, and, although less intense, the depolarization involves only some of the spins, thus partially conserving hyperpolarization for a series of small angle pulses (Figure 5.2).

For small Θ angle excitations, the hyperpolarization retained after n scans corresponds to

$M_0 \times (\cos \theta)^n$, where $\cos \theta = 98\%$ for $\theta = 10°$.

Compared to the maximal sensitivity S obtained with a flip angle of 90°, the sensitivity is

$S \times (\sin \theta)$ for each scan with a θ flip angle, which is equal to 17% of S for $\theta = 10°$.

This acquisition mode is then very valuable for monitoring evolution in the NMR tube during a few tens of seconds.

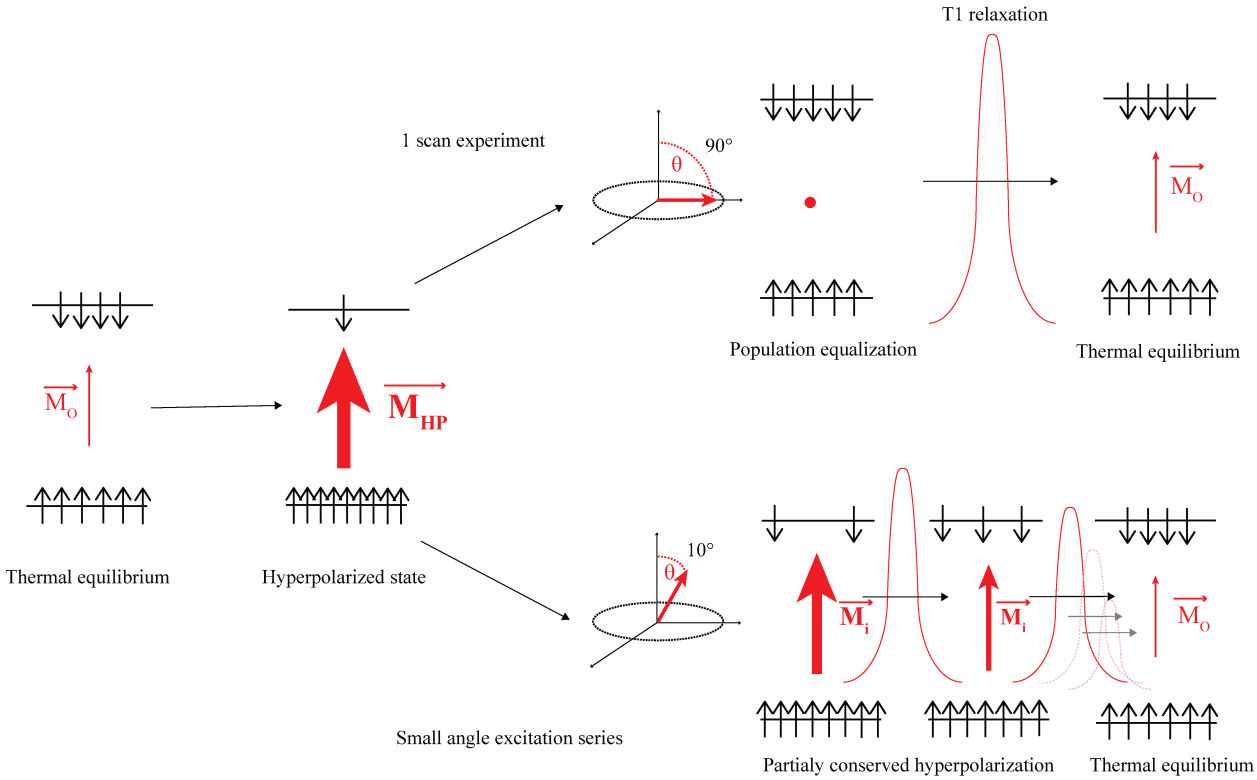


Figure 5.2 : Illustration of the difference between a single scan experiment and small angle excitation series. The single scan experiment produces an intense signal taking advantage of the whole hyperpolarization achieved but it is depleted afterwards. The small angle excitation series produces, for each scan, less intense signals but a large fraction of hyperpolarization is conserved for the following acquisition.

V.2.1 WW labeling strategy

The WW subunit is the N-terminal domain of Pin1 consisting of 39 residues and of a mass of 4.51 kDa (see Chapter II). The WW domain of Pin1 is comprised of amino acids 24 to 64 in Figure 6.3, because for purification purposes, a His-tag and cleavage site have been added to the construction, amounting to 64 amino acids. This His-tag plays in our favor as it also increases the τ_C and thus the T_1 of the protein. For these purposes, it was left uncleaved for the experiments. The theoretical molecular weight of this construction at pH 7 is 7143.74 kDa.

```
        10         20         30         40         50         60
GSSHHHHHH SSGLEVLFQG PSSMADEEKL PPGWEKRMSR SSGRVYYFNH ITNASQWERP SGNS
                               10         20         30         40
```

Figure 5.3 : Amino acid sequence of the WW domain construction used for the D-DNP experiments. Highlighted in grey is the His-tag and in cyan the linker, which stays attached to the WW domain after cleavage. The numbers in bold under the sequence represent the real numbering of the WW domain.

For these DNP experiments, random proton occurrence in a perdeuterated background was desired, so as to isolate individual protons as much as possible from dipolar interactions. Controlling proton density was achieved by producing the WW domain in deuterated solvent but with slightly differing conditions. For such purposes, a high cell-density protocol[15] was employed, with changes regarding added labeled or unlabeled glucose, and changes in M9+ medium preparation. The conditions are available in section V.4.1. This protocol generated 3 different levels of deuterated samples, compared to a fourth, unlabeled control sample produced using a regular *E.coli* based protocol (see Chapter II).

To verify that the WW domain could withstand the high temperatures of dissolution, it was expressed with ^{1}H-^{15}N-^{13}C labeling and two HN-HSQC spectra were recorded, before and after heating at 90°C for 5 minutes. The recording conditions are given in section V.4.2. and the NMR spectra obtained are displayed in Figure 5.4. The fold of the WW protein is lost for temperatures close to 60°C[16] but is instantly recovered when the temperature is set back to 25°C. Pin WW is recognized as a fast-folding domain with folding rates in the microsecond range[17]. We checked by NMR that the HN-HSQC spectra were qualitatively and quantitively superimposable, which attests that the thermic denaturation is absolutely reversible.

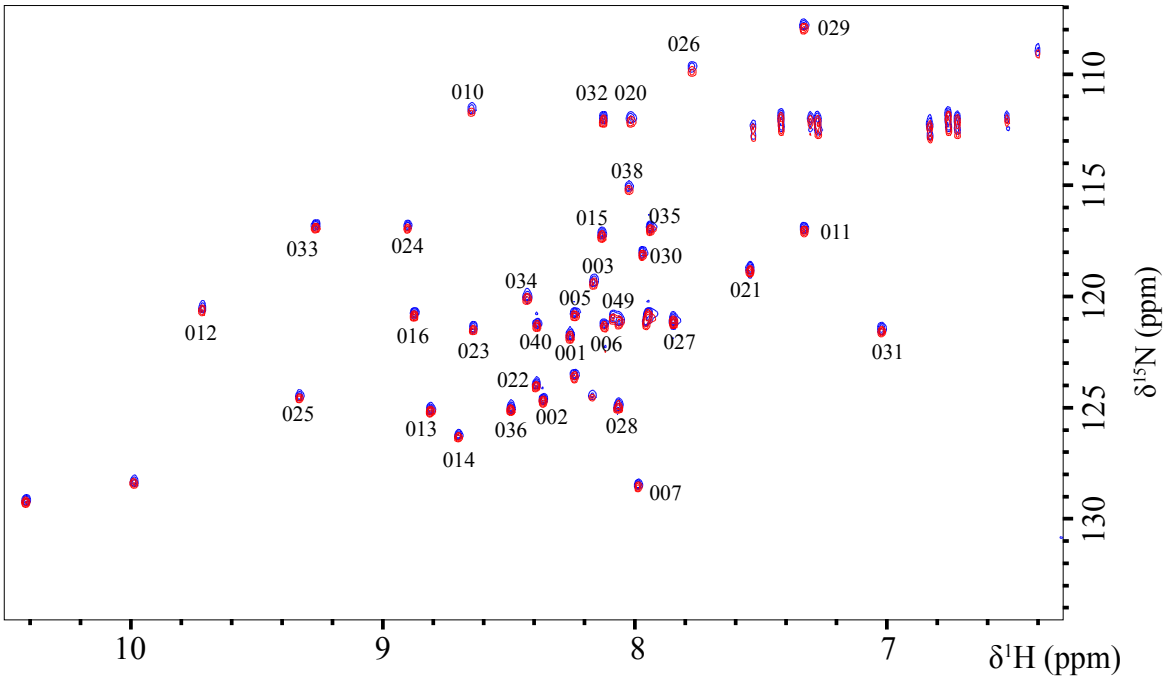


Figure 5.4: ^{15}N-^{1}H HSQC spectra of the WW domain U ($^{2}H_{80\%}$ -^{15}N-^{12}C) labeled. 330 µM, 30 mM Phosphate pH 8, 50 mM NaCl, 50% D_2O, 50% H_2O, 1% DSS. 25°C 500 MHz Cryoprobe Bruker spectrometer, 5mm tube. In blue before and in red after 5 minutes' incubation at 90°C.

To verify the ^{2}H labeling percentages of the different samples, mass spectrometry (MALDI-TOF) was employed see section V.4.5. WW samples were diluted in milliQ H_2O at 10 µM and mixed with matrix (HCCA) for acquisition (see V.4.5. for complete protocol). The Figure 5.5 displays the observed masses for the four samples. There is a total of 6 negatively charged residues (Asp + Glu) and 6 positively charged residues. With an atomic composition of 309 carbons, 98 nitrogen and 456 hydrogens at pH 7, the percentage of deuteration was calculated assuming 98% ^{15}N labeling:

= (Mass WW_labeled - MassWW_Non-labeled – 98-98% nb N) / (nb H non-exchangeable)

x100 = (7544.4 - 7144.7 - (0.98*98)) / (335)

The calculations suggest there is 80.1%, 90.6% and 92.6% ^{2}H labeling for the three deuterated samples. This is somewhat surprising because we had hoped for values closer to 99% at least for production 3 where the autoclave step was skipped to avoid the proton exchanges. However, different sources of protons remained in our expression media such as the ammonium chloride $^{15}NH_4Cl$ added without a lyophilization step in D_2O, in addition to the ^{1}H impurities in the commercial D_2O used. The samples were analyzed by mass spectrometry before and directly after HPLC to ensure that the masses were unchanged (see Annex 6).

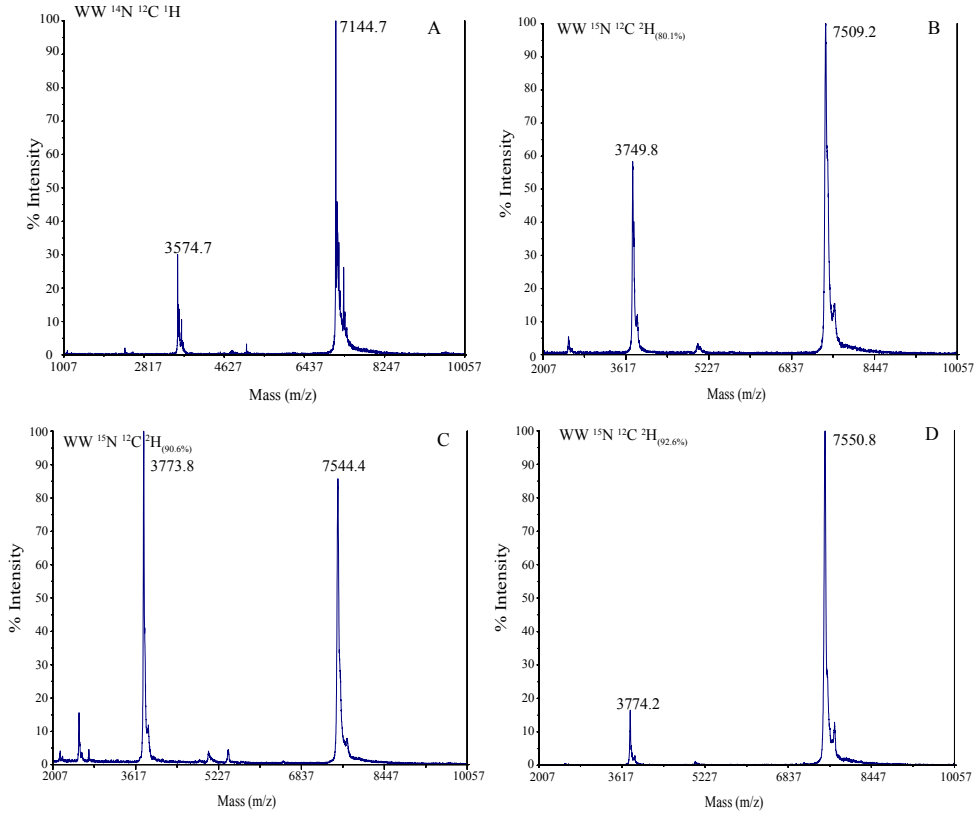


Figure 5.5 : Mass spectra of different samples of WW protein (uncleaved of the His-tag), expressed with different levels of labeling. A) unlabeled. B) ^{2}H-^{15}N labeled expressed with ^{1}H glucose in autoclaved D_2O M9+ medium. C) ^{2}H-^{15}N labeled expressed with ^{2}H glucose in autoclaved D_2O M9+ medium. D) ^{2}H-^{15}N labeled expressed with ^{2}H glucose, lyophilized M9+ salts and filtered D_2O.

V.2.2. Choice of the WW ligand.

Among the numerous ligands described for the Pin1 protein, we chose the Pintide peptide which was originally designed from a peptide library[18,19]. It is a six-residue peptide containing a phosphorylated serine: WFYpSPR. The affinity determined for the full length Pin1[19] was between 200 to 400 µM (100 mM imidazole pH 6.6, 100 mM NaCl, 5 mM DTT) which is more than one order of magnitude larger than previous studies performed with an extension of the Pintide sequence on the *C*-terminal part,[20] where affinity was determined at 17 µM for the full length Pin1 and 44 µM for the WW domain alone with Pintide WFYpSPFLE (pH 7.5 25 mM HEPES,100 mM NaCl, 1 mM DTT). This observed difference could also be due to changed buffer conditions which affects the charge of the phosphate and disturbs binding. A titration study was undertaken at our laboratory and revealed Pintide binding affinity of 9,4 $\pm$ 4 µM for full length Pin1 (30 mM Tris buffer pH 7, 50 mM NaCl, 2 mM DTT). The chemical shift perturbation of amide groups (recorded for ^{1}H-^{15}N spectra) are given below (Figure 5.6). The threshold bar, calculated as the sum of the average of all CSP and the standard deviation, allows the identification of the most representative CSP, localized between residues R14 and E35 for the WW domain, and residues R54, S71 and between residues G128 and S147 for the PPIase domain. As can be observed, the Pintide induces chemicals shift changes in several regions of the WW domain, but also in the WW-PPIase interface by allosteric effect, which suggests a global change of dynamics and/or structure of WW upon Pintide binding.

