

Novel Isotope Labeling Strategies for NMR-Based Structure Determination of High Molecular Weight Proteins

by

Gaddafi Ibrahim Danmaliki

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry
University of Alberta

© Gaddafi Ibrahim Danmaliki, 2023

Abstract

Solution NMR studies of high molecular weight proteins suffer from inadequate side-chain information and rapid signal decay. Selective ^1H , $^{13}\text{CH}_3$ -methyl labelling of perdeuterated proteins has been the standard approach addressing these limitations but suffers from a lack of information about the non-methyl-containing amino acids. Here, we present methods for introducing additional isolated ^1H - ^{12}C groups into proteins using inexpensive metabolic precursors and inhibitors.

Using fumarate as a carbon source for *Escherichia coli* (*E. coli*) in D_2O allows for stereospecific incorporation of ^1H at β_2 and ^2H at β_3 positions of the oxaloacetate family of amino acids (Asp, Asn, Met, and Lys). We demonstrate an efficient and inexpensive synthetic protocol for producing phenylpyruvate, hydroxyphenylpyruvate, and anthranilate precursors for labelling Phe (δ , ζ), Tyr (δ), and Trp (δ , η , ϵ), respectively. We labeled the imidazole ring of histidine by supplementing *E. coli* growth medium with the unlabelled amino acid.

The ^1H - ^{12}C groups introduced in eight non-methyl-containing amino acids complement methyl groups and can be connected to ^1H - ^{15}N amide positions using multi-dimensional through-space NOESY experiments. We demonstrate the utility of our approach by labelling the outer membrane protein PagP in DPC detergent micelles (>50 kDa), yielding rich new structural restraints for chemical shift assignments and high-resolution structure determination.

Solution NMR is the ideal technique for probing protein motion over various timescales. Most NMR studies frequently focus on quantifying only the backbone dynamics by measuring ^{15}N relaxation properties in uniformly labelled proteins. Side chain motion also plays a pivotal role in many biological processes. However, characterizing side chain dynamics remains difficult for

solution NMR. We introduce a new method for assessing protein side-chain motion using ^1H NMR relaxation rates, extending the dynamic analysis to exciting unstudied protein sites.

Preface

The introduction in chapter 1 and the conclusions in chapter 5 are my original work. The remaining chapters have been published previously or submitted for publication. The contributions made by the author, Gaddafi I. Danmaliki, and the coauthors of these studies are described below.

A portion of chapter 1 has been published as Danmaliki, G.I., Hwang P.M. Solution NMR spectroscopy of membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1862, 183356 (2020). GID wrote the manuscript, and PMH edited it.

Chapter 2 has been published as Danmaliki, G.I., Liu, P.B., Hwang, P.M. Stereoselective deuteration in aspartate, asparagine, lysine, and methionine amino acid residues using fumarate as a carbon source for *Escherichia coli* in D₂O. *Biochemistry* 56, 45, 6015–6029 (2017). PMH directed the research. PBL helped with protein purification. GID expressed and purified the human Pin1 WW domain protein and its constructs, tested several metabolic precursors and inhibitors, acquired all the 2D and 3D NMR data, assigned the protein chemical shifts, and analyzed the data under PMH supervision. GID wrote the manuscript, and PMH edited it.

We have submitted Chapter three to be published as Danmaliki, G.I., Yu, S., Braun, S., Zhao, Y.Y., Moore, J., Fahlman, R.P., West, F.G., Hwang, P.M. Cost-effective selective deuteration of aromatic amino acid residues produces long-lived solution ¹H-NMR magnetization in proteins. *Angew. Chem. Int.* (2022). PMH directed the research. SY synthesized the aromatic precursors under FGW supervision. GID expressed and purified the protein constructs, tested many metabolic precursors and inhibitors, acquired, and analyzed the NMR data under PMH supervision. SB and YYZ analyzed the protein samples by mass spectrometry under JM and RPF supervision. GID wrote the first draft of the manuscript under PMH supervision.

Chapter 4 has been published as Danmaliki, G.I., Hwang, P.M. Proton TOCSY NMR relaxation rates quantitate protein side chain mobility in the Pin1 WW domain. *J Biomol NMR* 76, 121–135 (2022). PMH directed the research, designed the NMR pulse sequence, and acquired the NMR data. GID purified the protein constructs, analyzed the NMR experiments, and wrote the first draft of the manuscript.

Dedication

To my late mother, Hajiya Sadiya Abdullahi. Your love, wisdom, and presence will forever be with us. Like you nurtured me when I was little, may Allah's mercy always be with you.

Acknowledgments

All praise is due to Allah, with whose grace all good works come to completion. Thank you for giving me the wisdom, strength, health, and numerous other blessings to accomplish this work.

My utmost gratitude goes to my supervisor Dr. Peter M. Hwang, for his guidance, friendship, and incredible patience. Thank you for your close and continued mentorship, teaching me about solution NMR. Working in your lab on some novel, exciting, and challenging projects has been the most rewarding scientific experience of my career. You have been generous to us through the Hwang Professional Corporation, and we are eternally grateful. We have shared many beautiful memories throughout my stay in Canada, and I will cherish them for the rest of my life. My appreciation also goes to the members of my supervisory committee: Dr. Michael Overduin and Dr. Joanne Lemieux, for their time, support, and insights offered during committee meetings and for providing me with valuable advice about my career.

My special thanks to our collaborators: Dr. Fredrick G. West and Dr. Shaohui Yu, for contributing immensely and selflessly synthesizing many of the aromatic precursors used in this thesis. A special thanks go to Dr. Richard Fahlman and Shelly Braun for analyzing our data using mass spectrometry.

Thanks to Dr. Brian D. Sykes for accommodating me in his lab at the initial stages of my program and enabling a conducive environment to operate his 500 MHz spectrometer. We have shared many beautiful memories celebrating birthdays, golf tournaments, lab lunches, and the fascinating discussions of having four wives. I would also like to thank Dr. Monica X. Li for teaching me how to operate the spectrometer and analyze NMR experiments. Thanks to Philip B. Liu for teaching me protein expression and purification. Phil has undoubtedly been a good friend and has made my stay at the lab enjoyable. Thanks to Dr. David Westaway and his lab members, particularly Dr. Serene Wohlgemuth, Dr. Ghazaleh Eskandari-Sedighi, and Dr. Zelin Fu, for guiding me during rotation at the Center for Prions and Protein Folding Diseases.

I want to use this opportunity to thank my lab mate and a good friend, Dr. Zabed Mahmud, his wife, Farzana Khan, and Faiyaz. We started this journey together, young and inexperienced, sticking together through thick and thin and evolving into the enlightened trainees we have now become. Thank you for your friendship and kind support to Hamida and I. My sincere gratitude

goes to the past and present colleagues at the lab: Dr. Somaya Zahran, Dr. Fangze Cai, Rafid M. Feisal, Kevin Mak, Rozhan Sonboli Peter Kim, Prab Paul Dhani, Violet, and Mohammad Hemadi. I appreciate you all for creating a conducive environment for me at the lab. I would also like to thank the friends I made in the department: Suad Rashid, Muhammad Bashir Khan, Charnpal Grewal, Dean Schieve, Nishad Rathod, Rashmi Panigrahi, and Hyeong Jin Kim. Meeting you has made my graduate studies a fulfilling experience.

I would also like to thank Dr. Aminu Bello and his wife, Mrs Sadiya A. Bello, for being my family's most vital support in Edmonton and connecting me to many Hausa-speaking families in Alberta. You have been my role models, and I'm forever grateful for meeting you during my graduate program. I would also like to thank Isah, and his wife, Hauwa, for helping us immensely in our most profound time of need in Edmonton.

I want to express my gratitude to the Natural Sciences and Engineering Research Council (NSERC) for funding my stay in Canada and the generous donors that supported me financially through the University of Alberta, the Faculty of Graduate Studies and Research, and the International Student Centre. I would also like to thank the Faculty of Medicine and Dentistry for the 75th Anniversary Graduate Student Award and the Usmanu Danfodiyo University Sokoto for laying the foundations to pursue an unconventional degree in North America.

I want to thank my father, Alhaji Ibrahim Danmalikin Gidan Goga, and my siblings for their lifelong support in my pursuit of excellence. Baba, you have always been a friend and a good role model, and I hope I have made you proud of what I have accomplished. Finally, I want to thank my wife, Hamida Bello, for her support, patience, love, and understanding. You have always been my pillar and a confidant. Thank you for taking good care of Ahmad and me.

Table of Contents

Chapter 1	1
Introduction	1
Overview.....	1
Solution NMR in structural biology	2
Solution NMR methods for tackling large protein systems.....	3
Deuteration with selective ^1H - ^{15}N protonation.....	3
Selective Methyl Group Protonation and Methyl-TROSY.....	5
Strategies for selective methyl group protonation in perdeuterated proteins	6
Isoleucine	7
Leucine and Valine	9
Methionine	10
Alanine and threonine.....	10
Combinatorial labelling of methyl-containing amino acids.....	11
Aromatic residue ^1H , ^{13}C -site-selective labeling	11
Stereo-Array Isotope Labeling (SAIL)	12
Our approach for overcoming the size limitation of solution NMR involves economic stereoselective deuteration of non-methyl-containing amino acid side chains.....	13
<i>Escherichia coli</i> bacteria as a model system for isotope labelling	14
Dynamics by Solution NMR.....	18
Backbone dynamics	18
Side chain dynamics	24
Summary.....	26
References.....	27
Chapter 2	36
Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for <i>Escherichia coli</i> in D_2O	36
Summary.....	36
Introduction.....	37
Experimental Procedures	38
Construct design.....	38
Production of (^{15}N , ^{13}C)-mutant human Pin1 WW domain in H_2O minimal M9 media	39

Production of (¹⁵ N, ¹² C)-mutant human Pin1 WW domain in D ₂ O-M9 minimal media.....	40
Rationale for isotope labeling using D ₂ O-“FROMP” minimal media.....	41
Production of mutant human Pin1 WW domain in D ₂ O-“FROMP” (fumarate, rhamnose, oxalate, malonate, and pyruvate) media	43
NMR Spectroscopy.....	43
Results and Discussion	44
Chemical shift assignment of Pin1 WW domain.....	44
Quantitation of ¹ H/ ² H incorporation in Pin1 WW domain grown in D ₂ O- ¹² C-glucose or D ₂ O-(FROMP) media.....	45
Pyruvate family (alanine, valine, leucine, isoleucine-γ2).....	48
Phosphoglycerate family (serine, cysteine, tryptophan, and glycine)	50
Oxaloacetate family (aspartate, asparagine, methionine, threonine, lysine, isoleucine)	51
α-ketoglutarate family (glutamate, glutamine, proline, arginine).....	54
Phosphoenolpyruvate + erythrose-4-phosphate family (tryptophan, phenylalanine, tyrosine) 56	
Ribose phosphate family (histidine)	59
Application of FROMP labeling to χ ₁ dihedral angle determination.....	60
Application of FROMP labeling technique for structure determination of large proteins	62
Conclusions.....	63
Acknowledgements.....	63
References.....	64
Chapter 3	69
Cost-effective selective deuteration of aromatic amino acid residues produces long-lived solution ¹H-NMR magnetization in proteins	69
Summary	69
Introduction.....	70
Experimental	71
Synthesis of selectively deuterated phenylpyruvate, hydroxyphenylpyruvate and anthranilate precursor compounds for aromatic residue labeling.....	71
Human Pin1 WW domain as an ideal system for monitoring precursor incorporation.....	73
Protocol for expressing Pin1 with metabolic precursors in D ₂ O	74
Results.....	74

Incorporation of selectively deuterated phenylpyruvate, hydroxyphenylpyruvate, and anthranilate precursors into Phe, Tyr, and Trp.	74
Phenylalanine and Tyrosine.....	74
Tryptophan.....	75
Incorporation of isolated ^1H - ^{12}C groups into the aromatic ring of histidine.....	76
Measurement of linewidth	77
Cycloserine prevents alanine-pyruvate interconversion.	78
The addition of unlabeled cysteine and methionine to <i>E. coli</i> growth media decreases isotope scrambling at threonine- γ^2	80
Methionine	82
Stereoselective ^1H - ^{12}C -labeling of beta methylene proton $\text{H}\beta^2$ in Asp, Asn, Lys, and other amino acid residues.....	82
Application of isolated ^1H - ^{12}C groups for structure determination of protein system > 30 kDa	83
Conclusion	85
Acknowledgement	86
References.....	86
Supplementary Information	89
Experimental Procedures	89
Synthesis of selectively deuterated phenylpyruvate and 4-hydroxyphenylpyruvate.....	89
Protocol for bacterial expression of Pin1 protein in 1 L H_2O -M9 minimal media.....	93
Protocol for bacterial expression of Pin1 protein with metabolic precursors and small molecule inhibitors in 1 L D_2O -M9 minimal media.....	93
Protocol for isoketovalerate and ketobutyrate deuteration	95
NMR spectroscopy.....	95
Amino acid digestion and Mass Spectrometry Analysis	95
Results.....	97
References.....	106
Chapter 4	107
Proton TOCSY NMR relaxation rates quantitate protein side chain mobility in the Pin1 WW domain.....	107
Summary.....	107

Introduction.....	108
Materials and Methods.....	113
Results and Discussion	114
Pin1 backbone dynamics.....	114
¹ H R _{DIPSI-2} relaxation rates	120
Cross-correlated relaxation rates of Pin1 residues.....	123
χ_1 dihedral angle estimated using maximum observed ³ J couplings	124
The estimation of maximum measured ³ J-coupling for Pin1 residues	127
Rigid Pin1 residues adopting a single dominant χ_1 dihedral angle rotamer	128
Pin1 residues averaging between two χ_1 dihedral angle rotamers.....	130
Methyl-containing residues.....	132
Correlations between maximum observed ³ J couplings and protein side-chain relaxation rates	133
Acknowledgements.....	136
References.....	136
Supplementary Information	138
Chapter 5	153
Conclusion and future directions	153
Limitations	155
Future directions	156
References.....	159
Bibliography	162

List of Tables

Table 2.1. Average percent isotopomers in the amino acids of mutant Pin1 WW domain grown in glucose/FROMP and D ₂ O.....	47
Table 2.2. Average percent isotopomers in methyl-containing residues of Pin1 WW domain grown in glucose/FROMP and D ₂ O	50
Table 3.S1. M9 Minimal Media Composition for Stereoselective Labeling of Proteins	94
Table 3.S2: Isotope Incorporation into Pin1 Amino Acids Monitored by Solution NMR and Mass Spectrometry	100
Table 3.S3: Linewidth measurement for the stereoselectively deuterated aromatic amino acids, methyl groups, and methylene proton of aspartate and asparagine.	106
Table 4.1. Chemical shifts and normalized values for maximum observed ³ J-couplings and ¹ H R _{DIPS12} relaxation rates for the mutant human Pin1 WW domain at the beta positions.	116
Table 4.S1. Chemical shifts and normalized values for maximum observed ³ J-couplings and ¹ H R _{DIPS12} relaxation rates for the mutant human Pin1 WW domain.	140

List of Figures

Figure 1.1 Effect of molecular size on signal decay in solution NMR.....	2
Figure 1.2. The range of biological assemblies above 100 kDa studied using the methyl-TROSY technique.....	6
Figure 1.3. Simplified biosynthetic steps involved in the metabolism of methyl-containing amino acids.	7
Figure 1.4. Expected chemical shifts of different methyl groups in 2D ^1H - ^{13}C correlation spectra.	8
Figure 1.5. Overview of α -ketoacid precursors used for labelling aromatic amino acids.....	12
Figure 1.6. The design concept of SAIL amino acids.....	13
Figure 1.7. The design concept for stereoselective deuteration of non-methyl-containing amino acids	14
Figure 1.8. The reaction catalyzed by PagP.	16
Figure 1.9. Solution structures of PagP in DPC and β -OG/SDS detergent micelles.....	16
Figure 1.10. The X-ray crystal structure of PagP in LDAO detergent micelles	17
Figure 1.11. Accessible timescales of NMR experiments.	19
Figure 1.12. Description of NMR experiments for measuring backbone dynamics	22
Figure 1.13. The three major χ_1 dihedral angle rotamers as viewed along the $\text{C}\alpha$ - $\text{C}\beta$ bond.	24
Figure 2.1 Mutant human Pin1 WW domain sequence.	39
Figure 2.2 Overview of amino acid biosynthetic pathways.	42
Figure 2.3. Stereoselective incorporation of ^2H from D_2O into the $\text{H}^{\beta 3}$ position in aspartate.	43
Figure 2.4. Natural abundance 2D ^1H - ^{13}C HSQC (aliphatic region).....	46
Figure 2.5. Natural abundance 2D ^1H - ^{13}C HSQC (aromatic region).....	48
Figure 2.6. Biosynthesis of serine from 3-phosphoglycerate.	51
Figure 2.7. Pathways for the biosynthesis of asparagine, lysine, threonine, and methionine.....	53
Figure 2.8. Biosynthesis of glutamate, glutamine, proline and arginine from α -ketoglutarate..	55
Figure 2.9. Average percent protonation observed at the aromatic sidechains of phenylalanine, tyrosine, tryptophan and histidine for Pin1(glucose) and Pin1(FROMP).....	57
Figure 2.10. Biosynthesis of tryptophan from chorismate.	57
Figure 2.11. Biosynthesis of phenylalanine and tyrosine.	58
Figure 2.12. Outline of the biosynthesis of histidine.	59