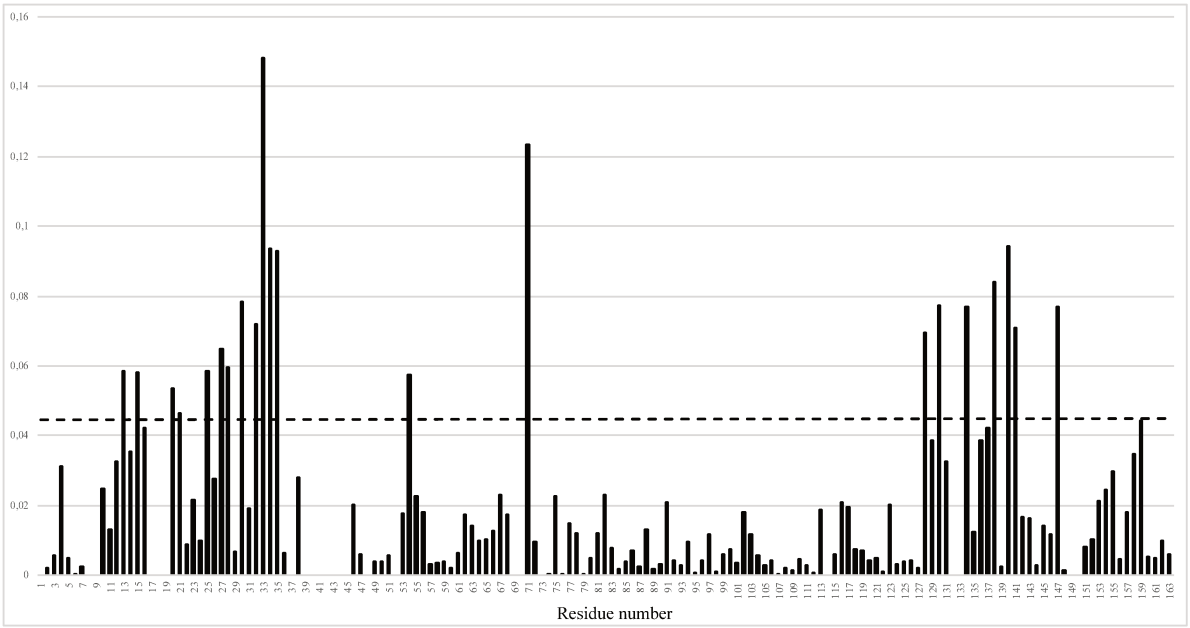


Figure 5.6: CSP of the WW domain (from the full length Pin1 titration) with the Pintide. 500 µM, 30 mM Phosphate pH 7.5, 50 mM NaCl, 10% D_2O, 90% H_2O, 1% DSS. 25°C 500 MHz Cryoprobe Bruker spectrometer, 5mm tube

The Pintide is commercially available and will be used, in future works, to demonstrate that polarization transfer can occur between the hyperpolarized WW domain and a peptide ligand. We used a model peptide similar to the Pintide, originating from the CDC25 natural Pin1 partner to screen the proximal protons less than 4Å of distance which could have an influence on relaxation and would be susceptible to absorb hyperpolarization of the protons of the protein. The PDB entry 1I8G[21] corresponds to a NMR ensemble of 10 structures, that have been analyzed as possible interaction models for the WW domain with the pCD25 ligand. We used this available pCDC25-WW complex to characterize the protein-ligand intermolecular vicinities between protons. pCDC25 is a phosphopeptide of 10 amino acids derived from Cdc25 which is a protein partner of Pin1. Intermolecular ^{1}H-^{1}H distances less than 4Å have been extracted from the structures and revealed that notably the HZ2 proton of W34 and the Hα proton of R17 of the WW domain are in the immediate vicinity of some ligand protons and could serve for transferring polarization (tables of ^{1}H-^{1}H distances are given in Annex 7). As the W34 presents the most important CSP after binding with the Pintide, the predicted contacts with the pCD25 peptide are sensitively similar. In addition, to allow a transfer a magnetization from the protein to the ligand, these WW protons in close contact with the peptide ligand should also be affected in terms of relaxation upon ligand binding. Using D-DNP on hyperpolarized protons of a protein sample could enable us to observe both the resonances of the interacting ligand and the amino acids in close contact with the ligand. Unfortunately, due to a long interruption of experimental activities in the laboratory from March to June 2020, we were not able to demonstrate this phenomenon. Nevertheless, we were able to prepare the protein sample for experimentation, as explained in the following sections.

V.2.3. T_1 measurements of the different WW labeled samples

Equilibrium is reached exponentially following: $M_z = M_0(1 - 2e^{-TI/T_1})$ (5.2) after a perturbation, where M_z depicts the magnetization as a function of TI, M_0 is the magnetization at thermal equilibrium, and T_1 is the longitudinal relaxation time to be fitted. To obtain the characteristic T_1 of given resonances, their intensities (proportional to M_z) are plotted as a function of TI using the inversion-recovery experiment. Longitudinal relaxation times of magnetization along the z-axis (T_1) can be measured using an inversion-recovery experiment. D1 - p1(180°) - TI – p2(90°) – FID. In this experiment, A 180° pulse (P1) is applied which will invert the spins and thus invert magnetization. Magnetization will be allowed to relax back to

M_0 (magnetization at thermal equilibrium) for a certain time (inversion time TI) which is incremented between two successive experiments. After the variable delay time TI, a 90° pulse (P2) is applied to the system to obtain transversal magnetization and the FID is recorded. The signal intensities contained within the FID are a function of the TI delay. The longer TI is, the more signal will have relaxed and this longitudinal relaxation rate is characterized by the T_1 value. In this experiment, a long delay D1 has to be set to attain thermal equilibrium between each scan.

1D ^{1}H spectra were recorded before and after each T_1 measurement experiment to ensure the sample was stable overtime. Four spectra are displayed (Figure 5.7), one for each labeling production each taken before the T_1 measurements. As one can observe, for equivalent protein concentrations and using the same NMR parameters, the signal intensity diminishes considerably when deuteration rates were increased. As the number of protons decreases, so does their signal, as S/N is proportional to concentration of protons.

The Table 5.1 summarizes the T_1 values found for several resonances or ensembles of resonances. On the figure 5.8, the labels of the resonances are displayed directly on the 1D spectra. Finally, typical inversion recovery data and the fit obtained are provided in Figure 5.9.

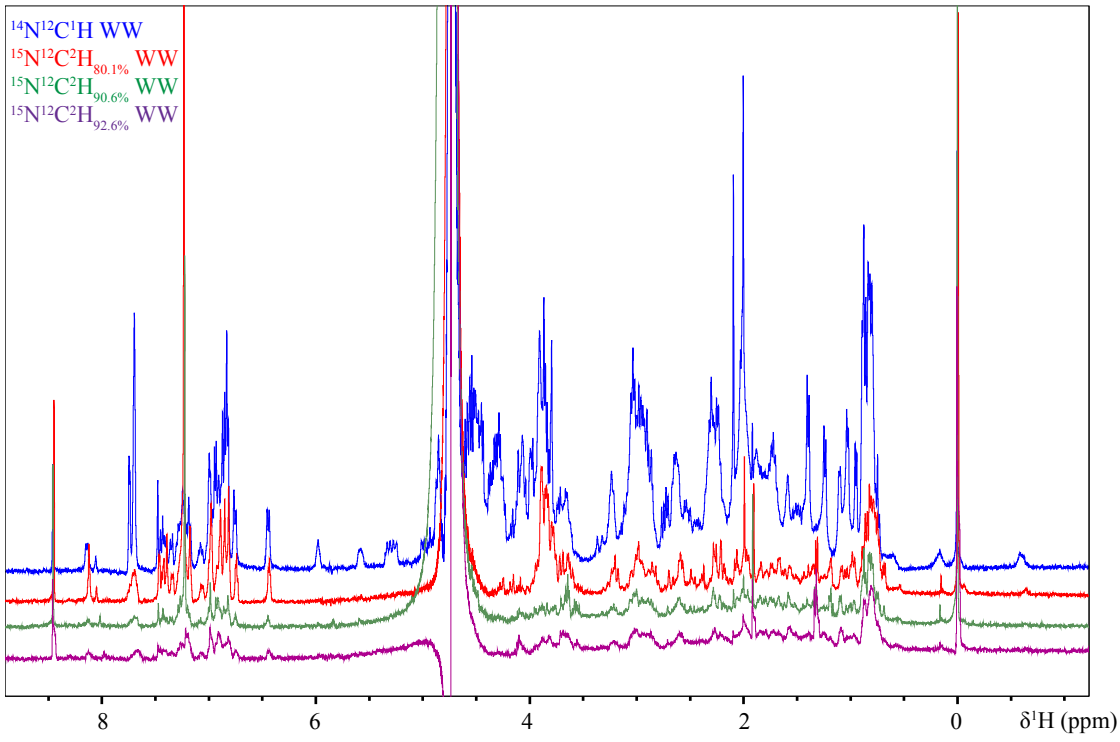


Figure 5.7 : 1D proton overlay of the four differently labeled WW productions. Blue: unlabeled WW, red : $^2H_{80.1\%}$-^{15}N, green: $^2H_{90.6\%}$-^{15}N, purple: $^2H_{92.6\%}$-^{15}N, 2 mg in 500 µL, 30 mM Phosphate pH 7.5, 50 mM NaCl, 100% D_2O, 1% DSS. 30°C 400 MHz.

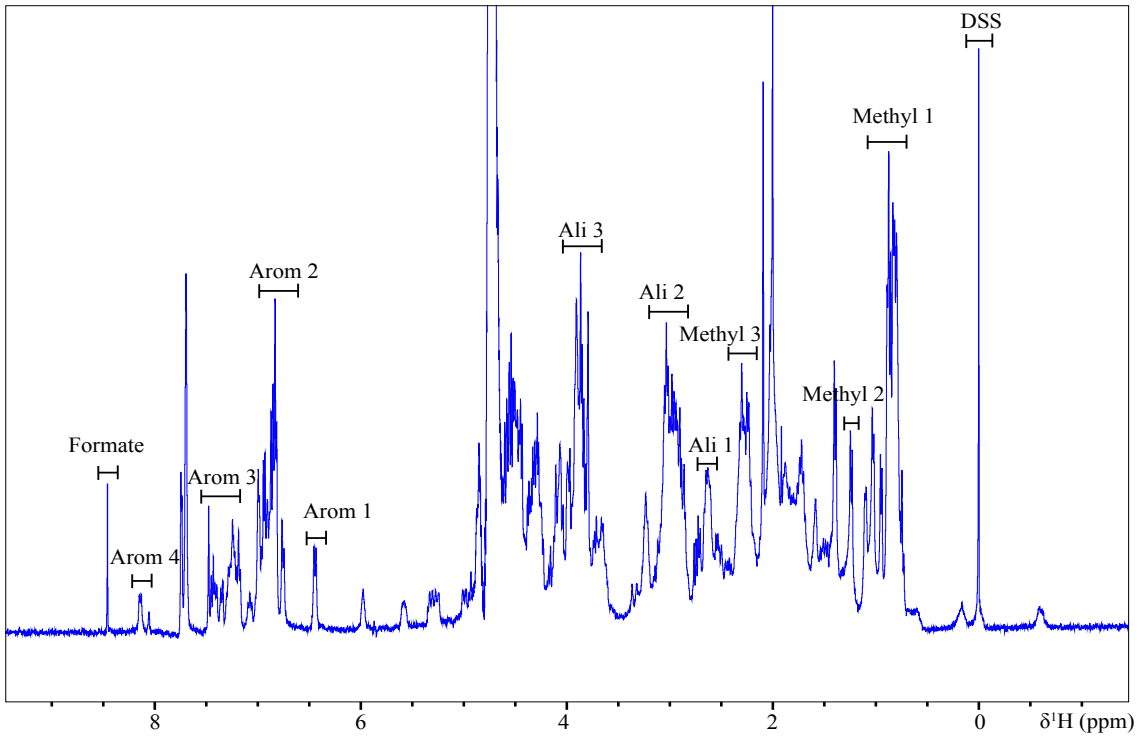


Figure 5.8 : 1D proton spectrum of unlabeled WW. The resonance labels are given above each. 2 mg in 500 µL, 30 mM Phosphate pH 7.5, 50 mM NaCl, 100% D_2O, 1% DSS. 30°C 400 MHz.

T1 sec	100% ^1H	ERROR	80.1% ^2H	ERROR	90.6% ^2H	ERROR	92.6% ^2H	ERROR
DSS	3,00	0,07	3,59	0,10	3,40	0,09	3,50	6,37
METHYL 1	0,54	0,01	0,93	0,03	1,12	2,22	1,34	2,69
METHYL 2	0,79	0,03	1,86	0,20	2,00	0,30		
METHYL 3	0,54	0,01	1,70	0,20	3,22	6,75	5,32	10,65
ALI 1	0,72	0,02	1,81	0,20	1,86	0,57	5,85	11,69
ALI 2	0,68	0,02	1,77	0,24	1,08	2,24	3,98	8,03
ALI 3	0,51	0,01	0,77	0,02	0,99	2,03	5,99	12,41
AROM 1	1,45	0,16	3,31	0,60	10,29	20,95		
AROM 2	1,48	0,05	3,92	0,14	8,26	0,43		
AROM 3	1,87	3,49	4,26	0,35	11,40	0,89		
AROM 4	1,64	0,34	3,18	0,52	10,25	19,31		

Table 5.1: Table of T_1 measured for different proton resonances using the four samples of WW domain.

For the unlabeled WW sample, where only the amide groups are deuterated in 100% D_2O, the relaxation times range between 0.5 s for the fastest relaxing protons (ALI 3) and 1.87 s for the slowest (AROM 3). The Hα (ALI 3) and other aliphatic protons (ALI 2 and ALI 1) have the shortest relaxation times. This is expected because they are densely surrounded by other protons within protein side chains, and thus are highly subject to dipolar interactions. It has to be noted that these T_1 relaxation times are already rather long because the amides are

deuterated, but, in a water-based solvent, the relaxation times would be even smaller in consequence, especially for the Hα. It is estimated that relaxation times of aliphatic protons increase significantly from 80% deuteration upwards[22]. Methyl protons have T_1 in the same range than Hα and Hβ protons (up to 0.78 s, METHYL 2) because they are surrounded by numerous protons, especially for those located in hydrophobic cores and are also subject to intra-methyl relaxation[23]. The longest T_1 relaxation times found were for aromatic protons with a mean of 1.6 s but going up to 1.86 s (AROM 3). Again, this was expected because proton-proton distances in aromatic groups are on average longer than in aliphatic chains. These values are in agreement with previous studies[24,25] whereby aromatic protons were found to have the longest T_1. Relaxation measurements on different protons of Y23 and C30 of the Bovine Pancreatic Trypsin Inhibitor at 327 K on a 300 MHz spectrometer gave relaxation times of 0.263 seconds and 0.435 seconds for the H^{δ} and H^{ε} of tyrosine respectively, and 0.3, 0.13 and 0.125 seconds for the Hα, Hβ and Hβ' of cysteine in 100% D_2O.

Once deuteration levels reach 80.1%, T_1 values more than double in average (although with some disparity, almost tripling for METHYL 3). A factor of ~4 is observed between the unlabeled and 90.6% deuterated samples, although a factor of 6 is visible for AROM 3 and METHYL 3. In general the T_1 times are considerably extended from 90% deuteration upwards. It is well known that for protein samples partially deuterated, the incorporation of residual protons is not fully randomized. It depends on several biochemical processes and some chemical groups in certain amino acids will be more prone to incorporate protons than others. This may explain why the evolution of METHYL1 and METHYL3 is very different upon deuteration (from relaxation times of 0.5 seconds when unlabeled to 1.1 and 3.2 seconds with 90.6% deuteration respectively). The corresponding T_1 values may reflect the local proton networks which can evolve very differently as the deuteration level is increased. However, at 90.6% and 92.6% deuteration, the signal was so low for some resonances that the T_1 could not be accurately measured (very high error bars) or not determined at all. As far as the DSS is concerned, very consistent T_1 values are measured which provides a valuable internal reference for these experiments.

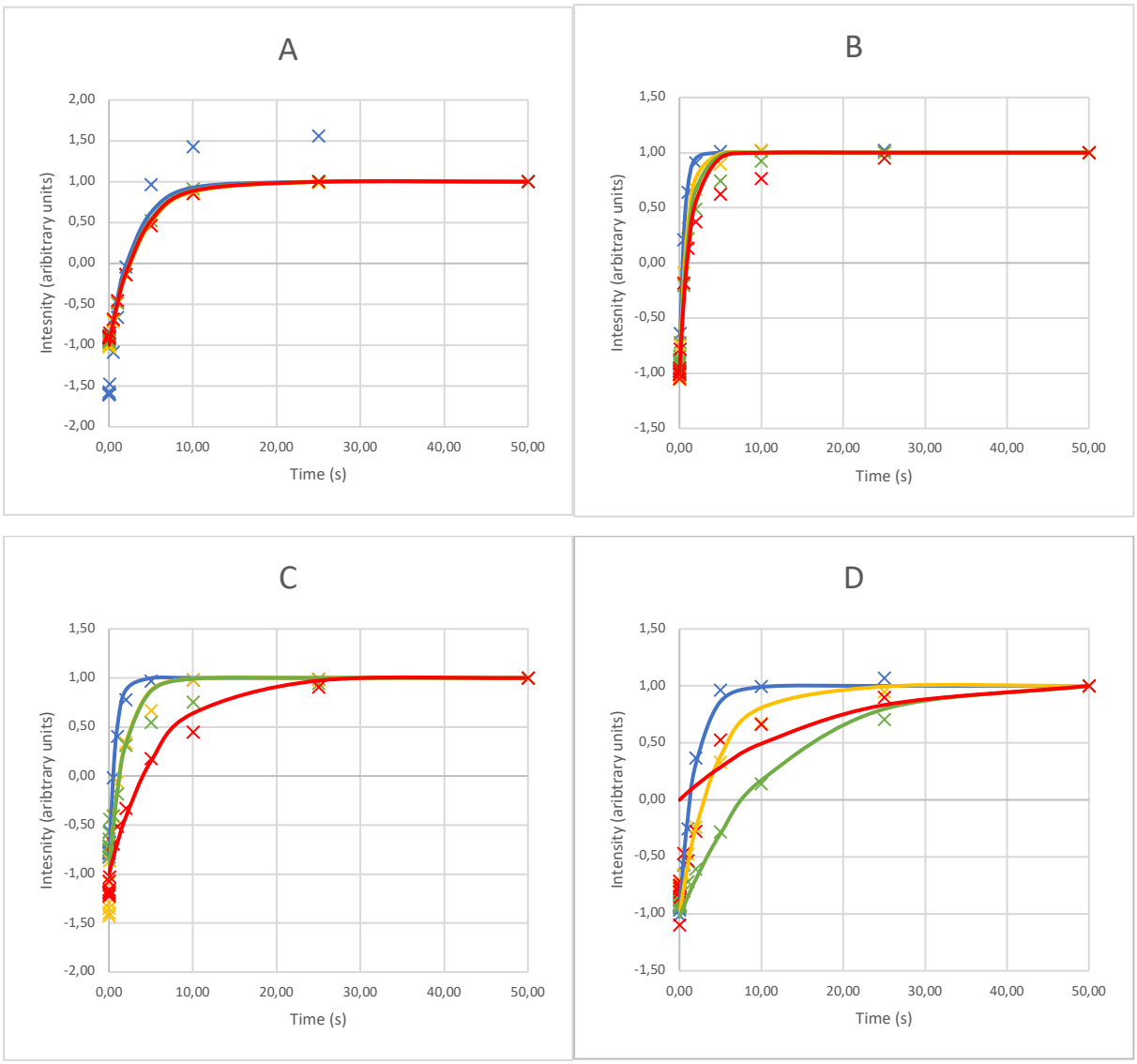


Figure 5.9 : Curve fits of 4 different resonances A : DSS, B : METHYL 1, C: ALI 1, D : AROM3 for differently labeled samples with increasing concentrations of ^{2}H. In blue unlabeled sample (100% ^{1}H), in yellow $^{2}H_{80.1\%}$, labeled sample, in green $^{2}H_{90.6\%}$, labeled sample, and in red $^{2}H_{92.6\%}$ labeled sample of WW protein.

Common D-DNP real time analyses are performed on ^{13}C nuclei with T1 of 4 to 10 s. On these systems, the time window for signal observation is typically 30 to 60 s. The proton T1 values on proteins deuterated at ~90% seem thus high enough to envision D-DNP experiments on proteins. T_1 would be even longer using higher field. It would be also possible to deuterate at a higher rate the protein, although the rarer will be the protons, the weaker will be the protein 1D spectra. 90% of deuteration appears to be a good compromise between sustaining proton concentration and sustaining hyperpolarization.

V.3. Protein D-DNP experimentations

V.3.1. D-DNP set up and "DNP-juice"

Dissolution DNP is set up such that the polarized sample can be rapidly transferred to the spectrometer for signal recording. In the polarizer, illustrated in Figure 5.10, the cryogenically frozen DNP sample is inserted, containing compounds necessary for efficient polarization. The sample is then cooled further to 1.2 K and irradiated with microwaves. After polarization, an injection of pressurized (10 bars) and heated (~130°C) D_2O (5 mL) is undertaken so as to push the sample up and into the magnetic tube. The pressure enables quick transfer to the spectrometer, in which a tube with a prepared buffered solution (500 µL) is waiting, at room temperature.

Figure 5.10 : Illustration of the D-DNP set up, with the polarizer on the left and the spectrometer on the right. In between them is a magnetic tunnel in which passes the dissolved sample to reach the spectrometer.

The "DNP-juice" is the solvent preparation that contains different compounds enabling homogenous and efficient polarization of the sample. Apart from the substrate of interest, it contains two other important compounds. Firstly, the radical, source of unpaired electrons, that permits polarization of the nuclei. There are different types of commercial radicals available. The most common nitroxide-based free radicals for biological experiments include AMUPol ((15-{[(7-oxyl-3,11-dioxa-7-azadispiro[5.1.5.3]hexadec-15-yl)carbamoyl][2-(2,5,8,11-tetraoxatridecan-13-ylamino)}-[3,11-dioxa-7-azadispiro[5.1.5.3]hexadec-7-yl])oxidanyl)[26], TEMPOL (4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl)[27], and TOTAPOL (1-(TEMPO-4-oxy)-3-(TEMPO-4-amino)propan-2-ol)[28]. The concentration of the free radical can range from 5 to 50 mM, depending on the type of nuclei that is polarized and its concentration. A low concentration of free radical would inefficiently polarize the nuclei, whereas an elevated concentration risks affecting the relaxation of the sample after dissolution, unpaired electrons

acting as a strong source of relaxation. For our experiments, 15 mM to 40 mM of TEMPOL was used for ^{1}H polarization.

Additionally, the sample requires a glassing solvent. The sample has to form an amorphous, non-crystalline solid at cryogenic temperatures, such that the radical is well distributed and in contact with the material to be polarized for a homogenous transfer. The choice of glassing solvent will therefore depend on the solubility of the sample. Common glassing agents used include glycerol, DMSO, ethanol, ethylene glycol and methanol in a solution of $^{1/2}H_2O$. The volume of the components can be adjusted depending on the solubility of the sample. For our experiments, DMSO was routinely used as it enabled high solubility of the protein, even at very high concentration.

Our sample preparation is undertaken by solubilizing the desired amount of protein (2.5 to 5 mg) in a small volume (100 µL) of DMSO and D_2O by vortexing. A solution of free radical (TEMPOL) in D_2O is added and the vortex is used again to ensure proper dissolution of the free radical and homogenous repartition of the compounds. A complete description is given in section V.4.6. The sample is irradiated with microwaves to excite the electron spins (equivalent to a radio-frequency pulse for nuclear spins) and the transfer of polarization from electrons is instantaneous, hyperpolarizing the nuclear spins. In the polarizer, 1° pulses are made every 5 seconds in order to follow the buildup of ^{1}H polarization (a 90° pulse would completely eliminate all acquired polarization, as explained previously section V.2).

V.3.2. Results

The polarization of the sample at the solid state, 1.2 K was undertaken on the differently labeled samples of the WW domain. The figure 5.11 displays, for the $^{2}H_{80.1\%}$-^{15}N-^{12}C labeled WW domain, the ^{1}H achieved after 190 minutes. The quantification of the polarization has not been performed since it required to wait very long time for the ^{1}H magnetization to come back to the thermal equilibrium after turning off the microwaves irradiation. Proton relaxation times are indeed very long at 1.2K. However, compared qualitatively to other samples usually polarized in the lab (mostly on ^{13}C nuclei), the mount of polarization is very satisfactory which indicate that the DNP juice is properly operating.

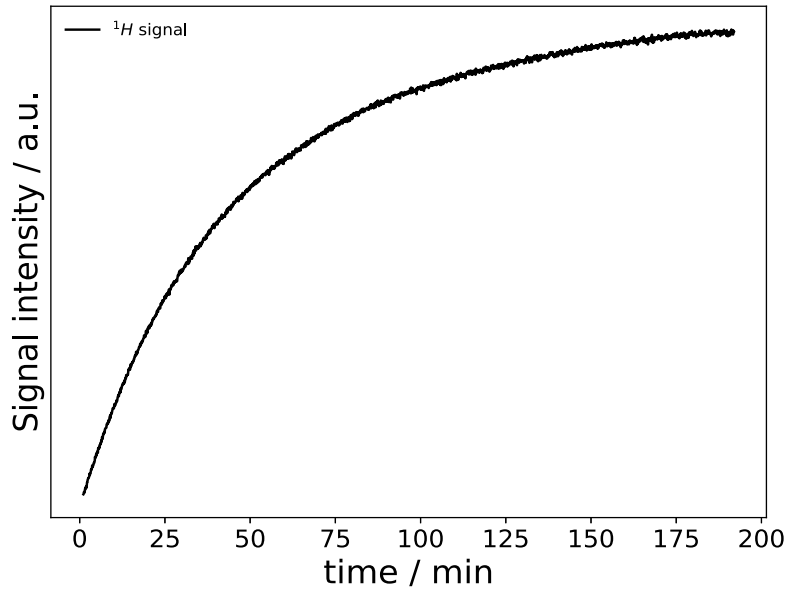


Figure 5.11: Solid polarization of the $^2H_{80.1\%}$-^{15}N- labeled WW domain in the polarizer achieved after 190 minutes. 6.7 T polarizer, the delay between each pulse is 5 s, with 1 degree squared pulses. Microwave irradiation is at 350 mW with a frequency of 187900 MHz, positive polarization. 5 mg WW, $^2H_{80.1}$%-^{15}N labeled, 15 mM TEMPOL, 60% : 40% D_2O:DMSO.

Proton polarization signal reaches a plateau at ~200 min, signaling that the maximum polarization has been reached in these conditions. Then, the dissolution is undertaken by injecting pressurized D_2O onto the sample, which will dissolve and rapidly pass through the magnetic tunnel to reach the spectrometer. However, this dissolution step has to be adapted to concentrated protein samples. A too low D_2O pressure will not insure a proper dissolution in the polarizer while to high pressure makes the protein sample froth. As a result, inhomogeneous liquid introduced in the NMR tube precludes the acquisition of well resolved spectra. ^1H NMR requires very homogeneous field in the sample volume compared to ^{13}C spectra that are less sensitive because of their much larger spectral width. Preliminary data were obtained on ^{13}C-labeled WW proteins but only a large resonance was observed in ^1H NMR, probably due to residual water protons and masking the protein resonances. The high concentration of protein in the DNP juice together with the very fast dissolution created a very dense mousse, lasting for tens of minutes. A new set-up is currently under investigation to limit this formation of mousse, using a chamber in the magnetic tunnel for achieving a phase separation.

Additionally, TEMPOL is more active at basic pH. The WW domain, after HPLC and lyophilization, remains very acidic because of the left-over TFA molecules. When the powder is dissolved in the low volumes of DNP-juice, the pH is extremely acidic (around 1). Correcting the pH to 10 using deuterated NaO^2H enabled enhanced polarization of the sample (Figure 5.12).

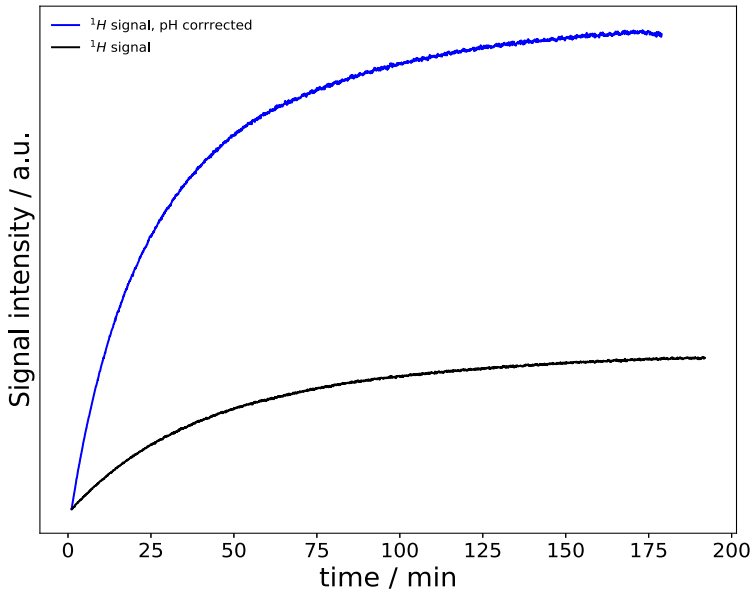


Figure 5.12: Solid polarization of the $^2H_{80.1\%}$-^{15}N labeled WW domain in the polarizer achieved after 190 minutes, in black when the DNP juice is not corrected for pH (estimated at around pH 2) in blue when the pH of the DNP-juice is corrected to pH 10. 6.7 T polarizer, the delay between each pulse is 5 s, with 1 degree squared pulses. Microwave irradiation is at 350 mW with a frequency of 187900 MHz, positive polarization. 5 mg WW, $^2H_{80.1}$%-^{15}N labeled, 15 mM TEMPOL, 60% : 40% D_2O:DMSO.