Figure 2.13. Application of FROMP labeling technique for stereospecific assignment of β -methylene protons.	61
Figure 3.1. (a) α -ketoacid precursors used for labeling Phe, Tyr, and Trp aromatic rings.	76
Figure 3.2: 2D ^1H , ^{13}C -HSQC spectra of Pin1 aromatic region.	77
Figure 3.3: 2D 1H-1H NOESY Spectra of Pin1 in D_2O	78
Figure 3.4: Cycloserine prevents alanine-pyruvate isotope scrambling.	80
Figure 3.5: Prevention of threonine- γ_2 isotope scrambling by supplementing unlabeled methionine and cysteine in deuterated <i>E. coli</i> growth media.	81
Figure 3.6. 3D NOESY spectra of PagP in DPC detergent micelles (50 kDa).	85
Figure 3.S1. Two dimensional ^1H - ^{15}N HSQC spectrum of Pin1.	97
Figure 3.S2. Biosynthesis of tryptophan from chorismate.	97
Figure 3.S3. Inhibition of aromatic amino acid biosynthesis.	98
Figure 3.S4. Biosynthesis of histidine starting from ATP and PRPP Precursors.	98
Figure 3.S5. Threonine, methionine, and lysine biosynthesis starting from aspartate.	99
Figure 3.S6. Stereoselective incorporation of ^2H from D_2O into the $\text{H}\beta^3$ position in aspartate... ..	99
Figure 4.1. Pulse scheme used to measure ^1H relaxation rates, using any multi-dimensional sequence preceded by a ^1H TOCSY sequence such as DIPSI-2.	110
Figure 4.2 Crystal structure of the human Pin1 protein rotated along its long axis showing the antiparallel β -strands in green (PDB id 2F21).	115
Figure 4.3 Effect of the DIPSI2 mixing scheme on rigid Y37 and a rotamer-averaging N44. .	123
Figure 4.4 Correlation plot of simulated cross-peak intensity ratio against ^3J couplings	126
Figure 4.5 The crystal structure of the human Pin1 protein shows side chains for which NMR J-coupling experiments have defined a dominant χ_1 dihedral angle.	127
Figure 4.6 The folding core of the human Pin1 WW domain shows the buried residues: Trp26 (sphere) and Asn40 (salmon stick).	129
Figure 4.7 NMR strip plots from 3D ^1H -TOCSY- ^{13}C -HSQC, HN(CO)H β , and HNH β for a rigid Pin1 residue, Tyr37, adopting the gauche- χ_1 rotamer (a), and averaging Lys28 sampling two χ_1 rotamers (b).	130
Figure 4.8 The alternative rotamers of Lys28 (a) and Asn44 (b).	132
Figure 4.9 Correlation plot of the normalized maximum observed ^3J coupling and normalized R_{DIPSI2}	135
Figure 4.S1 Backbone dynamics of Pin1 measured by ^{15}N T_1 and ^{15}N T_2 relaxation.	139

Figure 4.S2 Normalized β -methylene ^1H R_{DIPS12} relaxation rates plotted against Pin1 sequence.	149
Figure 4.S3 Normalized cross-correlated relaxation rates plotted as a function of Pin1 residues.	149
Figure 4.S4 Normalized ^3J -coupling experiments: ^3J $\text{H}\alpha\text{H}\beta$ (a), ^3J $\text{HNCOH}\beta$ (b), and ^3J $\text{HNH}\beta$ (c) plotted against the Pin1 sequence.	151
Figure 4.S5 The normalized maximum measured ^3J -couplings plotted against human Pin1 sequence.	151
Figure 4.S6 NMR strip plots from 3D ^1H -TOCSY- ^{13}C -HSQC, HNCOHB , and HNHB experiments for Asp 33 (a) and His41 (b).	152

Chapter 1

Introduction

A portion of this chapter has been published as Danmaliki, G.I., Hwang P.M. Solution NMR spectroscopy of membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1862, 183356 (2020)

Overview

Solution nuclear magnetic resonance (NMR) spectroscopy is the most powerful spectroscopic method for studying the structure and dynamics of proteins under physiologic conditions. However, obtaining structural information for high molecular weight systems above 30 kDa is technically challenging due to rapid signal decay that worsens with increasing molecular weights^{1,2}. Selective ¹H,¹³C-methyl labelling in a deuterated background has become the standard approach for mitigating rapid signal decay in solution NMR studies of high molecular weight proteins. However, this strategy suffers from a lack of information about non-methyl-containing amino acids. My thesis is focused on developing cost-effective isotope labelling strategies for introducing isolated, long-lived ¹H-¹²C magnetizations into proteins grown in deuterated *E. coli* media. Such magnetizations can be transferred through-space nuclear Overhauser enhancements (NOEs) to nearby ¹H,¹³C-methyl and ¹H,¹⁵N-amide groups in proteins, yielding rich new structural restraints for structure determination and additional spectral probes. While developing isotope labelling strategies for studying large protein systems, we also discovered a new method for analyzing protein side chain mobility. Protein side chain dynamics play a crucial role in many biologic processes, yet differentiating mobile from rigid side chains remains a technical challenge for solution NMR and other structural techniques. Our finding fills in a significant gap in protein structural biology.

Solution NMR in structural biology

Solution nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for elucidating protein structures and dynamics at atomic resolution.^{3,4} The advent of Fourier transform multi-dimensional NMR, combined with uniform isotopic labelling of proteins using NMR-active isotopes (^{15}N and ^{13}C), has allowed detailed structural and dynamic analyses of proteins with molecular weights up to 25 kDa.^{3,5} However, NMR studies of proteins over 30 kDa are complicated by rapid signal decay due to the slow tumbling of proteins, yielding spectra with poor resolution and low signal-to-noise^{1,2}. The fast signal decay (determined by the transverse relaxation rate) scales roughly with molecular weight (Figure 1.1).

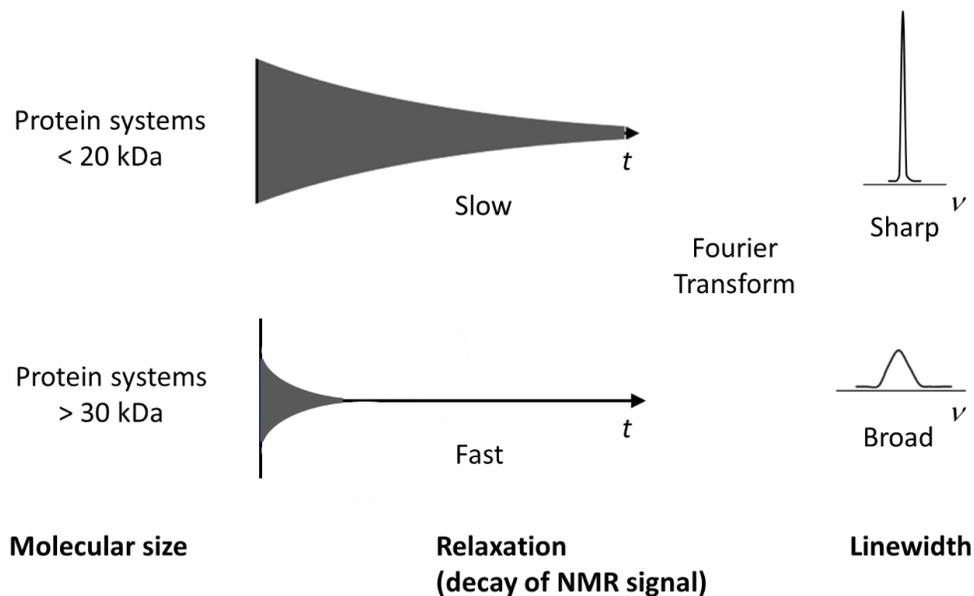


Figure 1.1 Effect of molecular size on signal decay in solution NMR. Small proteins (< 20 kDa) tumble rapidly, resulting in a slowly decaying NMR signal with a narrow linewidth. Large proteins (> 30 kDa) tumble slowly, yielding a fast-decaying NMR signal with a broad linewidth.

Technical developments in solution NMR, including high-field magnets equipped with cryogenic probes, transverse relaxation-optimized spectroscopy (TROSY) experiments, and recombinant protein expression with optimal isotope labelling, have extended the utility of solution NMR structural studies of proteins^{6,7}, allowing it to provide unique insights into folding, internal mobility, and binding of drugs, lipids, and ligands in native-like conditions.⁸⁻¹⁰

However, the role of solution NMR in structural biology must evolve considering other advancing techniques. X-ray crystallography has been the method of choice to determine protein structures at the highest resolution. Cryo-electron microscopy (cryo-EM) is becoming the dominant technique with current methodologies allowing high-resolution characterization of previously intractable biological systems like large multi-protein and integral membrane protein complexes. Computational methods can now accurately predict three-dimensional structures of proteins from amino acid sequences.^{11,12} Given the dominance of these techniques, one may wonder if solution NMR spectroscopy still has a relevant role in structural biology.

Solution NMR is an effective tool for determining the structures of small proteins and the dynamic movements of flexible protein regions, as well as for analyzing interactions between molecules. Additionally, it can be used to study proteins that are difficult to crystallize. Furthermore, NMR can be conducted on living cells or organisms, providing valuable insights into how proteins function in their natural environment. Since protein mobility plays a crucial role in biological activity, this can only be fully understood when studied at physiological temperatures, which can be easily accessed through solution NMR. However, the size of the protein is the main limitation when using solution NMR.

Solution NMR methods for tackling large protein systems

Deuteration with selective ^1H - ^{15}N protonation

Despite the utility of uniform ^{15}N and ^{13}C -labeling and triple resonance experiments, NMR studies of larger, fully protonated proteins are severely hampered by rapid signal decay. Relaxation at a given spin-1/2 nucleus (^1H , ^{15}N or ^{13}C in proteins) is caused by local fluctuating magnetic fields from other nearby nuclei (dipolar relaxation) or electrons (chemical shift anisotropy). Deuteration minimizes dipolar relaxation pathways by replacing ^1H nuclei in proteins with ^2H , which is still magnetic, but only about 15 times as effective as ^1H in inducing dipolar relaxation, providing sensitivity and resolution gains in most NMR experiments.¹³

Complete deuteration can be accomplished by expressing proteins in bacterial growth media containing D_2O and deuterated carbon source. The major problem with this approach is that most NMR pulse sequences start and end on ^1H nuclei. To overcome this limitation, proteins grown in D_2O are exchanged into H_2O , allowing re-protonation at all exchangeable H positions

(although transient unfolding of the protein may be needed to exchange buried parts). This selectively re-protonates all H atoms bound to nitrogen, allowing ^1H - ^1H nuclear Overhauser enhancement- (NOE-)derived distance restraints to be obtained for low-resolution structure calculation^{14,15}.

Transverse relaxation-optimized spectroscopy (TROSY) is a solution NMR strategy that exploits destructive interference between two interacting relaxation mechanisms.^{7,16} Dipolar relaxation is field invariant, whereas chemical shift anisotropy (CSA) scales with the square of the magnetic field strength. The destructive interference between dipolar coupling and CSA is serendipitously effective in both ^1H and ^{15}N atoms in backbone amide ^1H - ^{15}N groups at a high magnetic field (roughly 1 GHz).¹⁶ However, the effect is most significant for ^{15}N . In particular, pulse sequences that spend a long-time evolving transverse ^{15}N magnetization benefit most from TROSY, as occurs in multi-dimensional experiments in which magnetization is transferred from ^{15}N to ^{13}C via small 1- or 2-bond J-couplings (such as HNCA, HN(CO)CA, HN(CA)CB, HNCO, HN(CA)CO). Sensitivity gains from TROSY are most significant, with high levels of protein deuteration to attenuate long-range dipole-dipole interactions.^{16,17} Applying TROSY enables high-resolution NMR of large supramolecular complexes with molecular weights >30 kDa.^{4,17} 3D-TROSY versions of triple resonance experiments have allowed complete or near-complete backbone resonance assignment of several large proteins, establishing their global folds.^{14,18-21}

Like TROSY, Cross-correlated Relaxation-Induced Polarization Transfer (CRIPT), Cross-correlated Relaxation-Enhanced Polarization Transfer (CRINEPT), and Heteronuclear Multiple Quantum Coherence (HMQC) are NMR techniques that exploit the interference effect between dipolar relaxation and CSA to reduce transverse relaxation, extending the size limit of solution NMR to molecules with molecular weights beyond 100 kDa.²²⁻²⁴ Most other multi-dimensional NMR experiments rely on the transfer of magnetization from one nucleus to another via scalar spin-spin coupling using the insensitive nuclei enhanced by polarization transfer (INEPT) pulse sequence element.²⁵ The problem with INEPT elements is that a refocusing 180-degree pulse interconverts fast- and slow-relaxing magnetization, and for molecular weights above 100 kDa, transverse relaxation during these short ^1H - ^{15}N INEPT elements may become a limiting factor that compromises sensitivity. CRIPT, CRINEPT, and HMQC-based experiments

avoid extraneous magnetization-mixing pulses and thus better preserve slowly relaxing magnetization during magnetization transfers between ^1H and ^{15}N nuclei.^{6,22,26} However, it should be noted that in high molecular weight systems benefiting from non-INEPT transfers, signal decay tends to occur too quickly to obtain 3D or 4D NMR experiments needed for chemical shift assignment or structure determination.

Selective Methyl Group Protonation and Methyl-TROSY

For protein systems with a molecular weight above 100 kDa, obtaining NMR spectra of backbone amide groups becomes extremely difficult, even with deuteration and TROSY. Methyl groups give rise to a protein's most intense and longest-lived NMR signals. The protons in a methyl group are equivalent, contributing to the intensity of a single signal. Fast rotation around the methyl symmetry axis attenuates intra-methyl dipolar relaxation mechanisms by about one order of magnitude. Methyl groups in alanine, valine, leucine, isoleucine, threonine, and methionine are attractive targets for selective protonation since they populate hydrophobic protein cores, and they are essential for protein folding, stability, ligand binding, and interactions, making them valuable probes for obtaining structural and dynamic information.^{27,28} Protocols for methyl-selective protonation in these amino acids have been successfully developed.²⁹⁻³² Nuclear Overhauser enhancements (NOEs) can transfer magnetization between spatially close ^1HN - $^{13}\text{CH}_3$ and $^{13}\text{CH}_3$ - $^{13}\text{CH}_3$ groups to aid in the assignment process. These can also be combined with backbone ^1HN - ^1HN NOEs and are particularly important for structure determinations of supramolecular complexes.³³

Relaxation interference effects are optimized in methyl TROSY, which is fundamentally the same pulse sequence as HMQC.³⁴ Since methyl groups experience relaxation through a network of intra-methyl ^1H - ^1H and ^1H - ^{13}C dipolar interactions, methyl TROSY, whose effect is independent of magnetic field strength (unlike ^1H - ^{15}N TROSY), prevents mixing of slow- and fast-relaxing magnetization throughout the pulse sequence, resulting in spectra with improved signal-to-noise in a highly deuterated background.³⁴ Methyl TROSY experiments have provided structural, dynamic, and functional insights into protein systems up to 1 MDa in size (Figure 1.2).^{35,36} However, a significant challenge of selective methyl labelling in large protein systems is that the methyl groups are isolated in spin systems, making sequence-specific chemical shift

assignments difficult through NMR experiments. Extensive mutagenesis in combination with computational methods have been used to aid the assignment process.