V.3.3. Conclusions and perspectives

It is possible to achieve long T_1 relaxation times for protons of proteins using a sparse and random deuteration strategy. Moreover, we have achieved the preparation of a sample for D-DNP, with a small domain capable of efficient solubilization at high concentration and withstanding the extreme conditions of the DNP experiment. The WW domain of the Pin1 that we used is unfolded above ~60°C but this thermal denaturation is fully reversible, the Pin1 WW domain being recognized as a very fast folding structure. Hyperpolarization of the ^1H of WW has proved feasible within ~3h. There are however some instrumental issues that need to be addressed for this project to continue. As the sample integrity is compromised by the high pressure needed to transfer it to the spectrometer, other set up are explored for the dissolution step, as the use of a chamber for achieving a phase separation before the injection of the hyperpolarized protein into the NMR tube. Transfer of solid samples could be also explored[29]. A bullet containing the frozen, hyperpolarized sample is ejected from the polarizer using pressurized helium gas, and shot into a secondary magnet, where the bullet can be dissolved instantaneously. This strategy however poses other problems linked to fast relaxation as the magnetic field is close to 0 in the tunnel.

Once this issue of dissolution has been resolved, it could be envisaged to monitor the hyperpolarization of the protein, in the absence or in the presence of the peptide ligand Pintide for which we have already obtained NMR data. Protein to small molecule hyperpolarization transfer could be a mean to detect new ligand and characterize protein-ligand interaction even at very low concentration.

An alternative perspective would be to use the carbonyl group as a local probe for protein structure and dynamics study. There have been some experiments based on ^{13}C polarization of proteins[11]. Using a similar labeling method to the one presented in the previous Chapter, it could be possible to selectively label the ^{13}CO of residues in proteins in a perdeuterated background using cell-free. Carbonyl groups have ideal relaxing properties[30], with a T_1 that would be capable of reaching almost one minute in proteins in a deuterated environment. Thus, these nuclei could be used in hyperpolarization experiments for monitoring biological processes on the minute time scale.

V.4. Materials and methods

V.4.1. ^{2}H-^{15}N labeled WW domain expression

Following an expression protocol by Sivashanmugam et al.,[15], 250 mL of LB medium pH 7.8 (10 g casein peptone/L, 5 g yeast extract/L, 10g NaCl/L) was prepared in D_2O, in 50 mL batches in 250 mL Erlenmeyers and autoclaved. They were inoculated with 1 mL of glycerol stock bacteria with 1mg/mL of ampicillin and 0,3mg/mL of chloramphenicol. These were left overnight, shaking at 220 rpm 37°C. The next morning, when DO_{600} had reached between 5 and 10, cells were centrifuged 15 minutes at 3000g 4°C and resuspended in M9+ D_2O medium. M9+ medium was prepared the day before, either autoclaved, or salts lyophilized, in D_2O and filtered, pH corrected to 8,2 with the following components: 50 mM Na_2HPO_4, 25 mM KH_2PO_4 , 10 mM NaCl, 5 mM MgSO4, 0.2 mM CaCl2, 1,7 g/ L Yeast Nitrogen Base, . Supernatant was discarded and cells resuspended in 250 mL of M9+ medium. Cells were left shaking at 220 rpm 25°C for at least 2 hours. At this point, 1% glucose, 0,1% ^{15}N ammonium chloride and 1 mM IPTG in D_2O were added. Cells were returned to the incubator for overnight induction shaking at 220 rpm 25°C.

The next day, cells were centrifuged 20 minutes at 4000g 4°C. Supernatant was discarded, and cells were resuspended in 40 mL of lysis buffer. Cells were sonicated 20 times,

10 for 15 seconds and 10 for 20 seconds, 40 amp. The lysate was then centrifuged 30 minutes at 25000 g 4°C and the supernatant was used for purification by His-trap. The column was equilibrated with phosphate buffer 30 mM pH 7,8, 50 mM NaCl, 30 mM Imidazole and eluted with 500 mM imidazole buffer with a 0-80% gradient in 40 minutes at 1 mL per minute. By monitoring absorbance at 280nm, the protein eluted was recuperated in fractions. Yields ranged from 20 to 70 mg/L of culture.

The elution fractions were then purified by HPLC C4-300 column, 10-100% acetonitrile 0,1% TFA in 30 minutes. Fractions containing the WW protein eluted at 8 minutes and were recuperated. They were shock frozen and lyophilized. Dry powder was weighed the next day and resuspended in D_2O, shock frozen with liquid nitrogen and lyophilized again before keeping at -80°C.

V.4.2. Recording conditions for ^{1}H-^{15}N HSQC spectra of the WW domain

The WW sample was recorded at 500 MHz on a Bruker TCI cryoprobe spectrometer, 330 µM, 30 mM Phosphate pH 8, 50 mM NaCl, 50% D_2O, 50% H_2O, 1% DSS. 25°C 5mm tube

- ^{1}H-^{15}N HSQC sequence parameters: TD F2 2048, F1 128, NS 4, D1 1 sec, AQ 0.128 sec 0.036, SW 16ppm 35ppm, O1P : 4.7 117, RG 90.5

- processing parameters: SI 2048 512, sine bell apodisation in F2 and F1, SSB 2.2 , TDeff 0

V.4.3. Recording conditions for ^{1}H 1D measurements of the WW domain

For all WW samples, measurements were undertaken at 400 MHz Bruker spectrometer, 30°C, 30 mM Phosphate, pH 7.5, 50 mM NaCl, 1% DSS, 2 mg protein in 500 µL, 100% D_2O, 5 mm tube with the following parameters:

- ^{1}H 1D sequence parameters: TD F1 8192, NS 128, D1 30 sec, AQ 0.683 sec, SW 15 ppm, O1P : 4.63 ppm, RG 71.8

- processing parameters: SI 16384, shifted sine bell apodisation, SSB 2, TDeff 8192

V.4.4. Recording conditions for ^{1}H T_1 measurements of the WW domain

To measure the T_1, a series of 14 experiments with an increasing delay TI was recorded, with TI : 0.01 ms, 0.1 ms, 0.5 ms, 2 ms, 5 ms, 20 ms, 100 ms, 500 ms, 1 s, 2 s, 5 s, 10 s, 25 s, 50 s. The fit of curves was calculated for different protons using the DOSY module of NMRPipe and the equation A×(C - exp(-1.0*|alpha|*x)) with C = 0.5. In this equation, alpha accounts for $1/T_1$ and x corresponds to the TI values.

For all WW samples, measurements were undertaken at 400 MHz Bruker spectrometer, 30°C, 30 mM Phosphate, pH 7.5, 50 mM NaCl, 1% DSS, 2 mg protein in 500 µL, 100% D_2O, 5 mm tube with the following parameters:

- T_1 measurement sequence parameters: TD F1 8192, NS 128, D1 30 sec, TI : 0.01 m, 0.1 m, 0.5 m, 2 m, 5 m, 20 m, 100 m, 500 m, 1 s, 2 s, 5 s, 10 s, 25 s, 50 s, AQ 0.683 sec, SW 15ppm, O1P : 4.63 ppm, RG 71.8

- processing parameters: SI 16384, sine bell apodisation, offset 0.45, end 0.98, power 2, c 0.5, TDeff 8192

V.4.5. MASS spectrometry of the WW samples

Mass spectrometry was performed in a 4700 MALDI TOF/TOF proteomics analyzer (Applied Biosystems). The whole proteins were studied in linear mode and analyzed in reflector mode (MALDI TOF), Laser Yag 355 nm, 5 nsec impulses of 200 Hz. The matrix used was a-cyano-4-hydroxycinnamic (HCCA). After lyophilization, 10 µL WW samples were diluted in milliQ H_2O to about 1 mM. The samples were purified using custom-made tips containing Poros R2 (C8) resin. These small columns were washed with acetonitrile and equilibrated with TFA/water 0.1%. Samples (1 µL) diluted in 10 µL TFA/water 0.1% were applied to each column and washed with TFA/water 0.1% and bound peptides and protein were eluted with 0.5 µL of HCCA matrix. The purified material was spotted onto steel plates and analyzed in positive ion mode. Calibration was done using classical protein standards (Protein mix 1 or 2 from LaserBio Lab).

V.4.6. Sample conditions for WW domain polarization

5 mg of lyophilized WW, ^{2}H-^{15}N labeled was diluted in 100 µL of 15 mM TEMPOL, 60%:40% D_2O:DMSO respectively. The sample is then either frozen in 10 µL drops in liquid nitrogen to ensure a perfectly amorphous sample before being placed in the hold, although freezing the sample directly in the hold in liquid nitrogen was deemed sufficient and polarization not affected. A polarizer of 6.7 T was used for solid ^{1}H positive polarization, with microwave irradiation at 350 mW at a frequency of 187900 MHz. Dissolution was undertaken with 5 mL of D2O for transfer to a 10 mm tube in a Bruker spectrometer of 9.4 T (400 MHz).

1. Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **92,** 411–415 (1953).
2. Carver, T. R. & Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **92,** 212–213 (1953).
3. Ardenkjær-Larsen, J. H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M. H., Servin, R., Thaning, M. & Golman, K. Increase in Signal-to-Noise Ratio of >10,000 Times in Liquid-State NMR. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 10158–10163 (2003).
4. Bowen, S. & Hilty, C. Time-Resolved Dynamic Nuclear Polarization Enhanced NMR Spectroscopy. *Angew. Chem. Int. Ed.* **47,** 5235–5237 (2008).
5. Sadet, A., Weber, E. M. M., Jhajharia, A., Kurzbach, D., Bodenhausen, G., Miclet, E. & Abergel, D. Rates of Chemical Reactions Embedded in a Metabolic Network by Dissolution Dynamic Nuclear Polarisation NMR. *Chem. Eur. J.* **24,** 5456–5461 (2018).
6. Miclet, E., Abergel, D., Bornet, A., Milani, J., Jannin, S. & Bodenhausen, G. Toward Quantitative Measurements of Enzyme Kinetics by Dissolution Dynamic Nuclear Polarization. *J. Phys. Chem. Lett.* **5,** 3290–3295 (2014).
7. Harris, T., Szekely, O. & Frydman, L. On the Potential of Hyperpolarized Water in Biomolecular NMR Studies. *J. Phys. Chem. B* **118,** 3281–3290 (2014).
8. Chappuis, Q., Milani, J., Vuichoud, B., Bornet, A., Gossert, A. D., Bodenhausen, G. & Jannin, S. Hyperpolarized Water to Study Protein–Ligand Interactions. *The Journal of Physical Chemistry Letters* **6,** 1674–1678 (2015).
9. Kurzbach, D., Canet, E., Flamm, A. G., Jhajharia, A., Weber, E. M. M., Konrat, R. & Bodenhausen, G. Investigation of Intrinsically Disordered Proteins through Exchange with Hyperpolarized Water. *Angewandte Chemie International Edition* **56,** 389–392 (2017).
10. Olsen, G. L., Szekely, O., Mateos, B., Kadeřávek, P., Ferrage, F., Konrat, R., Pierattelli, R., Felli, I. C., Bodenhausen, G., Kurzbach, D. & Frydman, L. Sensitivity-enhanced three-dimensional and carbon-detected two-dimensional NMR of proteins using hyperpolarized water. *J Biomol NMR* **74,** 161–171 (2020).
11. Ragavan, M., Chen, H.-Y., Sekar, G. & Hilty, C. Solution NMR of Polypeptides Hyperpolarized by Dynamic Nuclear Polarization. *Anal. Chem.* **83,** 6054–6059 (2011).
12. Ragavan, M., Iconaru, L. I., Park, C.-G., Kriwacki, R. W. & Hilty, C. Real-Time Analysis of Folding upon Binding of a Disordered Protein by Using Dissolution DNP NMR Spectroscopy. *Angew. Chem.* **129,** 7176–7179 (2017).
13. Novakovic, M., Olsen, G. L., Pintér, G., Hymon, D., Fürtig, B., Schwalbe, H. & Frydman, L. A 300-fold enhancement of imino nucleic acid resonances by hyperpolarized water provides a new window for probing RNA refolding by 1D and 2D NMR. *Proc Natl Acad Sci USA* **117,** 2449–2455 (2020).
14. Kim, Y. & Hilty, C. in *Methods in Enzymology* **615,** 501–526 (Elsevier, 2019).
15. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. & Li, Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **18,** 936–948 (2009).
16. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. Structure-function-folding relationship in a WW domain. *Proceedings of the National Academy of Sciences* **103,** 10648–10653 (2006).
17. Dave, K., Jäger, M., Nguyen, H., Kelly, J. W. & Gruebele, M. High-Resolution Mapping of the Folding Transition State of a WW Domain. *Journal of Molecular Biology* **428,** 1617–1636 (2016).
18. Yaffe, M. B. Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: A Potential Mitotic Regulatory Mechanism. *Science* **278,** 1957–1960 (1997).
19. Jacobs, D. M., Saxena, K., Vogtherr, M., Bernadó, P., Pons, M. & Fiebig, K. M. Peptide Binding Induces Large Scale Changes in Inter-domain Mobility in Human Pin1. *Journal of Biological Chemistry* **278,** 26174–26182 (2003).
20. Verdecia, M. A., Bowman, M. E., Ping, K., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *nature structural biology* **7,** 5 (2000).
21. Wintjens, R., Wieruszeski, J.-M., Drobecq, H., Rousselot-Pailley, P., Buée, L., Lippens, G. & Landrieu, I. ^{1}H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* **276,** 25150–25156 (2001).
22. Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J. & Laue, E. D. An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *J. Am. Chem. Soc.* **118,** 407–415 (1996).
23. Kalk, A. & Berendsen, H. J. C. Proton magnetic relaxation and spin diffusion in proteins. *Journal of Magnetic Resonance (1969)* **24,** 343–366 (1976).
24. Benz, F. W., Roberts, G. C. K., Feeney, J. & Ison, R. R. Proton spin-lattice relaxation studies of the histidine residues of pancreatic ribonuclease. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **278,** 233–238 (1972).