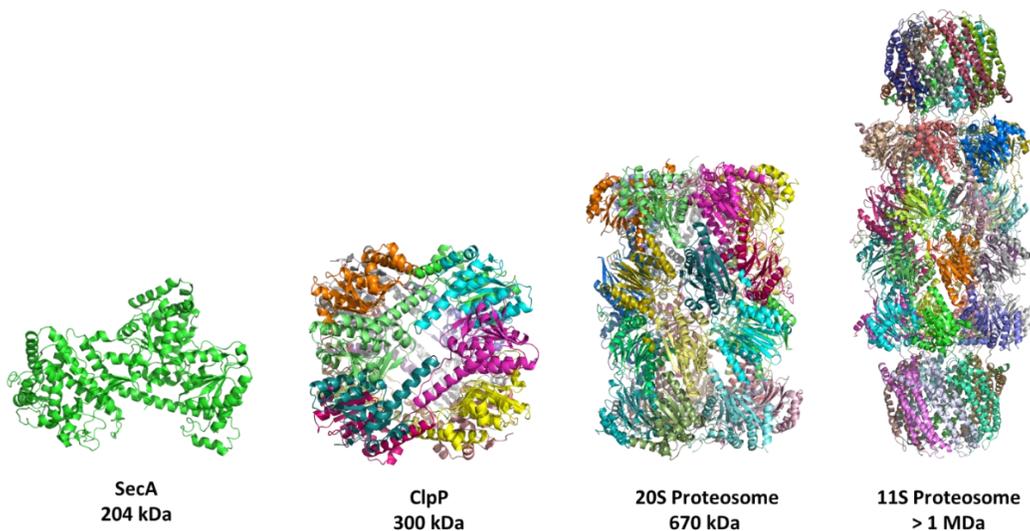


Figure 1.2. The range of biological assemblies above 100 kDa studied using the methyl-TROSY technique.

Strategies for selective methyl group protonation in perdeuterated proteins

Over the past 25 years, several strategies have been proposed for selective methyl group protonation in the six methyl-containing amino acids in proteins. These labelling approaches aim to produce deuterated proteins with selective $^{13}\text{CH}_3$ -labeling in methyl-containing amino acids. Such labelling schemes can be achieved by adding into deuterated *E. coli* growth media $^{13}\text{C}^1\text{H}_3$ -methyl-labeled amino acids or their metabolic precursors (Figure 1.3). Depending on the biosynthetic pathway of the desired methyl group, residues synthesized as late products of an irreversible metabolic reaction are the easiest to label (Figure 1.3).³⁷

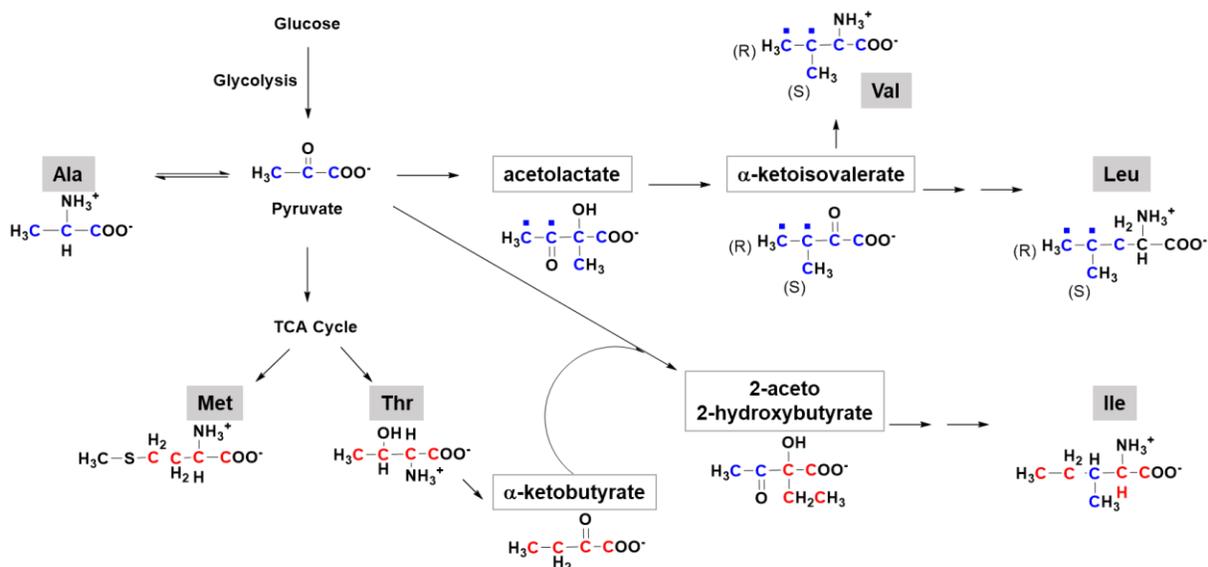


Figure 1.3. Simplified biosynthetic steps involved in the metabolism of methyl-containing amino acids. The carbon positions are colour coded according to their metabolic origin. Pyruvate-derived positions are shown in blue, whereas carbons derived from aspartate within the TCA cycle are highlighted in red. Amino acid residues are darkly shaded, while precursors frequently used in methyl labelling protocols are shown without fill. Two pyruvate molecules condense to form acetolactate, and blue dots represent carbon positions derived from one of the pyruvate molecules.

Isoleucine

Isoleucine contains two methyl groups: the γ_2 -methyl and the δ_1 -methyl, derived biosynthetically from pyruvate and α -ketobutyrate, respectively. *E. coli* synthesizes isoleucine by combining one molecule of pyruvate and one molecule of α -ketobutyrate (Figure 1.3). As a result, different labelling strategies have been developed targeting one or the other methyl group. Like leucine and valine, it is possible to supplement *E. coli* growth media with labelled isoleucine. However, this strategy has primarily been replaced by late α -ketoacid precursors of the final amino acid.^{29,38,39} The precursor-based methods are considerably cheaper, affording greater flexibility in the isotopic composition of the desired residue.

The earliest methyl-specific labelling protocol reported in literature involves the δ_1 -methyl group of isoleucine.⁴⁰ Gardner and Kay used the (3,3- $^2\text{H}_2$) ^{13}C 2-ketobutyrate precursor to label the δ_1 -methyl group of isoleucine without isotope scrambling. The precursor is derived

from ^{13}C -labelled threonine in two steps. Step one converts threonine into ($3\text{-}^2\text{H}$) α -ketobutyrate via the activity of threonine deaminase in D_2O . The second step converts ($3\text{-}^2\text{H}$) α -ketobutyrate into ($3,3\text{-}^2\text{H}_2$) α -ketobutyrate via a solvent exchange at position 3 in D_2O at elevated pH. The final product is incorporated into isoleucine by *E. coli* during protein synthesis (50-70 mg/L added 1 hour before induction).^{38,39} It is advantageous to supplement deuterated ^{13}C glucose to the *E. coli* medium to ensure complete ^{13}C labelling of the isoleucine side chain. Protonation at the Ile δ_1 methyl position is desirable since its signals occupy a unique region in the 2D ^{13}C - ^1H correlation spectra (Figure 1.4). Structurally, the isoleucine δ_1 -methyl group is typically about 3.7Å away from the protein backbone, facilitating the observation of NOEs between the methyl protons and amino acid side chains that are distant in the primary sequence. Since the first protocol, several alternative approaches have been developed for synthesizing α -ketobutyrate with various isotope labeling patterns commercially available from suppliers.^{41,42}

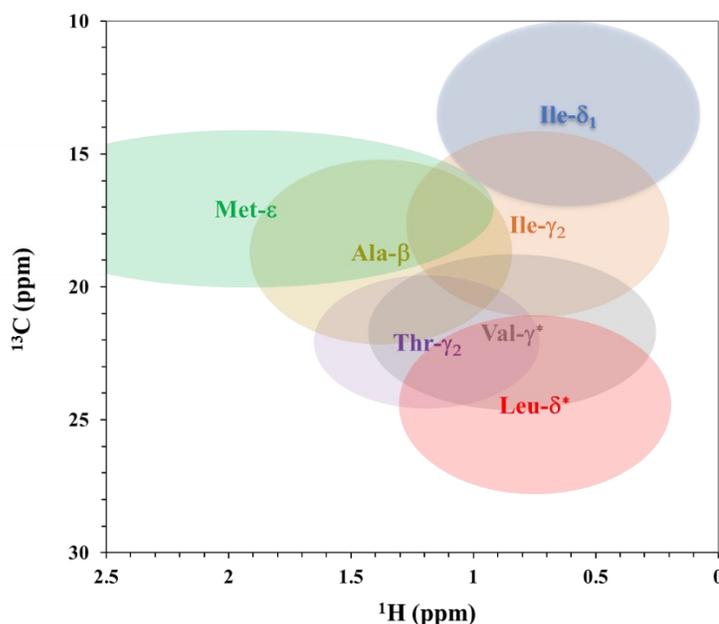


Figure 1.4. Expected chemical shifts of different methyl groups in 2D ^1H - ^{13}C correlation spectra. The coloured ellipses represent the average positions of methyl group signals. The average chemical shifts were taken from BioMagResBank (BMRB <https://www.bmrb.io/>). The boundaries of the ellipses are within two standard deviations of the chemical shifts.

The isoleucine γ_2 -methyl group is derived from pyruvate. Rosen *et al.* have shown the selective protonation of leucine, valine, and isoleucine- γ_2 methyl groups when *E. coli* is grown

on ^{13}C -pyruvate as the sole carbon source in D_2O .⁴³ However, isotope scrambling from the pyruvate methyl group dilutes signal intensities into numerous isotopomers (CH_3 , CH_2D , CHD_2 , and CD_3), compromising peak patterns and sensitivity. Ruschak and Ayala *et al.* have developed the specific labelling of the isoleucine- γ_2 methyl group using 2-hydroxy-2-ethyl-3-ketobutanoic acid (2-aceto-3-hydroxybutyrate) precursor.^{44,45} They showed that Ile- γ_2 methyl group could be selectively labelled (>95% efficiency) when 100 mg/L of the precursor is added to the *E. coli* medium before induction of protein synthesis in D_2O . The precursor is produced as a racemic mixture, but only the 2-(S)-aceto-2-hydroxybutyrate enantiomer is metabolized by *E. coli*. One drawback of this precursor is that some isotopic label is metabolized to the pro-(R) methyl group of leucine and valine, manifesting as weak peaks in 2D ^1H - ^{13}C correlation spectra. Ayala and coworkers showed that the isotope scrambling could be minimized by concurrently adding 200 mg/L of deuterated α -ketovalerate (leucine and valine precursor) to the growth medium.⁴⁵

Leucine and Valine

Leucine and valine are derived from two pyruvate molecules; the amino acids share a few common metabolic intermediates (Figure 1.3). The earliest protocol developed to label both methyl groups of leucine and valine using $[3\text{-}^2\text{H}]^{13}\text{C}$ α -ketoisovalerate [$^{13}\text{CH}_3$, $^{13}\text{CH}_3$] selectively labelled both proR and proS methyl groups.^{29,42,46} The use of this precursor to study large proteins proved inefficient because of intense intraresidue ^1H - ^1H dipolar interactions between prochiral methyl groups, limiting the sensitivity of NMR experiments and overcrowding in the 2D ^1H - ^{13}C correlation spectra. Subsequently, racemic $[3\text{-}^2\text{H}]^{13}\text{C}$ α -ketoisovalerate [$^{13}\text{CH}_3$, $^{12}\text{CD}_3$] precursors were used to label only one of the prochiral methyl groups in both amino acids.^{39,47} This precursor enhances the resolution and sensitivity of methyl TROSY spectra in large proteins. However, since it is synthesized as a racemic mixture, 50% of the prochiral methyl signal is detectable with both proR and proS sites visible by NMR at reduced sensitivity, preventing its use for the stereospecific assignment. Stereospecific labelling of leucine and valine was achieved using acetolactate (2-hydroxy-2- $[^{13}\text{C}]$ methyl-3-ketobutyrate)⁴⁸, a more upstream metabolite in Leu and Val biosynthesis (Figure 1.3). The precursor can be chemically synthesized to yield 100% proR or proS labelling of the methyl groups. Finally, protocols for labelling either leucine or valine have also been proposed. Kainosho's group showed synthetic

routes for labelling either of the prochiral methyl groups in leucine and valine,⁴⁹ while Lichtenecker used α -ketoisocaproate precursor to selectively label leucine residues.⁵⁰ Alternatively, supplementing acetolactate precursor with deuterated leucine selectively labels valine methyl groups.⁵¹

Methionine

The ϵ -methyl group of methionine is a valuable probe of protein structure and dynamics since it is uniquely positioned at the end of a long side chain unconnected to carbon but to a sulphur atom. These properties contribute to its signals resonating at an uncrowded region of the 2D ^1H - ^{13}C correlation spectrum (Figure 1.4). Isotope labelling of the methionine methyl group can be achieved by supplementing deuterated *E. coli* growth media with labelled methionine.⁵² Gelis and co-workers demonstrated the utility of this approach using 250 mg/L of [$^{13}\text{CH}_3$]-labelled methionine for a scrambling-free SecA (204 kDa) study.⁵² Fischer and colleagues developed an efficient synthetic protocol for producing 2-oxo-methionine precursor for labelling methionine residues in *E. coli*.³¹ It is relatively straightforward to react $^{13}\text{CH}_3\text{I}$ with homocysteine to produce methyl-labelled methionine, and this approach had been used long before protein perdeuteration was used for large proteins.^{53–55}

Alanine and threonine

Alanine is the smallest amino acid containing a methyl group and one of the most abundant amino acids in protein sequences. The β -methyl group of alanine is directly connected to the protein backbone serving as an excellent reporter of backbone structure and dynamics. Specific labelling of alanine or threonine is generally more difficult, as both residues are either synthesized in reversible reactions or are intermediates in the biosynthesis of other methyl-containing amino acids. Selective labelling of alanine- β and threonine- γ_2 methyl groups involves supplementing these residues directly into the *E. coli* minimal expression media.^{30,32} Ayala and colleagues added 800 mg/L of 2- ^{2}H , 3- $^{13}\text{CH}_3$ alanine for incorporation into perdeuterated proteins.³⁰ The reversible conversion of alanine to pyruvate via the enzymatic activity of alanine transaminase results in the supplemented alanine being scrambled to other pyruvate-derived methyl groups, i.e. isoleucine, leucine and valine. Co-addition of perdeuterated forms of α -

ketoisovalerate-d₇, succinate-d₄ and isoleucine-d₁₀ with labelled alanine reduces scrambling to <1%.

Of the six methyl-containing amino acids, threonine has one of the less hydrophobic side chains, residing within protein interiors or surfaces and participating in hydrogen bonding or signaling via phosphorylation. Velyvis showed an inexpensive biosynthetic strategy for producing L-[α -²H; β -²H; γ -¹³CH₃]-Thr, adding it directly to the *E. coli* expression medium for labelling the proteasome core particle (670 kDa).³² More recently, Ayala has shown an asymmetric synthetic route for producing methyl-specifically labelled threonine, demonstrating its application on the 82 kDa Malate Synthase G, adding to the limited number of threonine labelling strategies.⁵⁶

Combinatorial labelling of methyl-containing amino acids

The high level of perdeuteration required to produce proteins labelled explicitly at the methyl groups significantly reduces available NMR probes for structural studies compared to conventional uniform labelling. However, the six methyl-containing amino acids represent about 30-40% of residues found in proteins, 50% in protein hydrophobic cores, and about 30% in protein-protein interfaces.^{57,58} Specific labelling of the methyl-containing amino acids provides a good picture of protein structure and dynamics, maximizing the utility of the methyl NMR probe. As a result, several strategies have been developed for the combinatorial labelling of methyl groups, including ILV (Ile- δ_1 /Leu- δ^{proS} /Val- γ^{proS}), IMT (Ile- δ_1 /Met- ϵ /Thr- γ_2), AIM (Ala- β / Ile- δ_1 /Met- ϵ), AMLV (Ala- β /Met- ϵ / Leu- δ^{proS} /Val- γ^{proS}), AILV (Ala- β / Ile- δ_1 /Leu- δ^{proS} /Val- γ^{proS}), MILV (Met- ϵ /Ile- δ_1 /Leu- δ /Val- γ^{proS}), and MILVT (Met- ϵ /Ile- δ_1 /Leu- δ /Val- γ^{proS} / Thr- γ_2) etc. The combinatorial labelling strategy should ideally label more than one amino acid type simultaneously and target methyl cross peaks that do not overlap in the 2D ¹H-¹³C correlation spectrum. The availability of cost-effective metabolic precursors has made combinatorial labelling a robustly versatile and popular strategy.