25. Boulat, B. & Bodenhausen, G. Measurement of proton relaxation rates in proteins. *J Biomol NMR* **3,** 335–348 (1993).
26. Sauvée, C., Rosay, M., Casano, G., Aussenac, F., Weber, R. T., Ouari, O. & Tordo, P. Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chem. Int. Ed.* **52,** 10858–10861 (2013).
27. Zakrzewski, J. & Krawczyk, M. Reactions of Nitroxides. Part XII [1]. – 2,2,6,6-Tetramethyl-1-oxyl- 4-piperidyl Chloroformate – A New Reactive Nitroxyl Radical. A One-pot Synthesis of 2,2,6,6-Tetramethyl-1-oxyl-4-piperidyl N,N-Dialkyl-carbamates. *Zeitschrift für Naturforschung B* **66,** 493–498 (2011).
28. Song, C., Hu, K.-N., Joo, C.-G., Swager, T. M. & Griffin, R. G. TOTAPOL: A Biradical Polarizing Agent for Dynamic Nuclear Polarization Experiments in Aqueous Media. *J. Am. Chem. Soc.* **128,** 11385–11390 (2006).
29. Kouřil, K., Kouřilová, H., Bartram, S., Levitt, M. H. & Meier, B. Scalable dissolution-dynamic nuclear polarization with rapid transfer of a polarized solid. *Nat Commun* **10,** 1733 (2019).
30. Stevanato, G., Hill-Cousins, J. T., Håkansson, P., Roy, S. S., Brown, L. J., Brown, R. C. D., Pileio, G. & Levitt, M. H. A Nuclear Singlet Lifetime of More than One Hour in Room-Temperature Solution. *Angew. Chem. Int. Ed.* **54,** 3740–3743 (2015).

Conclusions and perspectives

Beyond the development of apparatus designed to increase the quality of NMR signals, such as spectrometers capable of attaining higher magnetic fields, there are advantages to optimizing the sample itself for protein solution NMR. Optimization of the isotopic labeling of proteins using deuteration and selective labeling strategies, in parallel to the use of various TROSY experiments, has demonstrated the quantifiable gains in resolution and sensitivity obtained compared to customary experiments, and has enabled considerable progresses in structure and dynamics studies, especially for larger biomolecular complexes. In addition, the development of heteronuclear three-dimensional experiments also contributes to the NMR progress, enabling the experimenter to acquire finer, more accurate data. Moreover, novel techniques such as hyperpolarization for solution NMR greatly benefit from the prolonged relaxation times offered by isotopic labeling strategies.

Three different labeling strategies have been explored here for different NMR uses, including the TROSY experiments, a new 3D pulse scheme and contributions for the development of D-DNP. In particular, cell-free expression was used for perdeuteration of proteins while preserving the protonation of the amide groups for NH-TROSY experiments. Using labeled algal mixes, a cocktail of inhibitors of enzymes implicated in metabolism of amino acids, and working in a H_2O based solvent, the optimized cell-free protocol we describe has made it possible to overcome problems of proton back-exchange, thus allowing the study of all the amides of large molecular complexes, even in hydrophobic cores, whilst freeing oneself of the issues related to protein denaturation and renaturation protocols. This labeling strategy finds uses not only for 2D NH-TROSY experiments but for all heteronuclear 3D experiments that start on the amide of proteins, for titration experiments of large complexes or interaction studies.

In a similar fashion, the optimized cell-free protocol was also employed for the specific labeling of glycines in a perdeuterated environment, for use in CH_2-TROSY experiments. The labeling strategy coupled to the use of CH_2-TROSY experiments allowed gains in resolution and sensitivity of 2.5 compared to a CH_2-HSQC experiment, offering much neater spectra for analysis. But in fact, deuteration of amides only, achieved by working in a deuterated solvent, gave significant gains in itself when using the CH_2-TROSY. When applying this to protein

interaction studies, it gave finer resonances of the CH_2 of glycines, with traceable and quantifiable chemical shifts which could serve for ^{1}H- ^{13}C titration spectra, useful for Kd determination and for the mapping of the biomolecular interactions. Considering that Cα positions are more reliable tracers than HN groups for interaction studies, this new approach could advantageously complement the ^{1}H-^{15}N spectra that are commonly used for such purposes. Unlike the CH_3-TROSY experiments however, the CH_2-TROSY does not offer the same gains in resolution and sensitivity, due to the location of the methylene of the glycines directly on the protein backbone and therefore the lower dynamics that accompany this location as well as the strong auto-relaxation phenomenon not so significant for methyls, found at the end of side chains. Nevertheless, the methylene of glycine residues can act as a noteworthy local probe of structure and dynamics of the backbone of proteins. In addition, this labeling strategy and insertion of the CH_2-TROSY experiment into 3D spin state selective experiments has enabled a number of interesting coupling constant measurements at the backbone level for more in-depth structural and dynamics studies of glycines in proteins.

Dissolution Dynamic Nuclear Polarization is an emerging technique in NMR which can increase signal to noise ratio of atoms by several orders of magnitude. However, it presents two major inconveniences of which the high temperature suffered by the polarized sample and the time it takes to get it to the spectrometer. Indeed, dissolution reaches near boiling temperatures which can irreversibly denature a protein sample and the transfer can last a few seconds. As the length of polarization depends on T_1 relaxation times, this can be a major drawback for protein studies at room temperature, their T_1 usually averaging around 1 second as widespread proton presence is a factor of intense dipolar interactions which decrease the T_1. The third labeling strategy explored here of random fractional deuteration by conventional protein expression techniques has allowed to considerably extend T_1 relaxation times of protons in proteins for applications in protein D-DNP. In a cheap and common manner, using a high cell density expression protocol already published and requiring limited amounts of medium for high protein yields, the WW domain of Pin1 was expressed with deuterium densities up to 92.6%, as estimated by mass spectrometry. This extended relaxation time of different protons by several factors according to our calculations. Moreover, the WW domain possesses thermodynamic properties compatible which high temperatures. We have defined the conditions for the preparation of a “DNP juice”, containing high concentration of proteins and allowing the hyperpolarization to take place. Although the dissolution process still has to be adapted to avoid the formation of bubbles during the transfer, the very high level of polarization

obtained on protein proton is very promising. This labeling strategy for D-DNP could open new routes for biomolecular studies, of which structural and dynamic studies, or ligand screening experiments.

Annex 1: Assignment table of the H23 protein (chemical shifts units in ppm)

AA	N°	HN	N	Cα	Cβ	C	Cα-1	Cβ-1	C-1
M	**1**								
S	**2**								
G	**3**								
S	**4**								
H	**5**								
H	**6**								
H	**7**								
H	**8**								
H	**9**								
H	**10**								
S	**11**								
S	**12**	8,41	118,80	55,92	61,19	172,29	55,54	61,12	171,58
G	**13**	8,31	111,35	42,62		171,59	55,97	61,19	172,29
I	**14**	7,89	120,20	58,80	36,06	173,80	42,62		171,58
E	**15**	8,44	124,90	54,39	27,27	174,39	58,79	36,01	173,82
G	**16**	8,24	110,29	42,86		171,78	54,36	27,25	174,41
R	**17**	8,05	120,98	53,74	27,85	174,28	42,85		171,80
G	**18**	8,33	110,20	42,76		171,30	53,76	27,86	174,31
R	**19**	7,94	120,81	53,43	28,13	173,33	42,70		171,32
L	**20**	8,10	123,88	52,44	39,49	174,12	53,38	28,14	173,35
I	**21**	7,93	123,22	58,08	35,90	173,05	52,46	39,45	174,13
K	**22**	8,31	127,15	53,46	30,42	173,02	58,07	35,89	173,06
R	**23**	8,24	124,35	50,91	27,87	170,14	53,50	30,40	173,03
P	**24**								
V	**25**	8,34	115,60	56,46	33,10	172,43	60,08	29,57	173,00
V	**26**	8,32	123,57	60,90	28,65	173,24	56,53	33,08	172,42
S	**27**	9,19	125,32	55,18	61,94	171,82	60,87	28,62	173,27
G	**28**	7,21	108,36	43,28		167,22	55,21	61,95	171,84
V	**29**	8,31	122,60	58,68	32,45	170,26	43,21		167,24
A	**30**	8,97	126,84	47,20	21,34	172,47	58,76	32,42	170,29
S	**31**	8,55	114,53	54,13	63,95	169,99	47,22	21,28	172,46
L	**32**	8,11	123,04	51,66	39,50	174,55	54,04	63,90	170,01
G	**33**	8,35	110,44	43,37		170,89	51,66	39,50	174,56
Y	**34**	7,58	119,01	55,06	39,51	173,87	43,38		170,90
E	**35**	9,04	122,43	54,07	27,40	175,96	55,09	39,51	173,90
E	**36**	9,37	126,20	58,64	26,38	176,30	54,10	27,41	175,99
Q	**37**	8,93	115,28	56,76	25,02	175,04	58,67	26,32	176,32
E	**38**	7,20	119,05	55,98	26,72	176,14	56,79	25,04	175,03
V	**39**	7,11	119,57	64,28	28,77	174,65	55,92	26,73	176,16
L	**40**	7,84	119,42	55,39	39,34	174,80	64,29	28,63	174,68

K	**41**	8,13	119,18	57,94	30,24	175,28	55,36	39,37	174,83
M	**42**	7,48	116,45	56,63	31,57	174,27	57,94	30,26	175,31
A	**43**	8,09	119,60	51,94	16,47	176,28	56,61	31,57	174,26
A	**44**	8,78	119,28	52,31	16,12	175,43	51,97	16,45	176,29
A	**45**	7,55	119,80	52,28	15,90	177,09	52,28	16,16	175,47
V	**46**	7,38	116,51	63,88	29,31	174,68	52,22	15,90	177,13
E	**47**	8,01	115,60	53,22	25,87	174,80	63,82	29,36	174,70
K	**48**	7,25	120,39	57,14	30,34	174,79	53,10	25,94	174,88
T	**49**	7,68	107,80	59,14	66,50	171,30	57,09	30,31	174,79
A	**50**	7,59	126,96	48,76	18,24	175,26	59,17	66,45	171,32
T	**51**	8,33	107,69	58,63	67,03	171,64	48,79	18,23	175,29
H	**52**	6,92	121,20	51,71	27,34	171,51	58,61	67,17	171,65
P	**53**								
I	**54**	8,54	119,37	62,65	34,54	173,90	63,64	30,10	176,27
A	**55**	7,17	123,52	52,52	15,81	176,75	62,64	34,51	173,93
K	**56**	7,72	115,53	55,97	29,03	175,88	52,41	15,81	176,78
A	**57**	7,69	121,33	52,35	15,24	178,10	56,02	29,02	175,91
I	**58**	7,59	120,97	63,54	35,43	174,25	52,27	15,26	178,10
V	**59**	8,08	120,46	64,73	29,06	175,83	63,57	35,41	174,28
N	**60**	8,61	118,13	53,43	35,53	175,32	64,63	29,05	175,82
E	**61**	7,73	122,90	56,17	26,45	175,84	53,38	35,53	175,34
A	**62**	7,66	121,08	52,96	15,24	176,43	56,19	26,40	175,87
E	**63**	8,25	118,10	56,61	26,62	178,55	52,95	15,22	176,48
S	**64**	8,31	118,87	58,88	60,15	172,71	56,62	26,53	178,56
L	**65**	7,13	120,90	51,82	40,02	173,08	58,81	60,22	172,75
N	**66**	7,96	116,54	51,50	34,44	171,89	51,77	40,06	173,10
L	**67**	7,74	116,92	51,40	39,36	173,73	51,35	34,50	171,91
K	**68**	8,41	125,42	52,35	29,28	173,95	51,36	39,35	173,75
T	**69**	8,24	117,25	55,88	65,90	170,20	52,32	29,29	173,98
P	**70**								
E	**71**	8,30	120,31	54,14	27,78	173,75	59,11	30,13	173,04
T	**72**	8,12	112,36	56,81	69,72	171,53	54,16	27,67	173,78
R	**73**	8,25	118,60	52,18	30,80	173,18	56,81	69,71	171,55
G	**74**	8,68	109,90	44,73	0,00	171,66	52,22	30,75	173,19
Q	**75**	8,05	117,40	53,82	26,03	173,71	44,72	0,00	171,62
L	**76**	9,01	126,42	52,62	42,77	172,53	53,88	26,09	173,72
T	**77**	8,77	121,36	59,24	68,57	170,55	52,65	42,76	172,54
E	**78**	8,78	129,53	50,11	28,35	170,67	59,22	68,54	170,57
P	**79**								
G	**80**	8,39	112,93	42,10		174,50	61,55	29,36	174,51
F	**81**	7,85	119,50	57,32	37,19	171,51	42,06	0,00	170,57
G	**82**	7,14	105,75	44,08	0,00	167,91	57,30	37,14	171,56
T	**83**	8,34	116,76	58,78	69,81	168,60	44,09	0,00	167,94

L	**84**	9,34	127,20	51,23	44,69	171,94	58,70	69,85	168,60
A	**85**	9,25	121,58	49,34	20,21	171,17	51,23	44,63	171,97
E	**86**	8,94	122,20	53,33	28,54	173,17	49,38	20,30	171,17
I	**87**	8,72	126,21	58,20	37,95	173,39	53,34	28,55	173,20
D	**88**	9,28	129,75	52,63	37,07	173,40	58,12	37,97	173,40
G	**89**	8,59	103,59	42,79		171,14	52,63	37,04	173,40
R	**90**	7,53	119,67	50,99	31,68	172,55	42,73		171,14
F	**91**	9,18	122,67	57,39	37,91	172,39	50,86	31,72	172,55
V	**92**	9,28	123,28	58,47	32,41	171,08	57,46	37,91	172,33
A	**93**	8,99	127,22	47,10	22,35	172,87	58,43	32,35	171,10
V	**94**	9,15	120,02	57,33	32,49	173,75	47,03	22,36	172,88
G	**95**	8,84	115,50	43,91		169,04	57,28	32,42	173,74
S	**96**	8,41	119,22	55,65	61,75	171,59	43,90		169,10
L	**97**	8,71	123,46	56,82	38,60	176,15	55,65	61,79	171,60
E	**98**	8,93	118,42	57,48	26,17	174,82	56,80	38,68	176,13
W	**99**	7,55	120,00	59,27	26,50	176,35	57,48	26,12	174,86
V	**100**	8,77	118,28	64,55	29,16	175,43	59,26	26,47	176,38
S	**101**	8,21	113,07	58,82	60,48	174,77	64,57	29,10	175,46
D	**102**								
R	**103**								
F	**104**								
L	**105**								
K	**106**	7,55	120,04	52,73	29,19	171,54	54,09	39,56	175,27
K	**107**	8,32	125,10	50,15	30,03	174,02	52,72	29,16	171,55
N	**108**	8,21	119,70	51,52	37,68	172,71	50,12	30,03	174,03
D	**109**	8,74	123,09	51,87	39,56	175,17	51,46	37,63	172,71
S	**110**	8,83	121,57	59,15	60,20	174,19	51,85	39,56	175,20
S	**111**	8,50	117,80	58,57	59,71	174,33	58,95	60,20	174,21
D	**112**	7,67	124,10	53,91	37,75	175,12	58,57	59,70	174,37
M	**113**	7,34	118,58	54,58	27,46	176,43	53,85	37,76	175,13
V	**114**	7,58	121,37	63,36	29,16	176,40	54,60	27,46	176,44
K	**115**	7,63	121,70	56,97	29,62	176,23	63,35	29,17	176,43
L	**116**	8,19	120,53	55,73	39,44	175,20	56,85	29,64	176,27
E	**117**	7,97	117,70	56,89	26,47	176,31	55,64	39,52	175,19
S	**118**	8,00	115,10	58,79	60,25	174,75	56,86	26,52	176,34
L	**119**	8,26	124,20	55,17	39,77	177,31	58,70	60,26	174,82
L	**120**	7,92	120,10	55,17	38,85	175,52	55,13	39,75	177,35
D	**121**	8,18	119,30	54,80	38,08	176,30	55,10	38,78	175,50
H	**122**	7,95	118,03	56,15	26,56	175,00	54,84	38,08	176,35
K	**123**	7,90	119,90	55,10	29,66	175,74	56,16	26,64	175,06
L	**124**	7,86	119,30	53,22	38,94	175,53	55,29	29,61	175,81
S	**125**	7,80	115,66	57,19	60,69	172,02	53,21	38,89	175,56
N	**126**	7,94	119,90	50,95	36,36	172,83	57,19	60,68	172,03

T	**127**	7,74	112,94	59,14	67,32	171,93	50,92	36,37	172,82
S	**128**	8,18	117,76	55,88	61,07	172,24	59,17	67,32	171,95
S	**129**	8,30	118,17	56,39	60,90	172,52	55,82	61,04	172,25
T	**130**	7,96	114,81	59,65	66,71	172,32	56,39	60,84	172,54
S	**131**	7,99	118,10	56,09	61,06	171,96	59,64	66,68	172,32
R	**132**	8,09	122,59	54,07	27,65	173,37	56,08	61,07	171,99
Y	**133**	7,92	120,00	55,21	36,10	173,29	54,09	27,67	173,38
S	**134**	8,05	117,17	55,33	61,16	170,67	55,22	36,08	173,26
K	**135**	7,96	122,70	52,98	31,46	172,77	55,34	61,14	170,70
T	**136**	8,62	121,02	59,73	67,74	169,80	53,01	31,48	172,77
V	**137**	8,41	127,29	58,62	30,95	172,54	59,71	67,70	169,84
V	**138**	8,77	123,54	57,26	31,48	171,17	58,67	31,00	172,54
Y	**139**	9,35	123,72	55,81	38,08	171,42	57,32	31,42	171,21
V	**140**	8,30	119,84	57,67	31,58	172,61	55,83	38,08	171,45
G	**141**	9,59	115,90	41,92		167,75	57,61	31,44	172,64
R	**142**	8,53	121,57	51,68	32,33	170,74	41,85		167,76
E	**143**	8,29	130,50	55,69	28,43	172,67	51,68	32,31	170,77
G	**144**	9,07	116,21	42,26		171,03	55,64	28,42	172,69
E	**145**	7,92	117,87	54,35	30,67	173,81	42,25		171,02
G	**146**	7,66	107,50	41,92		171,16	54,23	30,62	173,84
I	**147**	9,46	124,41	60,03	33,70	174,41	41,91		171,13
I	**148**	9,21	118,92	58,26	37,54	173,05	59,93	33,67	174,44
G	**149**	7,92	108,83	43,84		166,74	58,17	37,59	173,05
A	**150**	8,97	121,95	48,50	20,78	171,35	43,86		166,75
I	**151**	9,06	121,04	57,31	39,12	171,02	48,52	20,72	171,33
A	**152**	8,83	129,35	47,77	18,76	173,10	57,31	39,12	171,09
I	**153**	8,95	124,62	55,82	36,88	173,10	47,82	18,77	173,11
S	**154**	8,95	123,55	54,97	62,63	169,71	55,79	36,90	173,16
D	**155**	7,98	127,99	52,70	39,68	177,74	55,03	62,60	169,73

Annex 2 : Production and purification of the PreScission enzyme

1000 mL of LB medium pH 7.2 (10 g casein peptone/L, 5 g yeast extract/L, 10g NaCl/L) was prepared in H_2O, in 330 mL batches and autoclaved. Two 50 mL LB cultures were inoculated with 1 mL of glycerol stock BL21 (DE3) bacteria containing the pet15B plasmid with the corresponding sequence, 1mg/mL of ampicillin and 0,3mg/mL of chloramphenicol. These were left overnight, shaking at 220 rpm, 30°C. The next morning, when DO_{600} reached between 2 and 3, the 330 mL of LB medium were inoculated with the precultures to reach an OD_{600} of 0.05 with 1mg/mL of ampicillin and 0.3mg/mL of chloramphenicol. The cultures are left shaking at 200 rpm for three to four hours at 37°C, until the OD_{600} reaches between 0.6 and 0.8. Cell are induced with IPTG at 1 mM concentration, returned to the incubator for overnight induction shaking at 220 rpm, 25°C. The next day, cells are centrifuged 20 minutes at 4000g 4°C. Supernatant is discarded, and cells are resuspended in 40 mL of lysis buffer. Lysis buffer is constituted of 30 mM Phosphate buffer pH 7.8, 50 mM NaCl, 30 mM Imidazole, 1mM DTT, 1% Tween, 1 mM PMSF, 1mg/L Pepstatin A and 1 mg/L Leupeptin. Cells are sonicated 10 times for 15 seconds, 40 amp, with one minute on ice between sonication cycles. The lysate is incubated 15 minutes at room temperature with 2µg of benzonase and 2mM $MgCl_2$. The lysate is then centrifuged 30 minutes at 25000 g 4°C and the supernatant used for purification by His-trap column. The supernatant is filtered in a 0.45 µm filter. The column was equilibrated with phosphate buffer 20 mM pH 7.8, 500 mM NaCl, 30 mM imidazole, 1 mM DTT and eluted with phosphate buffer 20 mM pH 7.8, 500 mM NaCl, 500 mM imidazole, 1mM DTT with a 0-80% gradient in 40 minutes at 1 mL per minute. By monitoring absorbance at 280nm, the protein eluted can be recuperated in 2mL fractions with the elution peak at 32,5 % of elution buffer, estimated at 162,5 mM imidazole. Yields range from 10 to 70 mg/L of culture at this point.

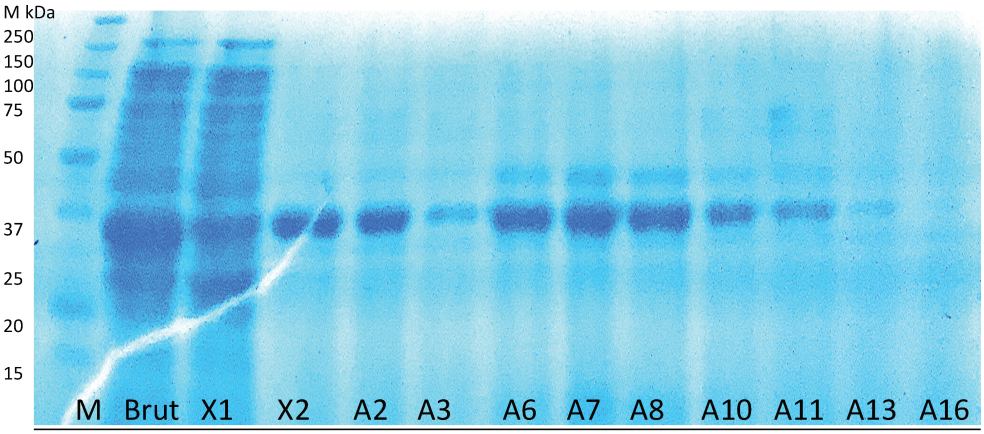
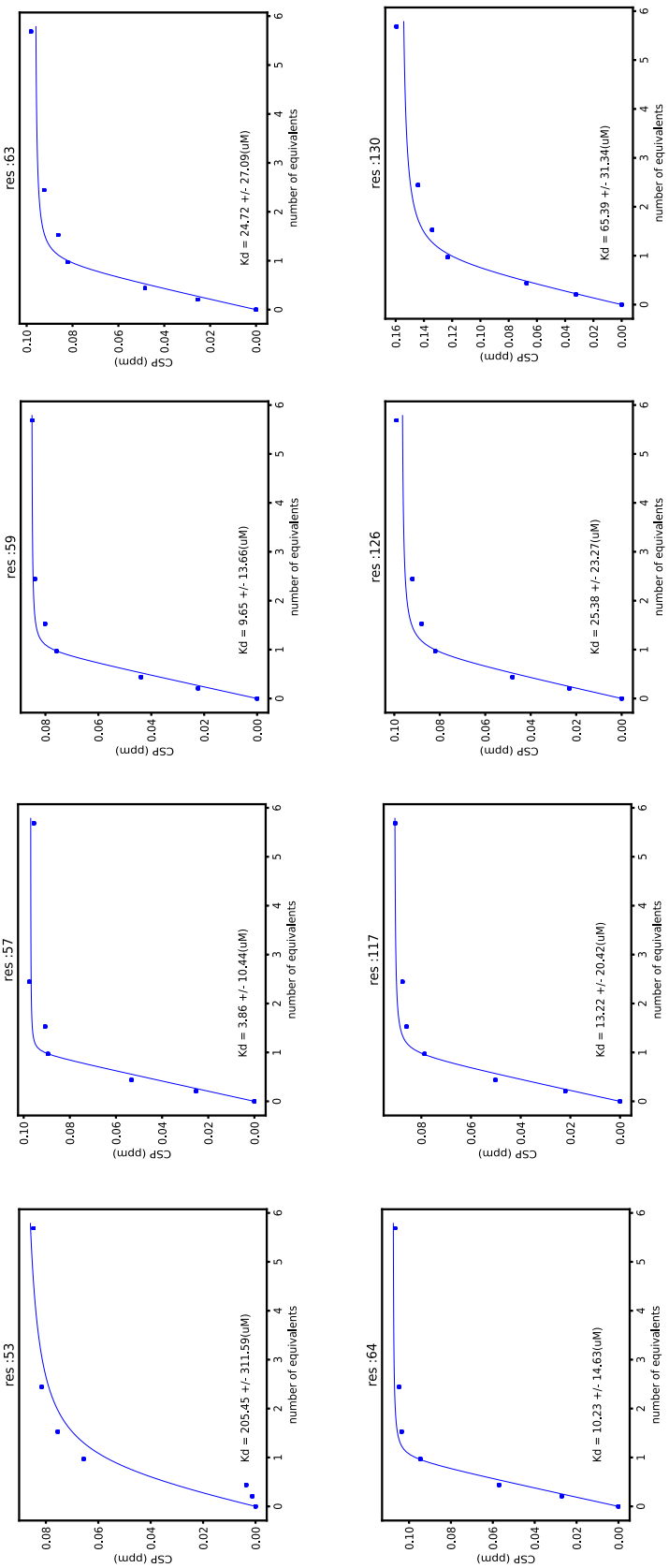
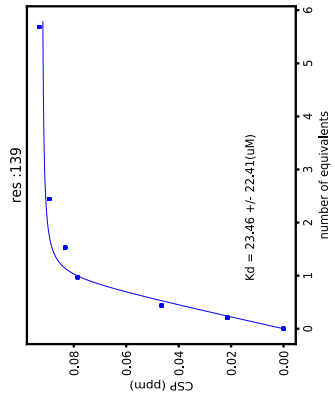


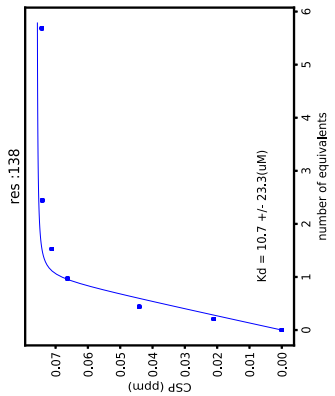
Figure 1 : 20% Polyacrylamide gel of the elution fractions of PreScission enzyme after His-trap Column. M: ladder, X1 : Loading elution, X2: rinsing elution, A1→ A16 elution fractions.

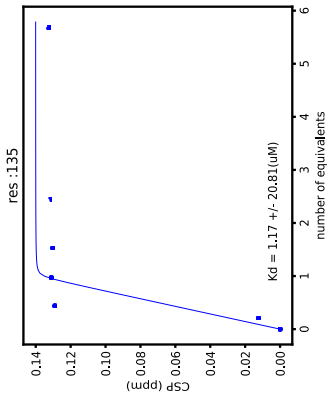
The elution fractions with the highest yields are pooled together and glycerol is added to 20%. Aliquots of 120 µL are prepared, with a final concentration of 1,1 mg/mL of PreScission estimated at 0,1 mg per aliquot.

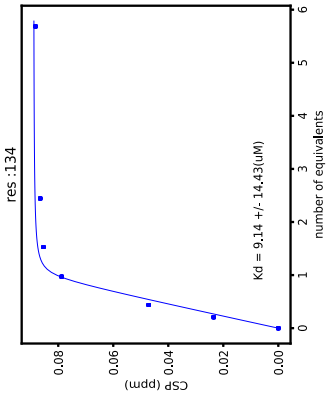
Annex 3: ^{1}H-^{15}N HSQC titration experiments of PPIase against increasing concentrations of CZ181 ligand

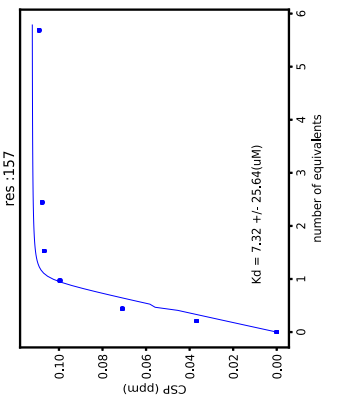


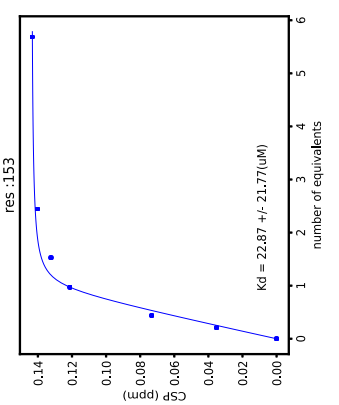

res :139
Kd = 23.46 +/- 22.41(uM)
CSP (ppm)
number of equivalents


res :138
Kd = 10.7 +/- 23.3(uM)
CSP (ppm)
number of equivalents


res :135
Kd = 1.17 +/- 20.81(uM)
CSP (ppm)
number of equivalents


res :134
Kd = 9.14 +/- 14.43(uM)
CSP (ppm)
number of equivalents


res :157
Kd = 7.32 +/- 25.64(uM)
CSP (ppm)
number of equivalents


res :153
Kd = 22.87 +/- 21.77(uM)
CSP (ppm)
number of equivalents

Annex 4: Table of proton-proton distances for glycine residues of H23

For each methylene proton of glycines of H23, numbered in parenthesis (and not part of the linker) are given, from left to right, the residue with the numbering found in the PDB data bank (pdb number : 5LBK), the proton of the residue, the distance in Å between the methylene proton and the proton of the residue, the distance as $1/(r^6)$. Follows 8 columns which are, for each labeling type (H_2O uniformly labelled, H_2O glycine specifically labeled, D_2O uniformly labeled and D_2O glycine specifically labeled) the sum of the $1/(r^6)$ that takes into account the gyromagnetic ratio and the percentage of the $^1H/^2H$ proportions (first column), and the second column the ratio: $1/(SUM)^{\frac{1}{6}}$.