Aromatic residue ¹H, ¹³C-site-selective labeling

A significant drawback of selective methyl group protonation is the inability to obtain structural information for non-methyl-containing amino acid side chains. Specifically, the side chains of aromatic amino acids are crucial in defining the hydrophobic packing and ligand

binding interfaces in proteins. They exhibit many long-range interproton NOEs important for determining protein folds.⁵⁹⁻⁶¹ However, the NMR analysis of aromatic ring signals using the conventional uniform labelling is complicated by severe signal overlap due to poor chemical shift dispersion, strong scalar couplings, and rapid signal decay.⁵⁹⁻⁶¹ Lichteneker and coworkers have developed a biosynthetic method for labelling aromatic amino acids in *E. coli* bacteria using α -ketoacid precursors.⁶²⁻⁶⁵ These α -ketoacid precursors, featuring an alternating [^1H - ^{13}C]-[^2H - ^{12}C]-[^1H - ^{13}C] ring patterns, incorporate an isolated ^1H - ^{13}C spin system in a deuterated background, simplifying chemical shift assignment of the ring signal and minimizing ^1H - ^1H and ^{13}C - ^{13}C scalar coupling (Figure 1.5). Phenylpyruvate and 4-hydroxy phenylpyruvate are the late metabolic precursors for labeling phenylalanine and tyrosine.⁶² Anthranilate, indole, and indolepyruvate precursors have each been used to label tryptophan.^{66,67} Histidine labeling is achieved using imidazole pyruvate.⁶⁵

Large signal enhancements can be obtained when the TROSY principle is applied to aromatic C-H groups.⁶⁸ Many NMR labs now combine methyl-specific protonation with the aromatic site-selective labelling to protein systems within 50-100 kDa. However, significant drawbacks of this labelling strategy are the cost associated with precursor synthesis and the rapid dipolar relaxation that still dominates any ^1H - ^{13}C spin system, preventing the routine application of the labelling strategy to large protein systems.

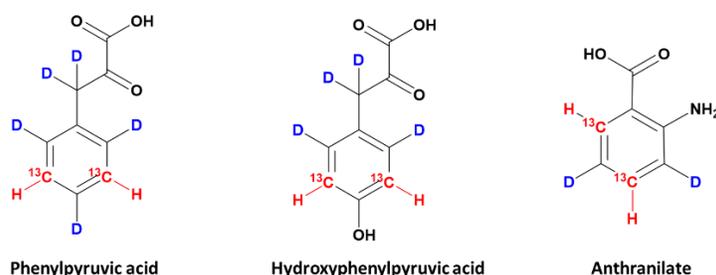


Figure 1.5. Overview of α -ketoacid precursors used for labelling aromatic amino acids.

Stereo-Array Isotope Labeling (SAIL)

Kainosho and coworkers have developed stereo array isotope labeling (SAIL), in which extensive chemical and enzymatic syntheses are employed to produce amino acids with optimal stereoselective ^2H , ^{13}C , and ^{15}N labeling.^{69,70} The method works well for structural and dynamic

studies of proteins with molecular weight <50 kDa. Due to the high cost of precursor synthesis, SAIL is always used with cell-free protein expression, allowing efficient use of the labelled precursors while preventing isotope label scrambling, which occurs in conventional bacterial protein expression systems. SAIL enhances spectral sensitivity and resolution at all amino acid residue positions, but the short-range ^1H - ^{13}C dipolar interaction still strongly dominates relaxation at non-methyl sites (Figure 1.6).

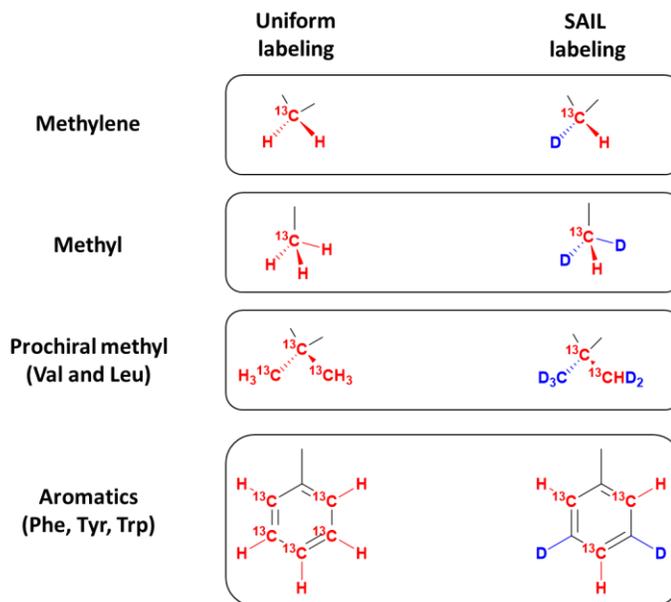


Figure 1.6. The design concept of SAIL amino acids.

Our approach for overcoming the size limitation of solution NMR involves economic stereoselective deuteration of non-methyl-containing amino acid side chains

Rapid signal decay in large proteins and inadequate side chain information are the two significant limitations of protein structure determination by solution NMR.⁷¹ Selective deuteration can overcome these problems, which slows signal decay and provides unambiguous stereospecific side chain chemical shift assignments. This thesis posits that only isolated ^1H - ^{12}C groups in a highly deuterated background possess sufficiently long-lived magnetization for non-methyl-containing amino acids in large proteins (Figure 1.7). Magnetization from ^1H - ^{12}C groups can be transferred via through-space NOEs to nearby ^1H - ^{15}N or ^1H - ^{13}C groups, providing the spectral resolution needed.

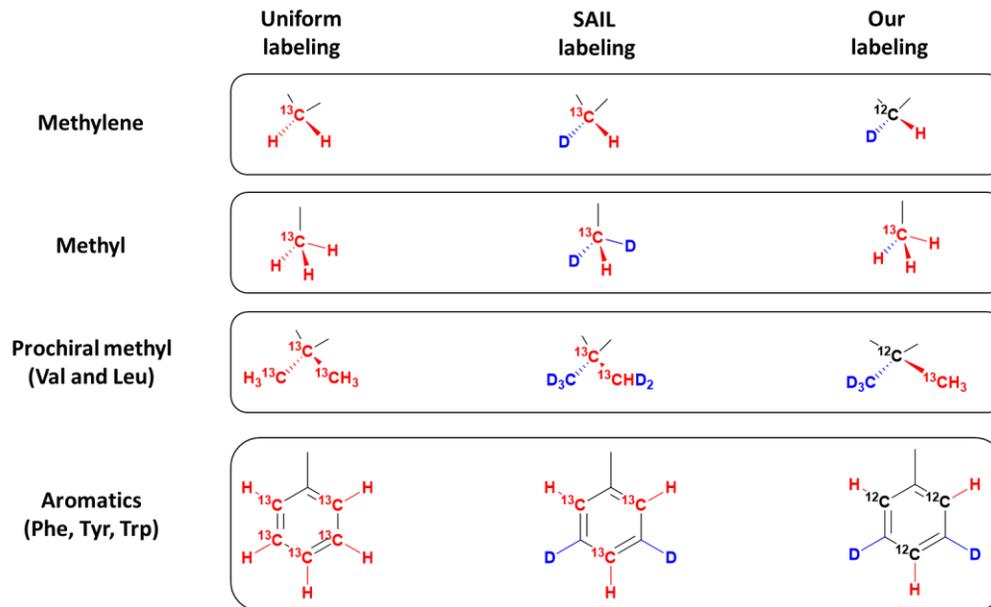


Figure 1.7. The design concept for stereoselective deuteration of non-methyl-containing amino acids

***Escherichia coli* bacteria as a model system for isotope labelling**

E. coli expression remains the most convenient means for producing proteins with complex isotope labelling patterns.^{72,73} The ease of genetic manipulation, rapid growth rate, and ability to overexpress large quantities of protein using simple, inexpensive growth media containing glucose, inorganic salts, and amino acid precursors make *E. coli* the first choice for protein expression.^{74–77} We have developed a model protein system, the thermostable mutant of the human Pin1 WW domain (the smallest folded protein domain containing all 20 amino acid residues except cysteine), for monitoring isotope incorporation into proteins expressed in *E. coli* bacteria. During my Ph.D., I assessed many different growth media for producing the desired stereospecific amino acid labelling pattern.

To demonstrate the utility of our approach, the second chapter of this thesis presents a method of introducing stereoselective ¹H-¹²C-labeling at the beta methylene proton (H β_2) of aspartate, asparagine, methionine, and lysine amino acid residues in an otherwise highly deuterated background by supplementing unlabeled sodium fumarate to the bacterial growth media.⁷¹ We used this labelling strategy for the unambiguous stereospecific assignment of the

beta methylene protons and showed its application for delineating side chain rotamers and dynamics in any protein.

The third chapter of the thesis establishes a new complementary method for labelling the aromatic rings of phenylalanine (H δ , H ζ), tyrosine (H δ), tryptophan (H δ , H η , H ϵ), and histidine (H δ) with isolated ^1H - ^{12}C groups using selectively deuterated phenylpyruvate, hydroxyphenylpyruvate, anthranilate, and unlabeled histidine as additives for *E. coli* D $_2\text{O}$ -based media. Our combined approaches improve the existing methyl labelling techniques, providing long-lived and highly selective ^1H - ^{12}C magnetization in seven non-methyl-containing amino acids in a highly deuterated background, facilitating high-resolution structure determination of large protein systems.

To demonstrate the utility of our approach, we applied the labelling strategy on a bacterial outer membrane enzyme PagP in dodecylphosphocholine DPC detergent micelles (>50 kDa), a protein our lab has extensive experience working on for a decade. PagP catalyzes the transfer of a palmitate chain from the sn-1 position of a phospholipid to the hydroxyl group of the N-linked R-3-hydroxymyristate chain on the proximal glucosamine unit of lipid A (Figure 1.8).^{14,78} The modification allows bacteria to evade host immune defences by providing resistance to antimicrobial peptides and attenuating the activation of lipopolysaccharide-mediated signalling through the Toll-like receptor 4 (TLR4) signal transduction pathway.^{79,80} PagP likely catalyzes reactions through a covalent intermediate using a variant of the classic Ser-His-Asp catalytic triad, an arrangement that should be druggable. This enzyme is a potential target for developing anti-infective agents and a tool for synthesizing lipid A-based vaccine adjuvants and endotoxin antagonists.⁸¹

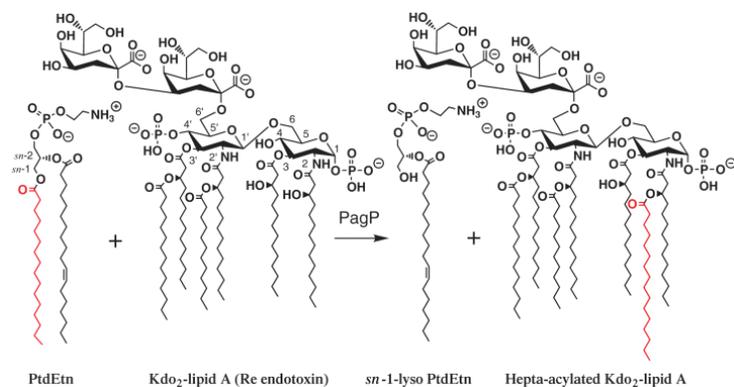


Figure 1.8. Reaction catalyzed by PagP. The acyltransferase PagP transfers a palmitate chain (red) from the *sn*-1 position of a phospholipid, such as phosphatidylethanolamine, to the free hydroxyl group at the 2-position of a distal glucosamine unit of the simplest form of hexa-acylated *E.coli* lipopolysaccharide (LPS), shown here as Kdo₂-lipid A, yielding a lyso-phospholipid and hepta-acylated LPS molecule.^{14,78}

Of the many enzymes involved in lipid A modification, PagP is the only known enzyme found in the outer membrane of *E. coli*. The solution structure of the EcPagP has been solved in DPC and n-octyl- β -D-glucoside (OG)/sodium dodecyl sulfate (SDS) detergents, revealing an 8-stranded antiparallel β -barrel preceded by an amino-terminal amphipathic α helix, with the active site residues located at the extracellular loops that are unstructured and dynamic (Figure. 1.9).¹⁴

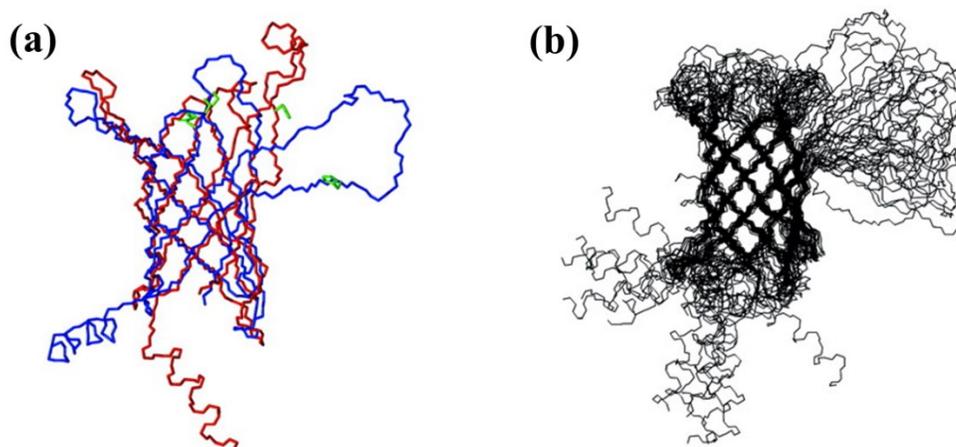


Figure 1.9. Solution structures of PagP in DPC and β -OG/SDS detergent micelles.¹⁴ (a) Lowest energy structure of PagP in DPC (blue) and β -OG/SDS (red) (b) 20 lowest energy structures of PagP in DPC. PagP is an eight stranded anti-parallel beta barrel preceded by an amino terminal amphipathic alpha helix. The core of the barrel is well defined, whereas the extracellular loops,

particularly the large loop1 is relatively disordered. Side chains of the residues important for catalysis, H33, D76 and S77, are shown in green.¹⁴

The X-ray crystal structure of PagP, solved in lauryl dimethylamine N-oxide (LDAO), showed an LDAO molecule buried in the barrel interior (Figure 1.10).⁸² This binding pocket in the core of the barrel serves as a hydrocarbon ruler that allows the enzyme to distinguish palmitate from other acyl chains found in phospholipids.⁸² The structural features of PagP pose interesting questions: 1) how can enzyme activity be supported in a protein where key catalytic residues are highly dynamic? 2) how do phospholipids which are not normally found in the outer leaflet of the outer membrane find their way to the active site of PagP. Despite the wealth of information about the structure and dynamics of PagP, little is known about its catalytic mechanism. The putative active site residues H33, D76 and S77 are not organized into a classic catalytic triad for catalysis.

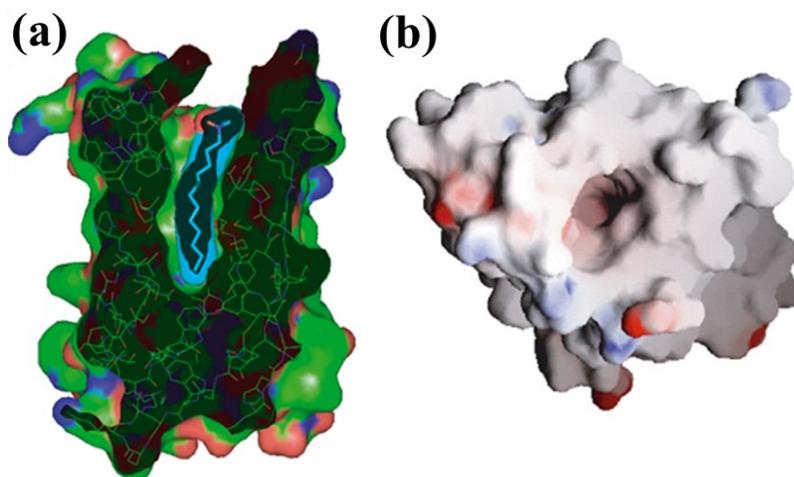


Figure 1.10. The X-ray crystal structure of PagP in LDAO detergent micelles shows an LDAO molecule occupying the substrate phospholipid binding site.⁸² (a) Cut off view of PagP showing LDAO detergent molecule buried in an interior binding pocket (b) Top view of PagP showing an empty binding pocket colored by electrostatic potential.⁸²

Like most membrane-bound enzymes of lipid metabolism, PagP requires a membrane-mimetic environment. However, only a select few detergents support enzymatic activity.⁸² The enzyme is active in DDM and CYFOS-7, but shows little or no activity in DPC or LDAO. LDAO and DPC have long hydrophobic tails, allowing them to inactivate PagP by occupying the

substrate's binding site, whereas DDM and CYFOS-7 have a bulky head group and a bulky cyclohexane tail, respectively, precluding them from binding the inner hydrophobic binding pocket, thus promoting enzymatic activity. The study of PagP dynamics in CYFOS-7, a detergent that promotes enzymatic activity, revealed that PagP exists in an equilibrium between two dynamically distinct states: a relaxed "R" form characterized by high degree of mobility and a tense "T" form characterized by more highly structured active site residues.⁸³

Attempts to crystallize and solve the solution structure of the tense form of PagP by conventional methods have been unsuccessful. A high-resolution structure by NMR will require the application of new isotope labeling methodologies. Selectively deuterated PagP structures could be solved in a bilayer, bicelles or nanodisc, which can help to resolve disordered loops, lipid interactions, or catalytic states. These systems can provide a more native-like environment compared to solutions in detergent, and thus, help to retain the structural and functional integrity of the protein, providing useful information about protein-lipid interactions and dynamics.

Dynamics by Solution NMR

Backbone dynamics

Proteins undergo various structural transitions at different timescales displaying physical motions ranging from bond vibrations, loop motions, and rotamer dynamics implicated in enzyme catalysis, ligand binding, conformational changes, protein folding, unfolding, etc. (Figure 1.11). NMR is the most versatile technique used to characterize protein dynamics. Among the numerous approaches available that provide access to internal motions over a wide range of timescale, the measurement of backbone ¹⁵N relaxation rates have proven to be the most popular for quantifying protein motion.^{84,85} The ¹⁵N relaxation experiments are based on a 2D ¹H-¹⁵N heteronuclear single quantum coherence (HSQC) experiment, which provides information about all ¹⁵N-¹H atom pairs in a protein. The 2D ¹⁵N HSQC shows backbone amide NH correlations for every amino acid type except proline in a uniformly labelled protein, allowing peak intensity changes to be monitored for site-specific information about the timescales and amplitudes of dynamic fluctuations. Backbone ¹⁵N relaxation provides information on internal motions faster than the overall tumbling motions (sub ns) and slower motions on the us-ms timescales that give rise to conformational exchange-induced relaxation.^{86,87} These investigations into the rapid sub-nanosecond motions most commonly

measure the longitudinal rate constant, R_1 , the transverse relaxation rate constant, R_2 , and the ^{15}N - $\{^1\text{H}\}$ nuclear Overhauser effect (NOE), from which the dipolar cross-relaxation rate between ^{15}N and its neighbouring proton can be extracted. These experiments usually indicate that the protein backbone is rigid in a folded domain except for long loops and tails.

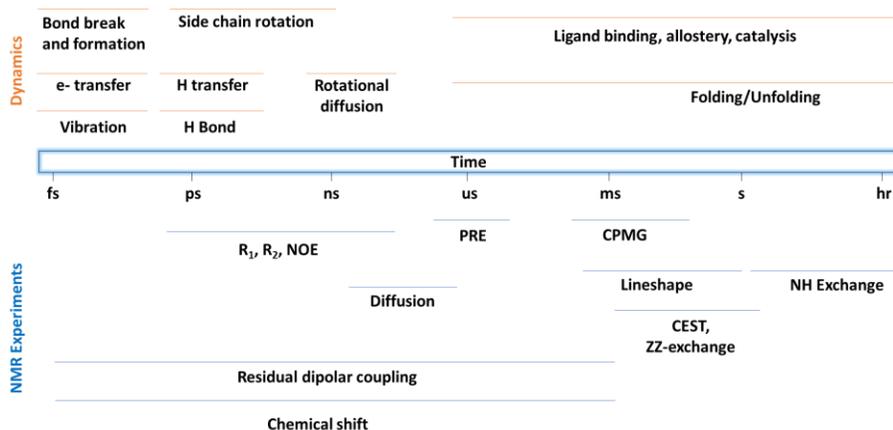


Figure 1.11. Accessible timescales of NMR experiments. The top portion corresponds to protein dynamics (adapted from).⁹

The longitudinal or spin-lattice relaxation R_1 is the process by which net magnetization returns to equilibrium. R_1 relaxation is commonly determined using the inversion recovery experiment (Figure 1.12a).^{85,88} This experiment applies a 180° pulse to tilt magnetization from the $+z$ to $-z$ axis, followed by a variable delay, which allows magnetization to relax before a 90° pulse flips magnetization from the z -axis to the transverse plane for detection. (Practically speaking, phase cycling to subtract out the equilibrium magnetization allows for a relaxation curve that decays to zero signal, requiring fewer fitting parameters). Insertion of an ^{15}N inversion recovery element into a 2D ^1H - ^{15}N HSQC sequence allows R_1 relaxation rates to be measured at most sites throughout the protein.

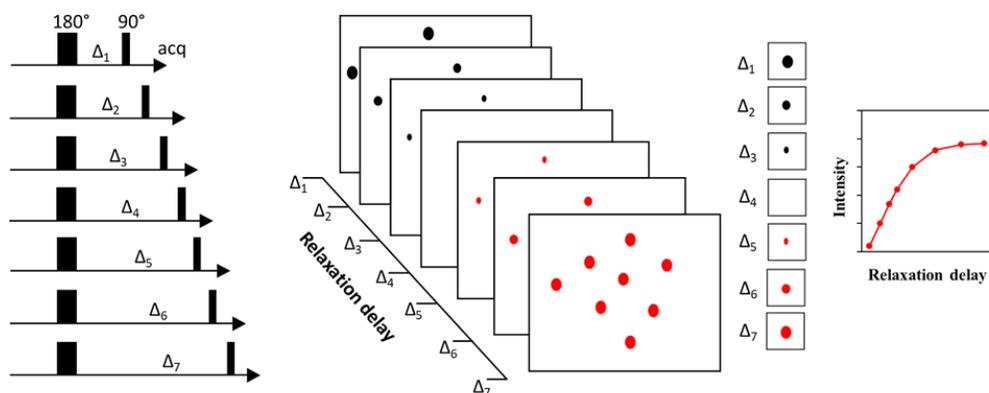
The transverse or spin-spin relaxation R_2 is the process by which the transverse component of magnetization decays or dephases. A 90° pulse tilts magnetization into the transverse plane, followed by a variable delay period, which is used to quantify R_2 relaxation rate. During this delay period the magnetization is locked into the transverse plane by a train of 180° pulses optimized to reduce the impact of pulse imperfections and off-resonance effects (Carr-Purcell-Meiboom-Gill or CPMG elements). An alternative is to measure transverse

relaxation in the rotating frame, i.e. $R_{1\rho}$.^{84,89} The $R_{1\rho}$ is determined by measuring magnetization decay under a spin-lock field, which generates a rotating magnetic field in the transverse plane perpendicular to the static magnetic field (Figure 1.12b).^{90,91} The $R_{1\rho}$ rate is then converted to an R_2 rate.

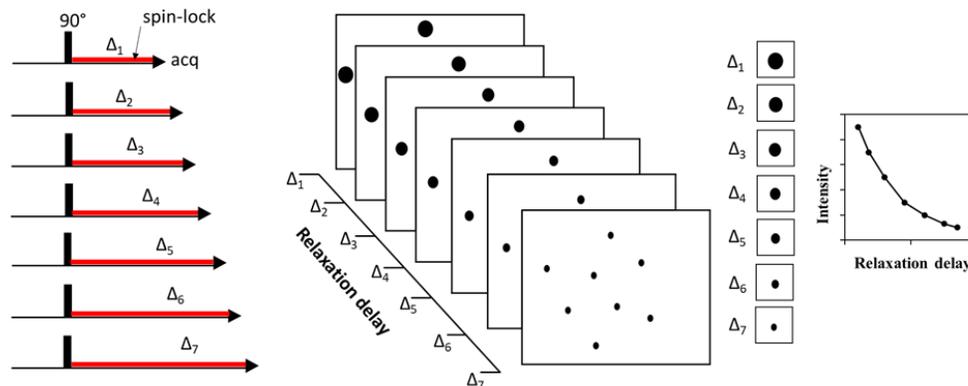
The heteronuclear 2D $^{15}\text{N}\{-^1\text{H}\}$ NOE experiment is another widely used NMR experiment for probing fast timescale (ps-ns) protein dynamics, allowing the quantitation of residue-specific fluctuations in proteins. The experiment differentiates unfolded and partially flexible regions from the protein-folded core. The heteronuclear NOE values are determined by measuring the NOE experiment (Figure 1.12c), which transfers magnetization through space via the dipolar coupling of ^1H to the attached ^{15}N .^{92,93} The experiment measures two different spectra with and without protein saturation, and the ratio of the intensities between the two experiments gives the NOE values for each residue. A significant decrease in the NOE intensity compared to the average NOE values in a protein corresponds to residues undergoing fast timescale motion.⁹³

The model-free analysis (Figure 1.12d) allows the extraction of dynamic information in a quantitative form from the R_1 , R_2 and the steady-state NOE relaxation data using the formalism initially developed by Lipari and Szabo.^{94,95}

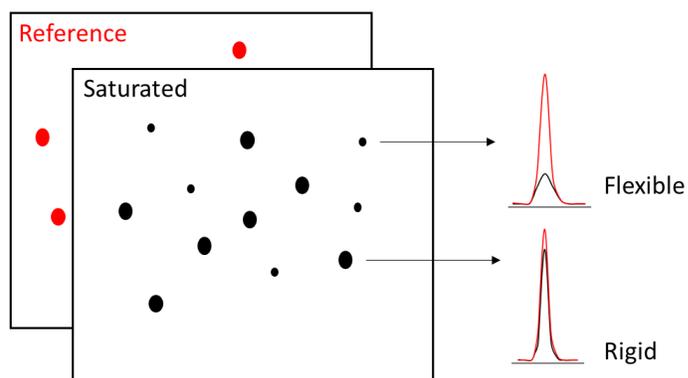
(a) Determination of R_1 rates



(b) Determination of $R_2/R_{1\rho}$ rates



(c) Determination of hetNOE values



(d) Model-free analysis

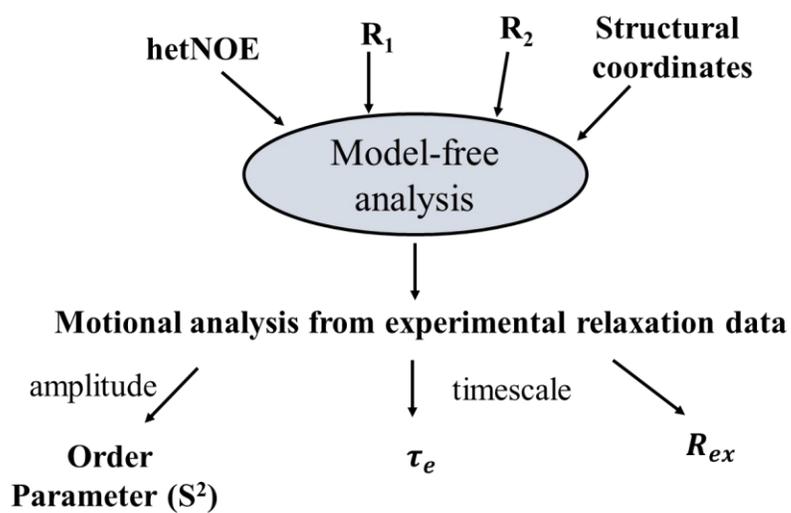


Figure 1.12. Description of NMR experiments for measuring backbone dynamics (a) the inversion recovery pulse scheme showing sample spectra at variable delay points and a chart of magnetization buildup back to equilibrium (b) $R_{1\rho}$ pulse scheme showing representative spectra at variable delays and a chart of magnetization decaying in the transverse plane (c) schematic diagram interpreting the heteronuclear NOE experiment (d) Motional analysis using the Model-free approach.

The R_1 and R_2 relaxation rates and the NOE enhancements of an amide ^{15}N spin are dominated by the dipolar interaction of the ^{15}N spin with its attached ^1H and by CSA as described by:^{88,96}

$$R_1 = \frac{1}{T_1} = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N)$$

$$R_2 = \frac{1}{T_2} = \left(\frac{d^2}{2}\right)[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] \\ + \left(\frac{c^2}{6}\right)[4J(0) + 3J(\omega_N)]$$

$$NOE = 1 + \left[\frac{\gamma_H}{\gamma_N} d^2 \{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\} T_1 \right]$$

The constants d^2 and c^2 are defined as

$$d^2 = 0.1\gamma_H^2\gamma_N^2h^2/(4\pi^2) \langle 1/r_{HN}^3 \rangle^2$$

$$c^2 = \left(\frac{2}{15}\right)\gamma_N^2B_0^2(\sigma_{\parallel} - \sigma_{\perp})^2$$

where R_1 is the longitudinal or the spin-lattice relaxation rate constant; R_2 is the transverse or spin-spin relaxation rate constant; γ_H and γ_N are the gyromagnetic ratios of the ^1H and ^{15}N nuclei; ω_H and ω_N are the ^1H and ^{15}N Larmor frequencies; h is the Planck's constant; r_{HN} is the internuclear ^1H - ^{15}N distance (1.02 Å); $J(\omega_i)$ is the spectral density at frequency ω_i ;

B_0 is the static magnetic field strength; σ_{\parallel} and σ_{\perp} are the parallel and perpendicular components of the assumed axially symmetric ^{15}N chemical shift tensor, which for peptide bonds has a value of about $(\sigma_{\parallel} - \sigma_{\perp}) = -160\text{ppm}$.⁹⁷

For a protein in solution, the spectral density function $J(\omega)$, depends on the overall motion of the macromolecule and the internal motions of the ^1H - ^{15}N bond vector. The model-free formalism uses a minimum number of motional parameters to describe $J(\omega)$ as follows:^{94,95}

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2)$$

Where S^2 is the order parameter which describes the degree of spatial restriction of the internal motion of the ^1H - ^{15}N bond vector; τ_m is the correlation time resulting from the tumbling motion of the entire molecule. The effective correlation time resulting from internal motions is described by τ_e , where $1/\tau = 1/\tau_m + 1/\tau_e$. This model assumes that the overall tumbling of the molecule is isotropic. The transverse relaxation rate constant may contain additional contributions from the chemical exchange process that arise from the μs - ms timescales, R_{ex} . As a result, an extended model-free spectral density function has been developed describing internal motions occurring in the fast and slow timescales.^{98,99} If the term containing the correlation time describing the faster of the two motions contributes a negligible amount to relaxation, the spectral density function becomes:⁸⁵

$$J(\omega) = S^2\tau_m/(1 + \omega^2\tau_m^2) + (S_f^2 - S^2)\tau/(1 + \omega^2\tau^2)$$

where the order parameter S^2 is expressed as the product of two order parameters representing the fast (S_f^2 , $\tau_f < 10\text{ ps}$) and slow (S_s^2 , $\tau_s < \tau_f < \tau_m$) motions, i.e. $S^2 = S_f^2 S_s^2$. The effective correlation time for the slow internal motions, τ_s , is included using the following relationship: $1/\tau = 1/\tau_s + 1/\tau_m$.

The R_{ex} term accounts for contributions to transverse relaxation from processes other than dipole-dipole and CSA. The equation describing R_2 relaxation is modified to account for the conformational exchange averaging processes as follows:

$$R_2 = R_{2(DD)} + R_{2(CSA)} + R_{ex}$$

The DD and CSA subscripts denote dipole-dipole and chemical shift anisotropy contributions to transverse relaxation.