GLY562 (28)		A	1/r^6	SUM H2O U	1/(SUM)^1/6	SUM H2O G	1/(SUM)^1/6	SUM D2O U	1/(SUM)^1/6	SUM D2O G	1/(SUM)^1/6
HA2											
VAL560	HG12	4,06	2,23E-04								
VAL560	HG13	4,54	1,14E-04								
SER561	H	4,73	8,90E-05								
SER561	HA	4,48	1,24E-04								
SER561	HG	4,67	9,60E-05								
GLY562	H	2,58	3,42E-03								
VAL563	H	2,30	6,72E-03								
VAL563	HA	4,33	1,53E-04								
VAL563	HB	4,58	1,08E-04								
VAL563	HG22	4,75	8,70E-05								
VAL563	HG23	3,63	4,34E-04								
ALA686	H	4,76	8,60E-05								
ALA686	HB3	4,36	1,45E-04	0,011	2,127	0,009	2,178	0,002	2,884	0	3,829
HA3											
VAL560	HG12	4,91	7,10E-05								
SER561	H	4,55	1,12E-04								
SER561	HA	4,66	9,80E-05								
SER561	HB3	4,39	1,39E-04								
SER561	HG	4,00	2,44E-04								
GLY562	H	2,61	3,13E-03								
VAL563	H	2,89	1,71E-03								
VAL563	HA	4,37	1,44E-04								
VAL563	HG23	4,68	9,50E-05								
ALA686	H	3,97	2,55E-04								
ALA686	HB1	4,57	1,10E-04								
ALA686	HB2	4,15	1,95E-04								
ALA686	HB3	3,12	1,08E-03	0,007	2,293	0,005	2,436	0,002	2,751	0	4,065
GLY567 (33)											
HA2											
LEU566	H	4,88	7,40E-05								

LEU566	HA	4,45	1,28E-04								
LEU566	HB2	4,81	8,00E-05								
LEU566	HB3	4,51	1,19E-04								
GLY567	H	2,67	2,74E-03								
TYR568	H	3,32	7,50E-04								
TYR568	HA	4,59	1,07E-04	0,004	2,549	0,003	2,601	0,001	3,525	0	4,596
HA3											
LEU566	H	4,94	6,80E-05								
LEU566	HB2	4,53	1,15E-04								
LEU566	HB3	4,52	1,17E-04								
GLY567	H	2,17	9,62E-03								
TYR568	H	3,17	9,81E-04								
TYR568	HA	4,28	1,64E-04								
TYR568	HD2	4,53	1,16E-04	0,01	2,15	0,01	2,167	0,001	3,3	0	3,901
GLY608 (74)											
HA2											
ARG607	HA	4,20	1,83E-04								
ARG607	HG2	4,69	9,40E-05								
ARG607	HG3	4,30	1,59E-04								
GLY608	H	2,25	7,71E-03								
GLN609	H	2,29	6,94E-03								
GLN609	HA	4,35	1,48E-04								
GLN609	HB3	4,38	1,42E-04								
GLN609	HG2	4,82	7,90E-05	0,014	2,036	0,013	2,056	0,001	3,083	0	3,693
HA3											
ARG607	H	4,77	8,50E-05								
ARG607	HA	4,19	1,83E-04								
ARG607	HB3	4,82	7,90E-05								
ARG607	HG2	4,04	2,31E-04								
ARG607	HG3	4,18	1,86E-04								
GLY608	H	2,72	2,49E-03								
GLN609	H	2,92	1,62E-03								
GLN609	HA	4,32	1,53E-04								
GLN609	HB3	4,94	6,90E-05	0,005	2,445	0,004	2,53	0,001	3,16	0	4,38
GLY614 (80)											
HA2											
PRO613	HA	4,33	1,52E-04								
GLY614	H	2,14	1,04E-02								
PHE615	H	3,24	8,63E-04								
PHE615	HA	4,43	1,32E-04	0,01	2,138	0,01	2,148	0,001	3,482	0	3,889
HA3											
PRO613	HA	4,33	1,51E-04								
PRO613	HB2	4,47	1,25E-04								

PRO613	HB3	4,86	7,60E-05								
GLY614	H	2,64	2,93E-03								
PHE615	H	3,26	8,31E-04								
PHE615	HA	4,48	1,24E-04								
PHE615	HD2	4,23	1,73E-04								
PHE615	HE2	4,87	7,50E-05	0,004	2,498	0,003	2,576	0	4,704	0	4,481
GLY616 (82)											
HA2											
GLU612	H	4,70	9,20E-05								
PHE615	H	4,25	1,69E-04								
PHE615	HA	4,75	8,70E-05								
PHE615	HB2	4,87	7,50E-05								
GLY616	H	2,52	3,87E-03								
THR617	H	2,66	2,82E-03								
THR617	HA	4,39	1,41E-04								
THR617	HB	4,51	1,19E-04								
THR617	HG1	4,71	9,20E-05								
GLY629	HA2	3,42	6,30E-04								
GLY629	HA3	3,18	9,62E-04								
SER630	H	4,55	1,13E-04								
TRP633	HE3	4,75	8,70E-05								
TRP633	HZ3	4,32	1,53E-04								
TRP633	HH2	4,99	6,50E-05	0,009	2,201	0,009	2,205	0,003	2,7	0,003	2,713
HA3											
GLU612	H	4,90	7,20E-05								
GLU612	HB2	4,93	6,90E-05								
PHE615	H	4,70	9,30E-05								
PHE615	HA	4,57	1,10E-04								
PHE615	HB2	4,49	1,23E-04								
PHE615	HB3	4,30	1,57E-04								
GLY616	H	2,66	2,84E-03								
THR617	H	2,41	5,10E-03								
THR617	HA	4,34	1,49E-04								
THR617	HB	4,93	7,00E-05								
VAL628	H	4,48	1,23E-04								
GLY629	H	4,74	8,80E-05								
GLY629	HA2	3,42	6,27E-04								
GLY629	HA3	2,47	4,37E-03								
SER630	H	3,99	2,47E-04								
TRP633	HB3	4,74	8,80E-05								
TRP633	HE3	3,22	9,01E-04								
TRP633	HZ2	4,87	7,50E-05								
TRP633	HZ3	2,92	1,61E-03								

TRP633	HH2	3,88	2,94E-04								
VAL634	HG22	4,95	6,80E-05								
VAL634	HG23	4,76	8,60E-05	0,017	1,981	0,008	2,231	0,009	2,192	0,001	3,454
GLY623 (89)											
HA2											
GLU620	HG2	3,29	7,94E-04								
GLU620	HG3	4,12	2,05E-04								
ASP622	HA	4,41	1,37E-04								
GLY623	H	2,17	9,71E-03								
ARG624	H	3,27	8,19E-04								
ARG624	HA	4,39	1,39E-04								
ARG624	HG2	4,78	8,40E-05	0,011	2,125	0,01	2,171	0,002	2,919	0	3,83
HA3											
GLU620	HG2	4,49	1,23E-04								
ASP622	HA	4,13	2,03E-04								
ASP622	HB3	4,96	6,70E-05								
GLY623	H	2,67	2,75E-03								
ARG624	H	3,26	8,32E-04								
ARG624	HA	4,50	1,20E-04								
ARG624	HB2	4,93	7,00E-05								
ARG624	HG2	4,01	2,41E-04	0,004	2,504	0,003	2,596	0,001	3,211	0	4,482
GLY629 (95)											
HA2											
GLY616	HA2	3,42	6,30E-04								
GLY616	HA3	3,42	6,27E-04								
THR617	H	4,54	1,14E-04								
VAL628	HA	4,40	1,39E-04								
VAL628	HB	4,87	7,50E-05								
VAL628	HG12	3,59	4,70E-04								
VAL628	HG13	4,69	9,40E-05								
GLY629	H	2,52	3,92E-03								
SER630	H	2,84	1,92E-03								
SER630	HA	4,24	1,72E-04								
SER630	HB3	4,82	8,00E-05								
TRP633	HE3	4,69	9,40E-05								
VAL634	HG21	4,59	1,07E-04								
VAL634	HG22	4,94	6,90E-05								
VAL634	HG23	4,28	1,63E-04								
VAL672	HA	4,28	1,63E-04								
VAL672	HG11	4,34	1,50E-04								
VAL672	HG12	4,13	2,01E-04								
VAL672	HG13	3,02	1,31E-03								
VAL672	HG22	4,74	8,80E-05								

TYR673	H	4,29	1,60E-04	0,01	2,149	0,006	2,365	0,005	2,436	0	3,768
HA3											
GLY616	H	4,73	8,90E-05								
GLY616	HA2	3,18	9,62E-04								
GLY616	HA3	2,47	4,37E-03								
THR617	H	3,97	2,56E-04								
VAL628	H	4,57	1,09E-04								
VAL628	HA	4,25	1,70E-04								
VAL628	HG12	4,40	1,37E-04								
GLY629	H	2,68	2,70E-03								
SER630	H	2,35	5,90E-03								
SER630	HA	4,38	1,42E-04								
SER630	HB3	4,71	9,10E-05								
TRP633	HB3	4,28	1,64E-04								
TRP633	HE3	3,11	1,10E-03								
TRP633	HZ3	3,75	3,60E-04								
VAL634	H	4,80	8,20E-05								
VAL634	HG21	3,84	3,14E-04								
VAL634	HG22	3,80	3,32E-04								
VAL634	HG23	3,12	1,08E-03								
VAL672	HG13	4,38	1,41E-04								
TYR673	H	4,82	7,90E-05	0,018	1,961	0,009	2,207	0,01	2,17	0,001	3,418
GLY675 (141)											
HA2											
VAL626	HG13	4,96	6,70E-05								
ALA627	H	4,66	9,70E-05								
VAL634	HG11	3,80	3,31E-04								
VAL634	HG12	3,96	2,61E-04								
VAL634	HG13	4,01	2,39E-04								
PHE638	HD2	4,67	9,60E-05								
PHE638	HE2	4,11	2,09E-04								
VAL674	H	4,80	8,20E-05								
VAL674	HA	4,29	1,60E-04								
VAL674	HB	4,72	9,10E-05								
VAL674	HG12	3,90	2,84E-04								
GLY675	H	2,71	2,52E-03								
ARG676	H	2,36	5,75E-03								
ARG676	HA	4,34	1,49E-04								
ARG676	HB2	4,31	1,56E-04								
ARG676	HB3	4,56	1,11E-04								
GLY680	HA2	4,83	7,90E-05								
ILE681	H	4,61	1,04E-04								
ILE681	HA	2,46	4,52E-03								

ILE681	HG22	4,96	6,80E-05								
ILE681	HG23	4,25	1,70E-04								
ILE681	HD12	4,57	1,10E-04								
ILE681	HD13	3,54	5,05E-04								
ILE682	H	2,65	2,87E-03								
ILE682	HG12	2,86	1,81E-03								
ILE682	HG13	4,18	1,87E-04								
ILE682	HG21	4,51	1,19E-04								
ILE682	HG22	4,84	7,70E-05								
ILE682	HG23	3,30	7,75E-04								
ILE682	HD11	4,76	8,60E-05								
ILE682	HD13	4,70	9,30E-05								
GLY683	H	4,49	1,21E-04	0,021	1,901	0,011	2,124	0,011	2,119	0,001	3,32
HA3											
ALA627	H	3,89	2,89E-04								
ALA627	HB3	3,88	2,91E-04								
VAL634	HA	4,99	6,50E-05								
VAL634	HB	4,85	7,70E-05								
VAL634	HG11	2,86	1,82E-03								
VAL634	HG12	2,73	2,43E-03								
VAL634	HG13	2,61	3,18E-03								
VAL634	HG22	4,94	6,90E-05								
PHE638	HD2	3,60	4,57E-04								
PHE638	HE2	2,87	1,78E-03								
PHE638	HZ	4,91	7,10E-05								
VAL674	HA	4,28	1,62E-04								
VAL674	HG12	4,68	9,50E-05								
GLY675	H	2,35	5,91E-03								
ARG676	H	2,87	1,80E-03								
ARG676	HA	4,36	1,46E-04								
GLY680	HA2	4,67	9,60E-05								
ILE681	H	4,92	7,10E-05								
ILE681	HA	3,08	1,17E-03								
ILE681	HG12	4,94	6,90E-05								
ILE681	HG23	4,79	8,30E-05								
ILE681	HD11	4,52	1,18E-04								
ILE681	HD12	4,38	1,41E-04								
ILE681	HD13	3,14	1,04E-03								
ILE682	H	4,04	2,31E-04								
ILE682	HG12	4,35	1,47E-04								
ILE682	HG23	4,72	9,00E-05	0,021	1,903	0,008	2,233	0,014	2,042	0,001	3,29
GLY678 (144)											
HA2											