Side chain dynamics

Protein side chain dynamics play a vital role in many biological processes. Unlike the generally rigid backbone, side chains can be quite flexible, with most residues in a protein accessing multiple rotameric states defined by the χ_1 dihedral angle (Figure 1.13).^{100,101} A typical 120° transition effectively remodels the protein surface. An understanding of protein side-chain motion is thus of greater interest. Yet, differentiating mobile from rigid side chains remains a technical challenge for solution NMR studies.

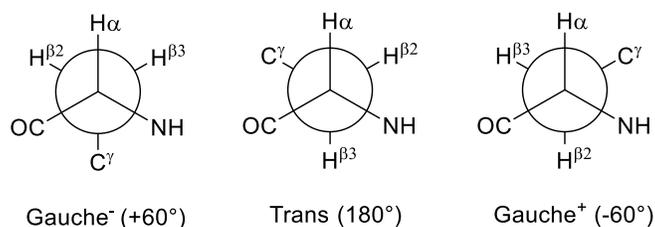


Figure 1.13. The three major χ_1 dihedral angle rotamers as viewed along the C_α-C_β bond.

Comprehensive studies of side chain relaxation are rare because of experimental challenges. In principle, ¹³C transverse relaxation time should offer more information on dynamics since carbon nuclei populate the protein backbone and side chain. In practice, however, the application of ¹³C relaxation has been greatly limited due to ¹³C-¹³C and ¹H-¹H J coupling and cross-correlated relaxation between ¹³C-¹H dipoles, which is not the case for backbone amides, which are considered as isolated two-spin systems.^{102,103} Earlier attempts to study side chain relaxation used ¹²C/¹³C and ¹H/²H fractional labelling to produce isolated ¹H-¹³C groups that behave similarly to ¹H-¹⁵N.¹⁰³⁻¹⁰⁶ However, these labelling methods are localized predominantly to a protein segment, providing an incomplete picture of overall dynamics.

Some methods have been developed to probe side-chain dynamics in a uniformly ¹³C-labeled, or ¹³C-, ¹⁵N-labeled protein involving pulse schemes for measuring cross-correlated relaxation between two C-H dipoles in a methylene or methyl group.^{102,107-109} The group of Lewis Kay developed a modified CBCACONH pulse scheme for measuring cross-correlated relaxation rates at the C_β positions in proteins.¹⁰⁷ Extending dynamic characterization beyond

C β , the group of Daiwen Yang further extended this methodology using a CC(CO)NH TOCSY-based 3D pulse sequence for measuring ^1H - ^{13}C dipole-dipole cross-relaxation at all CH $_2$ positions in proteins less than 15kDa in size. These methods extract cross-correlation rates based on deviations from the expected 1:2:1 ratio in the ^{13}C CH $_2$ triplet intensities, differentiating mobile from rigid CH $_2$ groups.¹⁰⁸ The group extended this scheme to methyl groups, relying on differences in peak intensities of the ^{13}C quartet component of methyl groups to extract cross-correlated relaxation rates to characterize methyl group dynamics. Moreover, differences in the multiplet components also allow the estimation of the magnitude of the ^{13}C nucleus CSA tensor.¹⁰⁹

Additional methods focus on ^2H relaxation, which is an excellent reporter of protein dynamics, with relaxation dominated by the strong quadrupolar interaction that is much larger than other relaxation effects.^{105,110} Yang and Kay developed an experiment for measuring side-chain ^2H T $_1$ and T $_{1\rho}$ relaxation rates using ^{15}N -, ^{13}C -labeled proteins in 50% D $_2\text{O}$. Following that experiment, Kay and coworkers further developed new pulse schemes for measuring deuterium relaxation in ^{13}C CH $_2\text{D}$ methyl groups of ^{15}N -, ^{13}C -labeled perdeuterated proteins, allowing five relaxation rates to be measured per deuteron.¹⁰⁵ Since then, measurement of ^2H relaxation rates has been mainly applied to study methyl group dynamics.^{105,110} Application to methine and methylenes has been limited to protein systems <15 kDa due to poor chemical shift dispersion and signal losses from larger proteins.

Because of spectral overlap in 2D ^1H - ^{13}C spectra, rapid signal losses due to long magnetization transfer steps, and rapid relaxation induced by dipolar interactions, side chain methods never achieved the popularity of ^{15}N relaxation methods. As a result, additional NMR methods are needed to measure side-chain dynamics in non-methyl-containing amino acids.

Although ^1H is the most abundant element in biomolecules, measuring ^1H relaxation is complicated by J couplings, cross-relaxation, and cross-correlated relaxation. We have developed a novel method for measuring ^1H relaxation rates in proteins. Chapter four of this thesis presents a simple NMR relaxation experiment that uses a radiofrequency field as a DIPSI2 mixing scheme at the beginning of any multi-dimensional pulse sequence to measure relaxation rates at all proton sites in proteins. The pulse scheme compels all ^1H atoms to behave similarly

with respect to strong coupling and maximum rotating frame ^1H - ^1H cross-relaxation. We applied the experiment to the human Pin1 WW domain, allowing for the measurement of relaxation rates at all protein sites. The fastest ^1H relaxation rates observed belong to rigid Pin1 residues, especially those found buried in the folding core. Flexible side chains undergoing rapid rotameric transitions at the χ_1 dihedral angle are associated with slow ^1H relaxation rates.

Moreover, the ^1H relaxation rates correlate well with two independent measures of side-chain dynamics: the cross-correlated relaxation rates in $^{13}\text{C}^1\text{H}_2$ methylene groups and ^3J couplings sensitive to the χ_1 dihedral angle ($^3\text{J}_{\text{H}\alpha,\text{H}\beta}$, $^3\text{J}_{\text{N},\text{H}\beta}$, and $^3\text{J}_{\text{CO},\text{H}\beta}$). These correlations indicate that ^1H relaxation rates can be a reliable indicator of side-chain motion. This simple methodology will allow side chain dynamics to be characterized with the same ease as backbone ^{15}N relaxation.

Summary

In this introductory chapter, I discussed some practical challenges of solution NMR in proteins. Even though X-ray crystallography and electron microscopy are the preferred methods to determine protein structures and their complexes, solution NMR is the ideal technique for monitoring the functional dynamics of proteins at various timescales, especially for systems that are difficult to crystallize or too small for electron microscopy. Solution NMR has primarily been considered a technique ideally suited for characterizing proteins smaller than 20 kDa. I have shown that NMR has made significant progress toward studying the structure and dynamics of large proteins, owing to current developments in protein production, sample preparation, pulse sequences, high-field magnets, and advances in isotope labelling strategies. The selective methyl group protonation, aromatic residue labelling, and SAIL labelling approach incorporate protons at specific protein sites, pushing the size limit of solution NMR to previously intractable biological systems. Despite these advances, structural determination of proteins by solution NMR remains challenging. Emerging techniques for labelling proteins cost-effectively and measuring side chain dynamics will undoubtedly bring additional advances to protein structural biology unique to solution NMR.

References

- (1) Yu, H. Extending the Size Limit of Protein Nuclear Magnetic Resonance. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (2), 332–334. <https://doi.org/10.1073/pnas.96.2.332>.
- (2) Foster, M. P.; McElroy, C. A.; Amero, C. D. Solution NMR of Large Molecules and Assemblies. *Biochemistry* **2007**, *46* (2), 331–340. <https://doi.org/10.1021/bi0621314>.
- (3) Oschkinat, H.; Griesinger, C.; Kraulis, P. J.; Sørensen, O. W.; Ernst, R. R.; Gronenborn, A. M.; Clore, G. M. Three-Dimensional NMR Spectroscopy of a Protein in Solution. *Nature* **1988**, *332* (6162), 374–376. <https://doi.org/10.1038/332374a0>.
- (4) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. NMR Analysis of a 900K GroEL-GroES Complex. *Nature* **2002**, *418* (6894), 207–211. <https://doi.org/10.1038/nature00860>.
- (5) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.* **1990**, *89* (3), 496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-5](https://doi.org/10.1016/0022-2364(90)90333-5).
- (6) Riek, R.; Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. Solution NMR Techniques for Large Molecular and Supramolecular Structures. *J. Am. Chem. Soc.* **2002**, *124* (41), 12144–12153. <https://doi.org/10.1021/ja026763z>.
- (7) Fernández, C.; Wider, G. TROSY in NMR Studies of the Structure and Function of Large Biological Macromolecules. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 570–580. <https://doi.org/10.1016/j.sbi.2003.09.009>.
- (8) Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein Dynamics and Function from Solution State NMR Spectroscopy. *Q. Rev. Biophys.* **2016**, *49*, e6. <https://doi.org/10.1017/S0033583516000019>.
- (9) Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **2004**, *104* (8), 3623–3640. <https://doi.org/10.1021/cr030413t>.
- (10) Bibow, S.; Hiller, S. A Guide to Quantifying Membrane Protein Dynamics in Lipids and Other Native-like Environments by Solution-State NMR Spectroscopy. *FEBS J.* **2019**, *286* (9), 1610–1623. <https://doi.org/10.1111/febs.14639>.
- (11) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.
- (12) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (13) Sattler, M.; Fesik, S. W. Use of Deuterium Labeling in NMR: Overcoming a Sizeable Problem. *Structure* **1996**, *4* (11), 1245–1249. [https://doi.org/10.1016/s0969-2126\(96\)00133-5](https://doi.org/10.1016/s0969-2126(96)00133-5).
- (14) Hwang, P. M.; Choy, W.; Lo, E. I.; Chen, L.; Forman-kay, J. D.; Raetz, C. R. H.; Bishop,

- R. E.; Kay, L. E.; Prive, G. G. Solution Structure and Dynamics of the Outer Membrane Enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U. S. A* **2002**, *99*, 13460–13565. <https://doi.org/10.1073/pnas.212344499>.
- (15) Fernández, C.; Hilty, C.; Wider, G.; Güntert, P.; Wüthrich, K. NMR Structure of the Integral Membrane Protein OmpX. *J. Mol. Biol.* **2004**, *336* (5), 1211–1221. <https://doi.org/10.1016/j.jmb.2003.09.014>.
- (16) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, 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. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (23), 12366–12371. <https://doi.org/10.1073/pnas.94.23.12366>.
- (17) Pervushin, K. Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a Technique in Structural Biology. *Q. Rev. Biophys.* **2000**, *33* (2), 161–197. <https://doi.org/10.1017/S0033583500003619>.
- (18) Fernández, C.; Adeishvili, K.; Wüthrich, K. Transverse Relaxation-Optimized NMR Spectroscopy with the Outer Membrane Protein OmpX in Dihexanoyl Phosphatidylcholine Micelles. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2358–2363. <https://doi.org/10.1073/pnas.051629298>.
- (19) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. Structure of Outer Membrane Protein A Transmembrane Domain by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8* (4), 334–338. <https://doi.org/10.1038/86214>.
- (20) Renault, M.; Saurel, O.; Czaplicki, J.; Demange, P.; Gervais, V.; Löhr, F.; Réat, V.; Piotto, M.; Milon, A. Solution State NMR Structure and Dynamics of KpOmpA, a 210 Residue Transmembrane Domain Possessing a High Potential for Immunological Applications. *J. Mol. Biol.* **2009**, *385* (1), 117–130. <https://doi.org/10.1016/j.jmb.2008.10.021>.
- (21) Bayrhuber, M.; Meins, T.; Habeck, M.; Becker, S.; Giller, K.; Villinger, S.; Vonrhein, C.; Griesinger, C.; Zweckstetter, M.; Zeth, K. Structure of the Human Voltage-Dependent Anion Channel. *Proc. Natl. Acad. Sci.* **2008**, *105* (40), 15370–15375. <https://doi.org/10.1073/pnas.0808115105>.
- (22) Riek, R.; Wider, G.; Pervushin, K.; Wüthrich, K. W. Polarization Transfer by Cross-Correlated Relaxation in Solution NMR with Very Large Molecules. *Proc. Natl. Acad. Sci. U. S. A* **1999**, *96*, 4918–4923. <https://doi.org/10.1073/pnas.96.9.4918>.
- (23) Brüschweiler, R.; Ernst, R. R. Molecular Dynamics Monitored by Cross-Correlated Cross Relaxation of Spins Quantized along Orthogonal Axes. *J. Chem. Phys* **1992**, *96*, 1758. <https://doi.org/10.1063/1.462131>.
- (24) Dalvit, C. ¹H to ¹⁵N Polarization Transfer via ¹H Chemical-Shift Anisotropy-¹H-¹⁵N Dipole-Dipole Cross Correlation. *J. Magn. Reson.* **1992**, *97*, 645–650. [https://doi.org/10.1016/0022-2364\(92\)90046-A](https://doi.org/10.1016/0022-2364(92)90046-A).
- (25) Morris, G. A.; Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **1979**, *101* (3), 760–762. <https://doi.org/10.1021/ja00497a058>.
- (26) Wider, G. NMR Techniques Used with Very Large Biological Macromolecules in Solution. *Methods Enzymol.* **2005**, *394* (1985), 382–398. [https://doi.org/10.1016/S0076-6879\(05\)94015-9](https://doi.org/10.1016/S0076-6879(05)94015-9).

- (27) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* **1996**, *263*, 627–636. <https://doi.org/10.1006/jmbi.1996.0603>.
- (28) Tugarinov, V.; Kay, L. E. Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins. *ChemBioChem* **2005**, *6* (9), 1567–1577. <https://doi.org/10.1002/cbic.200500110>.
- (29) 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 (δ 1). *J. Biomol. NMR* **1999**, No. 13, 369–374. <https://doi.org/10.1023/a:1008393201236>.
- (30) Ayala, I.; Sounier, R.; Usé, N.; Gans, P.; Boisbouvier, J. An Efficient Protocol for the Complete Incorporation of Methyl-Protonated Alanine in Perdeuterated Protein. *J. Biomol. NMR* **2009**, *43* (2), 111–119. <https://doi.org/10.1007/s10858-008-9294-7>.
- (31) Fischer, M.; Kloiber, K.; Häusler, J.; Ledolter, K.; Konrat, R.; Schmid, W. Synthesis of a ¹³C-Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. *ChemBioChem* **2007**, *8* (6), 610–612. <https://doi.org/10.1002/cbic.200600551>.
- (32) Velyvis, A.; Ruschak, A. M.; Kay, L. E. An Economical Method for Production of 2H,¹³CH₃-Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 KDa Proteasome. *PLoS One* **2012**, *7* (9), 1–8. <https://doi.org/10.1371/journal.pone.0043725>.
- (33) Sprangers, R.; Velyvis, A.; Kay, L. E. Solution NMR of Supramolecular Complexes: Providing New Insights into Function. *Nat. Methods* **2007**, *4* (9), 697–703. <https://doi.org/10.1038/nmeth1080>.
- (34) Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: Explanation and Experimental Verification. *Magn. Reson. Chem.* **2003**, *41* (10), 843–852. <https://doi.org/10.1002/mrc.1256>.
- (35) Wiesner, S.; Sprangers, R. Methyl Groups as NMR Probes for Biomolecular Interactions. *Curr. Opin. Struct. Biol.* **2015**, *35*, 60–67. <https://doi.org/10.1016/j.sbi.2015.08.010>.
- (36) Kay, L. E. Solution NMR Spectroscopy of Supra-Molecular Systems, Why Bother? A Methyl-TROSY View. *J. Magn. Reson.* **2011**, *210* (2), 159–170. <https://doi.org/10.1016/j.jmr.2011.03.008>.
- (37) Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. Methyl-Specific Isotopic Labeling: A Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122. <https://doi.org/10.1016/j.sbi.2015.03.009>.
- (38) Gardner, K. H.; Kay, L. E. Production and Incorporation of ¹⁵N, ¹³C, 2H (1H- Δ 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **1997**, *119* (32), 7599–7600.
- (39) Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, *1* (2), 749–754. <https://doi.org/10.1038/nprot.2006.101>.
- (40) Gardner, K. H.; Kay, L. E. Production and Incorporation of ¹⁵N, ¹³C, 2H (1H- Δ 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **1997**, *119* (32), 7599–7600. https://doi.org/10.1021/JA9706514/SUPPL_FILE/JA7599.PDF.
- (41) Lichtenecker, R.; Ludwiczek, M. L.; Schmid, W.; Konrat, R. Simplification of Protein

- NOESY Spectra Using Bioorganic Precursor Synthesis and NMR Spectral Editing. *J. Am. Chem. Soc.* **2004**, *126* (17), 5348–5349. <https://doi.org/10.1021/ja049679n>.
- (42) Gross, J. D.; Gelev, V. M.; Wagner, G. A Sensitive and Robust Method for Obtaining Intermolecular NOEs between Side Chains in Large Protein Complexes. *J. Biomol. NMR* **2003**, *25* (3), 235–242. <https://doi.org/10.1023/A:1022890112109>.
- (43) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* **1996**, *263* (5), 627–636. <https://doi.org/10.1006/jmbi.1996.0603>.
- (44) Ruschak, A. M.; Velyvis, A.; Kay, L. E. A Simple Strategy for ¹³C,¹H Labeling at the Ile- Γ 2 Methyl Position in Highly Deuterated Proteins. *J. Biomol. NMR* **2010**, *48* (3), 129–135. <https://doi.org/10.1007/s10858-010-9449-1>.
- (45) 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.* **2012**, *48* (10), 1434–1436. <https://doi.org/10.1039/c1cc12932e>.
- (46) Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W. NMR-Based Screening of Proteins Containing ¹³C-Labeled Methyl Groups. *J. Am. Chem. Soc.* **2000**, *122* (33), 7898–7904. <https://doi.org/10.1021/JA000350L>.
- (47) Tugarinov, V.; Kay, L. E. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *J. Biomol. NMR* **2004**, *28* (2), 165–172. <https://doi.org/10.1023/B:JNMR.0000013824.93994.1f>.
- (48) 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. *Angew. Chemie - Int. Ed.* **2010**, *49* (11), 1958–1962. <https://doi.org/10.1002/anie.200905660>.
- (49) Miyanoiri, Y.; Takeda, M.; Okuma, K.; Ono, A. M.; Terauchi, T.; Kainosho, M. Differential Isotope-Labeling for Leu and Val Residues in a Protein by E. Coli Cellular Expression Using Stereo-Specifically Methyl Labeled Amino Acids. *J. Biomol. NMR* **2013**, *57* (3), 237–249. <https://doi.org/10.1007/s10858-013-9784-0>.
- (50) Lichtenecker, R. J.; Coudevylle, N.; Konrat, R.; Schmid, W. Selective Isotope Labelling of Leucine Residues by Using α -Ketoacid Precursor Compounds. *ChemBioChem* **2013**, *14* (7), 818–821. <https://doi.org/10.1002/cbic.201200737>.
- (51) Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J. Specific Labeling and Assignment Strategies of Valine Methyl Groups for NMR Studies of High Molecular Weight Proteins. *J. Biomol. NMR* **2013**, *57* (3), 251–262. <https://doi.org/10.1007/s10858-013-9785-z>.
- (52) Gelis, I.; Bonvin, A. M. J. J.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **2007**, *131* (4), 756–769. <https://doi.org/10.1016/j.cell.2007.09.039>.
- (53) MELVILLE, D. B.; RACHELE, J. R.; KELLER, E. B. A Synthesis of Methionine Containing Radiocarbon in the Methyl Group. *J. Biol. Chem.* **1947**, *169* (2), 419–426. [https://doi.org/10.1016/s0021-9258\(17\)35042-1](https://doi.org/10.1016/s0021-9258(17)35042-1).
- (54) Gifford, J. L.; Ishida, H.; Vogel, H. J. Fast Methionine-Based Solution Structure

- Determination of Calcium-Calmodulin Complexes. *J. Biomol. NMR* **2011**, *50* (1), 71–81. <https://doi.org/10.1007/s10858-011-9495-3>.
- (55) Weininger, U.; Liu, Z.; McIntyre, D. D.; Vogel, H. J.; Akke, M. Specific $^{12}\text{C}\beta\text{D}2$ $^{12}\text{C}\gamma\text{D}2$ $^{13}\text{C}\epsilon\text{HD}2$ Isotopomer Labeling of Methionine To Characterize Protein Dynamics by ^1H and ^{13}C NMR Relaxation Dispersion. *J. Am. Chem. Soc.* **2012**, *134*, 18562–18565. <https://doi.org/10.1021/ja309294u>.
- (56) Ayala, I.; Chiari, L.; Kerfah, R.; Boissouvier, J.; Gans, P.; Hamelin, O. Asymmetric Synthesis of Methyl Specifically Labelled L-Threonine and Application to the NMR Studies of High Molecular Weight Proteins. *ChemistrySelect* **2020**, *5* (17), 5092–5098. <https://doi.org/10.1002/slct.202000827>.
- (57) Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. Interior and Surface of Monomeric Proteins. *J. Mol. Biol.* **1987**, *196* (3), 641–656. [https://doi.org/10.1016/0022-2836\(87\)90038-6](https://doi.org/10.1016/0022-2836(87)90038-6).
- (58) Conte, L. Lo; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol.* **1999**, *285* (5), 2177–2198. <https://doi.org/10.1006/jmbi.1998.2439>.
- (59) Takeda, M.; Ono, A. M.; Terauchi, T.; Kainosho, M. Application of SAIL Phenylalanine and Tyrosine with Alternative Isotope-Labeling Patterns for Protein Structure Determination. *J. Biomol. NMR* **2010**, *46* (1), 45–49. <https://doi.org/10.1007/s10858-009-9360-9>.
- (60) Torizawa, T.; Ono, A. M.; Terauchi, T.; Kainosho, M. NMR Assignment Methods for the Aromatic Ring Resonances of Phenylalanine and Tyrosine Residues in Proteins. *J. Am. Chem. Soc.* **2005**, *127* (36), 12620–12626. <https://doi.org/10.1021/ja051386m>.
- (61) Miyanoiri, Y.; Takeda, M.; Jee, J.; Ono, A. M.; Okuma, K.; Terauchi, T.; Kainosho, M. Alternative SAIL-Trp for Robust Aromatic Signal Assignment and Determination of the X2 Conformation by Intra-Residue NOEs. *J. Biomol. NMR* **2011**, *51* (4), 425–435. <https://doi.org/10.1007/s10858-011-9568-3>.
- (62) Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late Metabolic Precursors for Selective Aromatic Residue Labeling. *J. Biomol. NMR* **2018**, *71* (3), 129–140. <https://doi.org/10.1007/s10858-018-0188-z>.
- (63) Lichtenecker, R. J.; Weinhäupl, K.; Schmid, W.; Konrat, R. α -Ketoacids as Precursors for Phenylalanine and Tyrosine Labelling in Cell-Based Protein Overexpression. *J. Biomol. NMR* **2013**, *57* (4), 327–331. <https://doi.org/10.1007/s10858-013-9796-9>.
- (64) Lichtenecker, R. J. Synthesis of Aromatic $^{13}\text{C}/^2\text{H}$ - α -Ketoacid Precursors to Be Used in Selective Phenylalanine and Tyrosine Protein Labelling. *Org. Biomol. Chem.* **2014**, *12* (38), 7551–7560. <https://doi.org/10.1039/c4ob01129e>.
- (65) Schörghuber, J.; Geist, L.; Platzer, G.; Konrat, R.; Lichtenecker, R. J. Highly Selective Stable Isotope Labeling of Histidine Residues by Using a Novel Precursor in E. Coli-Based Overexpression Systems. *ChemBioChem* **2017**, *18* (15), 1487–1491. <https://doi.org/10.1002/cbic.201700192>.
- (66) Schörghuber, J.; Sára, T.; Bisaccia, M.; Schmid, W.; Konrat, R.; Lichtenecker, R. J. Novel Approaches in Selective Tryptophan Isotope Labeling by Using Escherichia Coli Overexpression Media. *ChemBioChem* **2015**, *16* (5), 746–751. <https://doi.org/10.1002/cbic.201402677>.
- (67) Schörghuber, J.; Geist, L.; Bisaccia, M.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Anthranilic Acid, the New Player in the Ensemble of Aromatic Residue Labeling

- Precursor Compounds. *J. Biomol. NMR* **2017**, *69* (1), 13–22.
<https://doi.org/10.1007/s10858-017-0129-2>.
- (68) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in ¹³C-Labeled Proteins. *J. Am. Chem. Soc.* **1998**, *120* (25), 6394–6400.
<https://doi.org/10.1021/ja980742g>.
- (69) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal Isotope Labelling for NMR Protein Structure Determinations. *Nature* **2006**, *440* (7080), 52–57. <https://doi.org/10.1038/nature04525>.
- (70) Kainosho, M.; Miyanoiri, Y.; Terauchi, T.; Takeda, M. Perspective: Next Generation Isotope-Aided Methods for Protein NMR Spectroscopy. *J. Biomol. NMR* **2018**, *71* (3), 119–127. <https://doi.org/10.1007/s10858-018-0198-x>.
- (71) Danmaliki, G. I.; Liu, P. B.; Hwang, P. M. Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for Escherichia Coli in D₂O. *Biochemistry* **2017**, *56* (45).
<https://doi.org/10.1021/acs.biochem.7b00991>.
- (72) Pandey, A.; Shin, K.; Patterson, R. E.; Liu, X. Q.; Rainey, J. K. Current Strategies for Protein Production and Purification Enabling Membrane Protein Structural Biology. *Biochem. Cell Biol.* **2016**, *94* (6), 507–527. <https://doi.org/10.1139/bcb-2015-0143>.
- (73) Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R. An Efficient and Cost-Effective Isotope Labeling Protocol for Proteins Expressed in Escherichia Coli. *J. Biomol. NMR* **1998**, *11*, 97–102.
<https://doi.org/10.1023/a:1008222131470>.
- (74) Schlegel, S.; Hjelm, A.; Baumgarten, T.; Vikström, D.; De Gier, J. W. Bacterial-Based Membrane Protein Production. *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier 2014, pp 1739–1749. <https://doi.org/10.1016/j.bbamcr.2013.10.023>.
- (75) Wagner, S.; Klepsch, M. M.; Schlegel, S.; Appel, A.; Draheim, R.; Tarry, M.; Wijk, K. J. Van; Slotboom, D. J.; Persson, J. O.; Gier, J. De. Tuning Escherichia Coli for Membrane. *Proc. Natl. Acad. Sci. U. S. A* **2008**, *105* (38), 14371–14376.
<https://doi.org/10.1073/pnas.0804090105>.
- (76) Zhang, Z.; Kuipers, G.; Niemiec, Ł.; Baumgarten, T.; Slotboom, D. J.; De Gier, J.-W.; Hjelm, A. High-Level Production of Membrane Proteins in E. Coli BL21(DE3) by Omitting the Inducer IPTG. *Microb Cell Fact* **2015**, *14*, 142.
<https://doi.org/10.1186/s12934-015-0328-z>.
- (77) Hattab, G.; Warschawski, D. E.; Moncoq, K.; Miroux, B. Escherichia Coli as Host for Membrane Protein Structure Determination: A Global Analysis. *Sci. Rep.* **2015**, *5*, 12097.
<https://doi.org/10.1038/srep12097>.
- (78) Bishop, R. E.; Gibbons, H. S.; Guina, T.; Trent, M. S.; Miller, S. I.; Raetz, C. R. H. Transfer of Palmitate from Phospholipids to Lipid A in Outer Membranes of Gram-Negative Bacteria. *EMBO J.* **2000**, *19* (19), 5071–5080.
<https://doi.org/10.1093/emboj/19.19.5071>.
- (79) Guo, L.; Lim, K. B.; Poduje, C. M.; Daniel, M.; Gunn, J. S.; Hackett, M.; Miller, S. I. Lipid A Acylation and Bacterial Resistance against Vertebrate Antimicrobial Peptides. *Cell* **1998**, *95* (2), 189–198. [https://doi.org/10.1016/S0092-8674\(00\)81750-X](https://doi.org/10.1016/S0092-8674(00)81750-X).
- (80) Tanamoto, K.; Azumi, S. Salmonella -Type Heptaacylated Lipid A Is Inactive and Acts

- as an Antagonist of Lipopolysaccharide Action on Human Line Cells . *J. Immunol.* **2000**, *164* (6), 3149–3156. <https://doi.org/10.4049/jimmunol.164.6.3149>.
- (81) Bishop, R. E. The Lipid A Palmitoyltransferase PagP: Molecular Mechanisms and Role in Bacterial Pathogenesis. *Mol. Microbiol.* **2005**, *57* (4), 900–912. <https://doi.org/10.1111/j.1365-2958.2005.04711.x>.
- (82) Ahn, V. E.; Lo, E. I.; Engel, C. K.; Chen, L.; Hwang, P. M.; Kay, L. E.; Bishop, R. E.; Prive, G. G. A Hydrocarbon Ruler Measures Palmitate in the Enzymatic Acylation of Endotoxin. *EMBO J.* **2004**, *23* (15), 2931–2941. <https://doi.org/10.1038/sj.emboj.7600320>.
- (83) Hwang, P. M.; Bishop, R. E.; Kay, L. E. The Integral Membrane Enzyme PagP Alternates between Two Dynamically Distinct States. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9618–9623. <https://doi.org/10.1073/pnas.0402324101>.
- (84) Kay, L. E.; Torchia, D. A.; Bax, A. Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **1989**, *28* (23), 8972–8979. <https://doi.org/10.1021/bi00449a003>.
- (85) Farrow, N. A.; Muhandiram, R.; Pascal, S. M.; Kay, L. E.; Singer, A. U.; Forman-Kay, J. D.; Kay, C. M.; Gish, G.; Pawson, T.; Shoelson, S. E. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry* **1994**, *33* (19), 5984–6003. <https://doi.org/10.1021/bi00185a040>.
- (86) Kleckner, I. R.; Foster, M. P. An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814* (8), 942–968. <https://doi.org/10.1016/j.bbapap.2010.10.012>.
- (87) Palmer Arthur G, I. I. Probing Molecular Motion by NMR. *Curr. Opin. Struct. Biol.* **1997**, *7* (5), 732–737.
- (88) Kay, L. E.; Torchia, D. A.; Bax, A. Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **1989**, *28* (23), 8972–8979. <https://doi.org/10.1021/bi00449a003>.
- (89) Farrow, N. A.; Muhandiram, R.; Pascal, S. M.; Kay, L. E.; Singer, A. U.; Forman-Kay, J. D.; Kay, C. M.; Gish, G.; Pawson, T.; Shoelson, S. E. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry* **1994**, *33* (19), 5984–6003. https://doi.org/10.1021/BI00185A040/SUPPL_FILE/BI00185A040_SI_001.PDF.
- (90) Palmer, A. G.; Massi, F. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* **2006**, *106* (5), 1700–1719. <https://doi.org/10.1021/cr0404287>.
- (91) Massi, F.; Johnson, E.; Wang, C.; Rance, M.; Palmer, A. G. NMR R1ρ Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* **2004**, *126* (7), 2247–2256. <https://doi.org/10.1021/ja038721w>.
- (92) Renner, C.; Schleicher, M.; Moroder, L.; Holak, T. A. Practical Aspects of the 2D 15N-{1H} NOE Experiment. *J Biomol NMR* **2002**, *23* (1), 23–33.
- (93) Kawale, A. A.; Burmann, B. M. Characterization of Backbone Dynamics Using Solution NMR Spectroscopy to Discern the Functional Plasticity of Structurally Analogous Proteins. *STAR Protoc.* **2021**, *2* (4). <https://doi.org/10.1016/j.xpro.2021.100919>.
- (94) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am.*