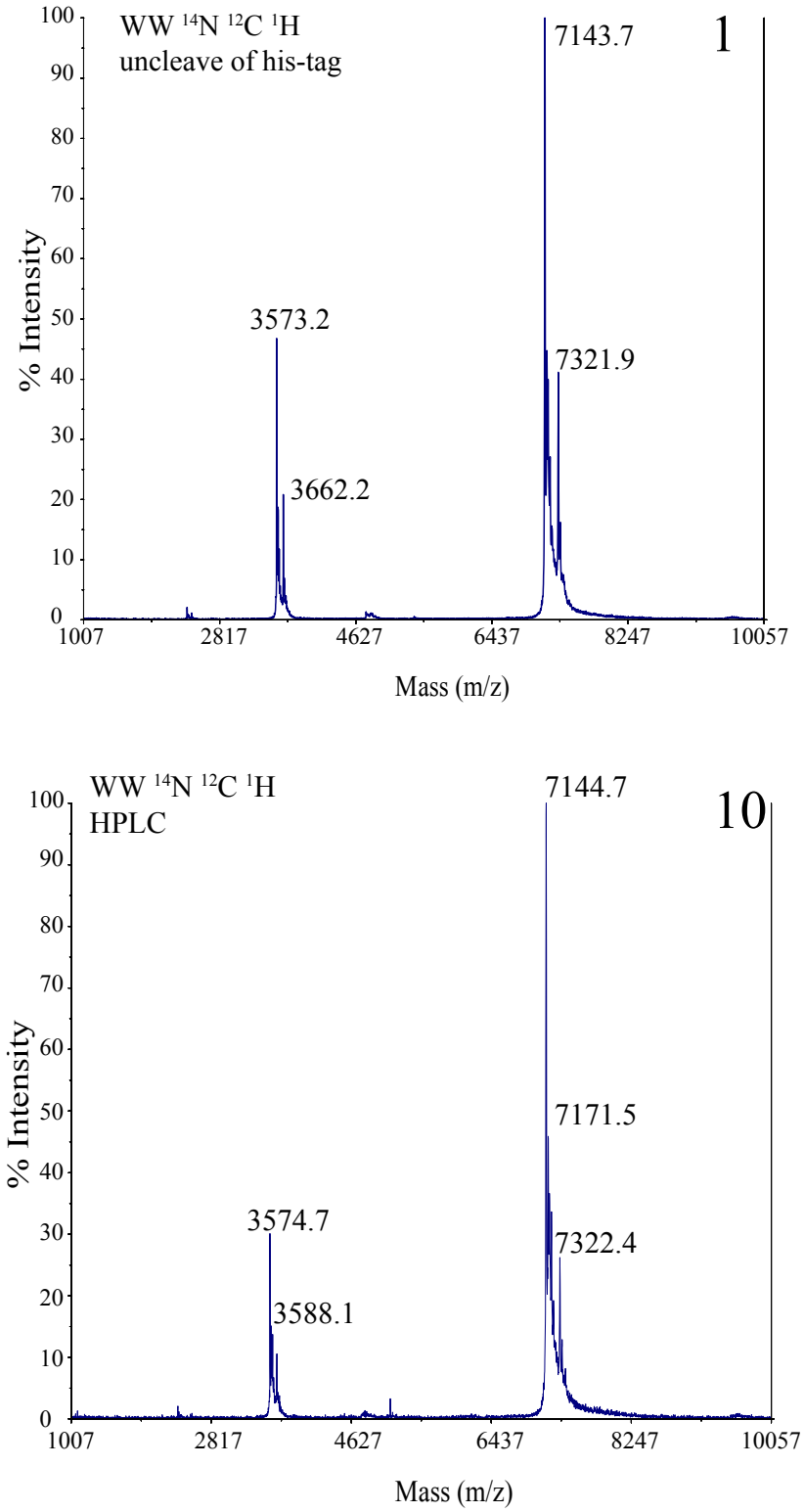
LYS640	H	4,41	1,37E-04								
LYS640	HB2	3,10	1,13E-03								
LYS640	HB3	3,47	5,78E-04								
LYS640	HG2	3,86	3,01E-04								
LYS640	HD3	4,38	1,42E-04								
GLU677	HA	4,36	1,45E-04								
GLU677	HG2	4,18	1,87E-04								
GLU677	HG3	4,70	9,30E-05								
GLY678	H	2,15	1,00E-02								
GLU679	H	3,30	7,76E-04								
GLU679	HA	4,46	1,27E-04								
GLU679	HG3	4,87	7,50E-05								
GLY680	H	4,86	7,60E-05	0,013	2,069	0,01	2,152	0,003	2,626	0	3,699
HA3											
ARG624	HD2	4,06	2,24E-04								
ARG624	HH11	4,11	2,07E-04								
ARG624	HH12	4,37	1,43E-04								
LYS640	HB2	4,54	1,15E-04								
LYS640	HB3	4,63	1,02E-04								
LYS640	HG2	4,72	9,10E-05								
GLU677	HA	4,34	1,50E-04								
GLU677	HB3	4,43	1,33E-04								
GLU677	HG2	4,24	1,71E-04								
GLY678	H	2,66	2,80E-03								
GLU679	H	3,23	8,74E-04								
GLU679	HA	4,46	1,28E-04								
GLU679	HB2	4,90	7,20E-05								
GLU679	HG2	4,67	9,60E-05								
GLU679	HG3	3,84	3,12E-04	0,005	2,398	0,003	2,579	0,002	2,81	0	4,233
GLY680 (146)											
HA2											
SER635	HA	4,67	9,60E-05								
PHE638	HB2	3,38	6,72E-04								
PHE638	HB3	4,46	1,26E-04								
PHE638	HD2	3,40	6,49E-04								
LYS640	HB3	3,99	2,48E-04								
LYS640	HG2	4,85	7,70E-05								
LYS640	HG3	4,37	1,44E-04								
LYS641	HA	4,28	1,63E-04								
GLY675	HA2	4,83	7,90E-05								
GLY675	HA3	4,67	9,60E-05								
ARG676	H	3,99	2,48E-04								
GLU679	H	4,06	2,23E-04								

GLU679	HB2	4,85	7,70E-05								
GLY680	H	2,17	9,51E-03								
ILE681	H	2,96	1,48E-03								
ILE681	HA	4,28	1,63E-04								
ILE681	HG12	4,43	1,33E-04								
ILE681	HD13	4,86	7,60E-05	0,013	2,059	0,01	2,138	0,003	2,622	0	3,681
HA3											
PHE638	HB2	4,78	8,40E-05								
PHE638	HD2	4,96	6,70E-05								
LYS640	HB3	4,15	1,95E-04								
LYS640	HG2	4,30	1,59E-04								
LYS640	HG3	3,62	4,46E-04								
LYS641	HA	3,82	3,20E-04								
ASN642	H	4,04	2,29E-04								
ASN642	HA	4,53	1,15E-04								
ASN642	HD21	4,02	2,35E-04								
ASN642	HD22	4,27	1,65E-04								
ARG676	H	4,80	8,20E-05								
GLU679	H	4,38	1,42E-04								
GLU679	HA	4,48	1,24E-04								
GLU679	HB2	4,79	8,30E-05								
GLU679	HB3	4,71	9,20E-05								
GLY680	H	2,68	2,73E-03								
ILE681	H	2,35	5,93E-03								
ILE681	HA	4,38	1,41E-04								
ILE681	HB	4,69	9,40E-05								
ILE681	HG12	4,37	1,44E-04	0,011	2,131	0,008	2,221	0,003	2,682	0	3,805
GLY683 (149)											
HA2											
ALA564	H	4,93	7,00E-05								
ALA564	HB1	4,58	1,09E-04								
ALA564	HB3	4,15	1,97E-04								
SER565	H	4,54	1,14E-04								
SER565	HA	2,54	3,76E-03								
SER565	HB2	4,50	1,21E-04								
SER565	HB3	4,44	1,31E-04								
LEU566	H	4,10	2,10E-04								
LEU566	HG	4,22	1,76E-04								
LEU566	HD11	4,93	7,00E-05								
LEU566	HD12	4,85	7,70E-05								
VAL573	HG11	4,07	2,20E-04								
VAL573	HG13	4,81	8,10E-05								
VAL573	HG21	4,19	1,85E-04								

VAL573	HG22	4,95	6,80E-05								
LEU650	HD11	3,27	8,14E-04								
LEU650	HD12	4,52	1,17E-04								
LEU650	HD13	3,53	5,17E-04								
LEU650	HD21	4,01	2,40E-04								
LEU650	HD22	4,24	1,73E-04								
VAL674	H	4,93	7,00E-05								
VAL674	HB	4,71	9,10E-05								
ILE681	HG21	4,97	6,60E-05								
ILE681	HG22	4,02	2,37E-04								
ILE681	HG23	4,70	9,30E-05								
ILE682	H	4,62	1,03E-04								
ILE682	HA	4,45	1,29E-04								
ILE682	HG13	4,53	1,16E-04								
GLY683	H	2,57	3,44E-03								
ALA684	H	2,55	3,62E-03								
ALA684	HA	4,42	1,35E-04								
ALA684	HB2	4,23	1,75E-04								
ALA684	HB3	4,36	1,45E-04	0,015	2,011	0,007	2,274	0,008	2,217	0,001	3,502
HA3											
ALA564	H	4,98	6,50E-05								
ALA564	HB3	4,96	6,70E-05								
SER565	HA	2,96	1,49E-03								
SER565	HB2	4,28	1,63E-04								
SER565	HB3	4,09	2,15E-04								
LEU566	H	4,61	1,05E-04								
VAL573	HB	4,53	1,16E-04								
VAL573	HG11	2,65	2,86E-03								
VAL573	HG12	4,12	2,04E-04								
VAL573	HG13	3,24	8,69E-04								
VAL573	HG21	3,04	1,27E-03								
VAL573	HG22	3,61	4,55E-04								
VAL573	HG23	4,55	1,13E-04								
LEU650	HD11	4,70	9,30E-05								
VAL674	H	4,55	1,12E-04								
VAL674	HB	3,57	4,86E-04								
VAL674	HG21	4,84	7,70E-05								
VAL674	HG23	4,74	8,80E-05								
ILE681	HG22	4,83	7,90E-05								
ILE682	H	4,48	1,24E-04								
ILE682	HA	4,64	1,01E-04								
ILE682	HB	4,70	9,30E-05								
ILE682	HG12	4,91	7,10E-05								

ILE682	HG13	3,53	5,16E-04								
GLY683	H	2,62	3,07E-03								
ALA684	H	2,50	4,11E-03								
ALA684	HA	4,23	1,74E-04								
ALA684	HB2	4,86	7,60E-05								
ALA684	HB3	5,00	6,40E-05	0,017	1,98	0,007	2,272	0,01	2,157	0,001	3,438

Annex 6: MALDI-TOF spectra of the WW domain before and after HPLC



Mass spectra of different samples of WW protein (uncleaved of the His-tag) before (upper panel) or after (lower panel) HPLC purification following the expression and purification protocol described in Chapters II and V and the MASS preparation method described in

Annex 7: Screen of the proximal protons between the CDC25 peptide similar and the WW domain, less than 2.5Å of distance to assess protons of proteins that could be influenced by ligand interaction.

10 complex structures between the WW domain and the pCD25 ligand are available from the PDB (number pdb: 1I8G[1]) Proton distances below 2.5 Å are listed in the table. Are given from left to right the name of the proton for the ligand, the name of the proton from the protein, the distance between them, the number of ligand protons less than 2.5 Å for one same protein proton, the frequency of interaction across the models as well as their average distance. In green are highlighted the two highest frequencies of interaction.

1. Wintjens, R., Wieruszeski, J.-M., Drobecq, H., Rousselot-Pailley, P., Buée, L., Lippens, G. & Landrieu, I. ^{1} H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* **276,** 25150–25156 (2001).

LIGAND	PROTEIN	DISTANCE	N° protons	Frequency	Average
TPO5HA	16SERHB2	2,09	3,00	4,00	2,33
TPO5HG23	16SERHB2	2,48			
PRO6HD2	16SERHB2	2,41			
TPO5HB	16SERHG	2,45	4,00	5,33	2,41
TPO5HA	16SERHG	2,44			
TPO5HB	16SERHG	2,35			
TPO5HA	16SERHG	2,40			
TPO5HG22	17ARGH	2,10	2,00	2,67	2,29
TPO5HG23	17ARGH	2,47			
LEU4HB3	17ARGHA	2,19	6,00	8,00	2,33
LEU4HD11	17ARGHA	2,36			
PRO3HB2	17ARGHA	2,48			
LEU4HB3	17ARGHA	2,23			
LEU4HB3	17ARGHA	2,24			
LEU4HD21	17ARGHA	2,49			
LEU4HB3	17ARGHB3	2,47	3,00	4,00	2,38
LEU4HB2	17ARGHD3	2,26			
PRO3HG3	17ARGHD3	2,42			
LEU4HD21	17ARGHE	2,02	2,00	2,67	2,24
LEU4H	17ARGHE	2,46			
LEU4HD12	17ARGHG2	2,40	2,00	2,67	2,33
TPO5HG21	17ARGHG2	2,25			

LEU4HG	17ARGHG3	2,48	1,00	1,33	2,48
LEU4HB2	17ARGHH11	2,16	2,00	2,67	2,31
LEU4HD12	17ARGHH11	2,46			
PRO3HA	17ARGHH21	1,94	5,00	6,67	2,31
LEU4H	17ARGHH21	2,29			
PRO3HA	17ARGHH21	2,47			
PRO3HB2	17ARGHH21	2,42			
PRO3HA	17ARGHH21	2,45			
PRO6HD2	23TYRHB2	2,41	2,00	2,67	2,42
PRO6HG2	23TYRHB2	2,44			
PRO6HD2	23TYRHB3	2,29	5,00	6,67	2,37
PRO6HG2	23TYRHB3	2,41			
PRO6HD2	23TYRHB3	2,35			
PRO6HG2	23TYRHB3	2,38			
PRO6HD2	23TYRHB3	2,43			
PRO6HG2	23TYRHD1	2,47	1,00	1,33	2,47
PRO6HG2	23TYRHD2	2,25	3,00	4,00	2,29
PRO6HG2	23TYRHD2	2,21			
PRO6HD2	23TYRHD2	2,43			
LEU4HB3	23TYRHE1	2,31	1,00	1,33	2,31
PRO6HG3	32SERHB2	2,24	1,00	1,33	2,24
THR8HA	32SERHG	2,27	2,00	2,67	2,19
PRO6HG2	32SERHG	2,12			
VAL7HA	33GLNHA	2,36	3,00	4,00	2,33
VAL7HG21	33GLNHA	2,43			
VAL7HA	33GLNHA	2,21			
THR8HG21	33GLNHB2	2,58	2,00	2,67	2,49
THR8HG23	33GLNHB2	2,39			
THR8HG22	33GLNHB3	2,36	2,00	2,67	2,42
THR8HB	33GLNHB3	2,48			
THR8HB	34TRPH	2,20	2,00	2,67	2,24
VAL7HG21	34TRPH	2,29			
THR8HB	34TRPHB2	2,46	1,00	1,33	2,46
VAL7HG13	34TRPHD1	2,15	4,00	5,33	2,39
VAL7HA	34TRPHD1	2,50			
VAL7HB	34TRPHD1	2,50			
VAL7HA	34TRPHD1	2,41			
VAL7HG23	34TRPHE1	2,47	4,00	5,33	2,25
VAL7H	34TRPHE1	2,19			

VAL7HG11	34TRPHE1	2,11			
VAL7HA	34TRPHE1	2,25			
TPO5HG21	34TRPHZ2	2,44	12,00	16,00	2,32
TPO5HB	34TRPHZ2	2,49			
TPO5HG21	34TRPHZ2	2,16			
TPO5HA	34TRPHZ2	2,26			
TPO5HB	34TRPHZ2	2,28			
TPO5HG21	34TRPHZ2	2,36			
TPO5HG23	34TRPHZ2	2,31			
TPO5HB	34TRPHZ2	2,17			
TPO5HB	34TRPHZ2	2,32			
TPO5HG22	34TRPHZ2	2,47			
TPO5HB	34TRPHZ2	2,44			
TPO5HB	34TRPHZ2	2,19			
			75,00	100,00	2,34