- Chem. Soc.* **1982**, *104* (17), 4559–4570. <https://doi.org/10.1021/ja00381a010>.
- (95) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104* (17), 4546–4559. <https://doi.org/10.1021/ja00381a009>.
- (96) Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon Press, Oxford, 1961.
- (97) Hiyama, Y.; Niu, C. H.; Silverton, J. V.; Bavoso, A.; Torchia, D. A. Determination of ¹⁵N Chemical Shift Tensor via ¹⁵N-²H Dipolar Coupling in Boc-Glycylglycyl[¹⁵N]Glycine Benzyl Ester. *J. Am. Chem. Soc.* **1988**, *110* (8), 2378–2383. <https://doi.org/10.1021/ja00216a006>.
- (98) Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Wingfield, P. T. Analysis of the Backbone Dynamics of Interleukin-1 β Using Two-Dimensional Inverse Detected Heteronuclear ¹⁵N-¹H NMR Spectroscopy. *Biochemistry* **1990**, *29* (32), 7387–7401. <https://doi.org/10.1021/bi00484a006>.
- (99) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J. Am. Chem. Soc.* **1990**, *112* (12), 4989–4991. https://doi.org/10.1021/JA00168A070/ASSET/JA00168A070.FP.PNG_V03.
- (100) Li, F.; Grishaev, A.; Ying, J.; Bax, A. Side Chain Conformational Distributions of a Small Protein Derived from Model-Free Analysis of a Large Set of Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, *137* (46), 14798–14811. <https://doi.org/10.1021/jacs.5b10072>.
- (101) Mittermaier, A.; Kay, L. E. X1 Torsion Angle Dynamics in Proteins, From Dipolar Couplings. *J. Am. Chem. Soc.* **2001**, *123* (28), 6892–6903. <https://doi.org/10.1021/ja010595d>.
- (102) Yang, D. Probing Protein Side Chain Dynamics Via ¹³C NMR Relaxation. *Protein Pept. Lett.* **2011**, *18* (4), 380–395. <https://doi.org/10.2174/092986611794653932>.
- (103) Zhang, X.; Sui, X.; Yang, D. Probing Methyl Dynamics from ¹³C Autocorrelated and Cross-Correlated Relaxation. *J. Am. Chem. Soc.* **2006**, *128* (15), 5073–5081. <https://doi.org/10.1021/ja057579r>.
- (104) Muhandiram, D. R.; Yamazaki, T.; Kay, L. E.; Sykes, B. D. Measurement of ²H T₁ and T_{1 ρ} Relaxation Times in Uniformly ¹³C-Labeled and Fractionally ²H-Labeled Proteins in Solution. *J. Am. Chem. Soc.* **1995**, *117* (46), 11536–11544. <https://doi.org/10.1021/ja00151a018>.
- (105) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in ¹³C-Labeled and Fractionally ²H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **2002**, *124* (22), 6439–6448. <https://doi.org/10.1021/ja012497y>.
- (106) Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. Internal Dynamics of Human Ubiquitin Revealed by ¹³C-Relaxation Studies of Randomly Fractionally Labeled Protein. *Tech. Protein Chem.* **1997**, *8* (C), 715–725. [https://doi.org/10.1016/S1080-8914\(97\)80070-1](https://doi.org/10.1016/S1080-8914(97)80070-1).
- (107) Yang, D.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. A Study of Protein Side-Chain Dynamics from New ²H Auto-Correlation and ¹³C Cross-Correlation NMR Experiments: Application to the N-Terminal SH3 Domain from Drk. *J. Mol. Biol.* **1998**, *276* (5), 939–954. <https://doi.org/10.1006/jmbi.1997.1588>.
- (108) Zheng, Y.; Yang, D. Measurement of Dipolar Cross-Correlation in Methylene Groups in

- Uniformly ^{13}C -, ^{15}N -Labeled Proteins. *J. Biomol. NMR* **2004**, 28 (2), 103–116.
<https://doi.org/10.1023/B:JNMR.0000013826.82936.0e>.
- (109) Liu, W.; Zheng, Y.; Cistola, D. P.; Yang, D. Measurement of Methyl ^{13}C - ^1H Cross-Correlation in Uniformly ^{13}C -, ^{15}N -, Labeled Proteins Weidong. *J. Biomol. NMR* **2003**, 27, 351–364. <https://doi.org/10.1023/a:1025884922203>.
- (110) Skrynnikov, N. R.; Millet, O.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* **2002**, 124 (22), 6449–6460.
<https://doi.org/10.1021/ja012498q>.

Chapter 2

Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for *Escherichia coli* in D₂O

A version of this chapter has been previously published as: Danmaliki, G.I., Liu, P.B., Hwang, P.M. Stereoselective deuteration in aspartate, asparagine, lysine, and methionine amino acid residues using fumarate as a carbon source for *Escherichia coli* in D₂O. *Biochemistry* 56, 45, 6015–6029 (2017). PMH directed the research. PBL helped with protein purification. GID expressed and purified the human Pin1 WW domain protein and its constructs, tested several metabolic precursors and inhibitors, acquired all the 2D and 3D NMR data, assigned the protein chemical shifts, and analyzed the data under PMH supervision. GID wrote the manuscript, and PMH edited it.

Summary

Perdeuteration with selective ¹H,¹³C-enrichment of methyl groups has enabled solution NMR studies of large (>30 kDa) protein systems. However, we propose that for all non-methyl positions, only magnetization originating from ¹H-¹²C groups is sufficiently long-lived, and it can be transferred via through-space NOEs to slowly relaxing ¹H-¹⁵N or ¹H-¹³C methyl groups to achieve multidimensional solution NMR. We demonstrate stereoselective ¹H,¹²C-labeling by adding relatively inexpensive unlabeled carbon sources to *Escherichia coli* growth media in D₂O. Using our model system, a mutant WW domain from human Pin1, we compare deuteration patterns in 19 amino acids (all except cysteine). Protein grown using glucose as the sole carbon source had high levels of protonation in aromatic rings and the H β positions of serine and tryptophan. In contrast, using our FROMP media (fumarate, rhamnose, oxalate, malonate, pyruvate), stereoselective protonation of H β_2 with deuteration at H α and H β_3 was achieved in Asp, Asn, Lys, and Met residues. In solution NMR, stereospecific chemical shift assignments for H β are typically obtained in conjunction with χ_1 dihedral angle determinations using 3-bond J-coupling (³JN-H β , ³JCO-H β , ³JH α -H β) experiments. However, due to motional averaging, the assumption of a pure rotameric state can yield incorrect χ_1 dihedral angles with incorrect

stereospecific assignments. This was the case for three residues in the Pin1 WW domain (Lys28, Met30, and Asn44). Thus, stereoselective ^1H , ^{12}C -labeling will be useful not only for NMR studies of large protein systems, but also for determining side chain rotamers and dynamics in any protein system.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for the elucidation of protein structure and dynamics at atomic resolution, contributing about 10% of all the structures deposited in the Protein Data Bank¹. The development of multinuclear multidimensional solution NMR combined with uniform isotopic enrichment (^{13}C , ^{15}N) supports the determination of protein structures up to 25 kDa^{2,3}. However, for systems more than 25 kDa, NMR studies are hampered by rapid signal decay, determined by the transverse relaxation rate, which scales roughly with molecular weight. One strategy for minimizing the transverse relaxation rate is to replace the ^1H nuclei in a protein with ^2H to minimize ^1H - ^1H and ^1H - ^{13}C dipolar relaxation⁴⁻⁷. However, this approach reduces the number of available NOE-based ^1H - ^1H distance restraints required for high-resolution structure determination⁸. Reincorporating ^1H at specific sites in a highly deuterated background is therefore necessary. Several approaches have been developed over the past decades to achieve this goal.

Random fractional deuteration has been utilized by several groups,^{4,9} and depending on the size of the protein, a deuteration level of 50-90% provides a good compromise between spectral quality and available distance information necessary for high-resolution structure determination. The disadvantage of this technique is that it produces numerous isotopomers that disperse signals into multiple peaks due to deuterium isotope shift, compromising sensitivity and resolution^{4,9}.

Methyl groups are ideal molecular probes for solution NMR spectroscopy studies of large proteins. Protocols for selective protonation of methyl groups of Ala, Thr, Ile, Leu, Val, and Met in perdeuterated background have been developed¹⁰⁻¹⁷. Rapid rotation about the methyl symmetry axis attenuates ^1H - ^{13}C dipolar relaxation. An inability to obtain structural information for all non-methyl-containing amino acid side chains is the major limitation of this approach. For instance, the side chains of the aromatic amino acids are very important for hydrophobic packing

in proteins and ligand binding interfaces. ^{12}C reverse labeling of the aromatic side chains of phenylalanine and tyrosine simplifies spectral crowding and overlap, improving sensitivity while providing long range NOEs^{18,19}. However, one drawback of this approach is that strong ^1H - ^1H dipolar relaxation still occurs between vicinal protons in the aromatic rings.

The stereo-array isotope labeling (SAIL) technique developed by Kainosho and co-workers is a creative approach that provides a unique solution to rapid transverse relaxation in NMR²⁰. In this technique, all twenty amino acids are chemically and enzymatically synthesized and then incorporated into protein with a cell-free expression system. The amino acids are synthesized with a stereospecific and regiospecific pattern fully ^{13}C labelled, with only one ^1H attached to every ^{13}C atom and all other sites completely deuterated. The technique greatly enhances spectral sensitivity and resolution, but the major drawbacks are that it is extremely labor-intensive and costly, and fast dipolar relaxation still dominates any non-methyl ^1H - ^{13}C group even in a highly deuterated background.

Despite advances to judiciously incorporate ^1H into proteins in an otherwise highly deuterated background, new methods are needed to obtain structural information from all 20 amino acid residues in a protein. Here, we present a protocol for producing proteins in *Escherichia coli* (*E. coli*) using inexpensive non-isotopically enriched carbon sources, taking advantage of the inherent amino acid biosynthetic pathways in *E. coli* to incorporate ^1H stereospecifically into proteins produced in D_2O -based media. This approach eliminates the dominant sources of dipolar relaxation (^1H - ^1H and ^1H - ^{13}C dipoles), producing isolated ^1H - ^{12}C groups in a largely deuterated background. This provides additional ^1H magnetization that can be transferred via through-space NOEs to nearby ^1H - ^{15}N and ^1H - ^{13}C methyl groups, which can then be resolved via multinuclear NMR to provide the necessary structural restraints for high-resolution structure determination of large protein systems.

Experimental Procedures

Construct design

We have applied our labeling technique to the WW domain from the human Pin1 protein, the smallest folded protein domain we could identify containing all 20 amino acids except cysteine (about 30 residues). It has been characterized by extensive biophysical studies and

mutagenesis. The rate-limiting step for the folding of this protein is the formation of a “loop 1” structure (between the first two β -strands) consisting of an unusual four-residue type-II turn inserted within a larger six-residue loop that is implicated in protein-protein interactions^{21,22}. Replacing the wild-type loop 1 with a shorter sequence has been shown to improve both folding and thermal stability by an order of magnitude²¹. The wild-type protein has a low yield when expressed in *E. coli*, typically 2 mg/L, explaining why most studies utilized peptide synthesis over recombinant expression^{21,23,24}. Due to these reasons, we designed, optimized, and expressed a mutant human Pin1 WW domain construct in *E. coli* possessing a shortened loop 1,²¹ and a His-tag connected by a long N-terminal linker (Figure 2.1). An expression vector for this was synthesized by ATUM (formerly DNA2.0), with a high copy number origin of replication, ampicillin selection marker, strong ribosome binding site, lacI repressor gene, and a T5 promoter controlled by two flanking lacO sites that allow induction by the addition of IPTG (isopropyl β -D-1-thiogalactopyranoside) in *E. coli*. The T5 promoter is recognized by *E. coli* RNA polymerase, so the construct can be expressed in any strain. Codons were optimized by ATUM for expression in *E. coli*, and the 53-amino acid construct yielded ~75 mg/L when expressed in 1 L minimal M9 media (10 g glucose).

10
20
30
40
50
 MGHHHHHHSS GGSSGMADEE KLPPG**WEKRM** S-ADGRVYYFN HITNASQWER PSG

Figure 2.1 Mutant human Pin1 WW domain sequence. Red lettering denotes modifications made to the native amino acid sequence: an N-terminal His-tag followed by a flexible linker, as well as an Ala-Asp substitution/deletion for Ser-Gly-Arg in the middle of the protein in loop 1²¹. The three anti-parallel β -strands are shown in bold.

Production of (¹⁵N, ¹³C)-mutant human Pin1 WW domain in H₂O minimal M9 media

To facilitate complete stereospecific resonance assignments, fully protonated uniformly (¹⁵N,¹³C)-labeled mutant human Pin1 WW domain was expressed in *E. coli* BL21(DE3) with slight modifications to methods described earlier^{3,25,26}. The protein was expressed in 1 L M9 minimal media, containing 9 g Na₂HPO₄ and 2.5 g KH₂PO₄ dissolved in 950 mL H₂O, yielding a pH 7.3-7.4. A 50 mL solution containing 1 g ¹⁵NH₄SO₄, 3g ¹³C-glucose, 1 mL 1 M MgSO₄, 1 mL 0.1M CaCl₂, 1 mL 5 % ampicillin solution, 1 mg biotin, and 100 mg thiamine was filtered and added to the autoclaved 950 mL M9 salts.

Four to six transformed colonies were inoculated into 10 mL LB media containing 100 mg/L ampicillin and incubated at 37 °C. After reaching $A_{600} \sim 1.0$, the cells were diluted into 1 L minimal M9 media and allowed to grow further until $A_{600} \sim 0.8$. The cells were induced with 1 mM IPTG and allowed to grow for six hours post-induction. The cells were centrifuged at 5000 rpm (4420 x g) for 20 min and harvested. The cell pellet was re-suspended in 20 mL lysis buffer: 50 mM Tris (pH 8.0), 10 mM $MgSO_4$, 10 ug/mL DNase I. 200 mg sodium deoxycholate and 20 mg lysozyme, each pre-dissolved in 1 mL distilled water, was added to the lysis buffer to disrupt cell membranes and cell walls. The lysate was centrifuged at 15000 rpm (27,000 x g) for 15 minutes, and the supernatant was syringe-filtered with a 0.45 μm cut-off filter. The supernatant was applied to a Qiagen Ni-NTA column equilibrated with binding buffer (20mM Tris-HCl, 300mM NaCl, 10 mM imidazole), and then washed with the same buffer containing 80 mM imidazole. Protein was eluted with the same buffer containing 250 mM imidazole. Fractions were assessed by SDS-PAGE, and fractions containing pure protein were dialyzed against 6.5 mM ammonium bicarbonate for three days and then lyophilized. Mass spectrometry was used to confirm the identity of the protein and showed that the initial methionine in the sequence had been removed. To facilitate the stereospecific assignments of methyl groups, a mutant human Pin1 WW domain sample was produced using 10% uniformly (1H , ^{13}C)-enriched glucose and 90% unenriched glucose as described previously²⁷.

Production of (^{15}N , ^{12}C)-mutant human Pin1 WW domain in D_2O -M9 minimal media

To assess the specific $^1H/2H$ incorporation into individual amino acids, mutant human-Pin1 WW domain was expressed in D_2O , using protonated natural abundance ^{12}C -glucose rather than ^{13}C -glucose as described in the previous section. The expression and purification process was basically the same as the previous section, except that cells were diluted into D_2O M9 media prior to induction. The bacterial cells were grown in D_2O media for at least one doubling time, which was three hours. Cells were induced at $A_{600nm} \sim 0.8$ using 1 mM IPTG and allowed to grow for six hours post-induction. The purification was the same as in the previous section, yielding isotopically enriched protein that we will denote as Pin1(glucose). The level of $^1H/2H$ incorporation in every amino acid was determined by comparing the peak intensities from a 1H - ^{13}C HSQC spectrum of Pin1(glucose) with that of a fully protonated sample, Pin1(unlabeled).

The HSQC was performed with long ^{13}C indirect acquisition times (20 ms) without constant time, which was made possible because only natural abundance (1%) ^{13}C isotope was used.

Rationale for isotope labeling using D_2O -“FROMP” minimal media

The major concern in producing perdeuterated proteins is the phenomenon of isotopic “scrambling”, whereby metabolic precursors are processed variably through multiple pathways before they are incorporated into amino acids, resulting in multiple isotopomers, like CH_2 , CHD , and CD_2 .^{11,28} The presence of multiple isotopomers spreads signals into multiple peaks, compromising both sensitivity and resolution. One potential cause of scrambling is the interconversion between phosphoenolpyruvate, pyruvate, and oxaloacetate at the juncture of the key metabolic pathways: glycolysis, tricarboxylic acid (TCA) cycle, and gluconeogenesis (Figure 2.2)²⁹. Oxalate has been shown to inhibit many of the enzymes at this key metabolic juncture including the gluconeogenic enzymes, phosphoenolpyruvate carboxykinase (PEPCK), malic enzyme, and phosphoenolpyruvate synthetase (Figure 2.2)³⁰⁻³². We reasoned that oxalate could thus be used to limit scrambling, so long as precursors are supplied on both sides of the blockade: pyruvate on the side of the TCA cycle and rhamnose on the glycolysis/gluconeogenesis side (Figure 2.2). Rhamnose was chosen instead of glucose because it is metabolized into L-lactaldehyde (which can be converted to pyruvate) and dihydroxyacetone phosphate. Gluconeogenic synthesis of glucose from dihydroxyacetone phosphate and subsequent metabolism to erythrose-4-phosphate via the pentose phosphate pathway would result in a higher degree of deuteration at the aromatic amino acids than if glucose were supplied in the media. (see Figure 2.2).

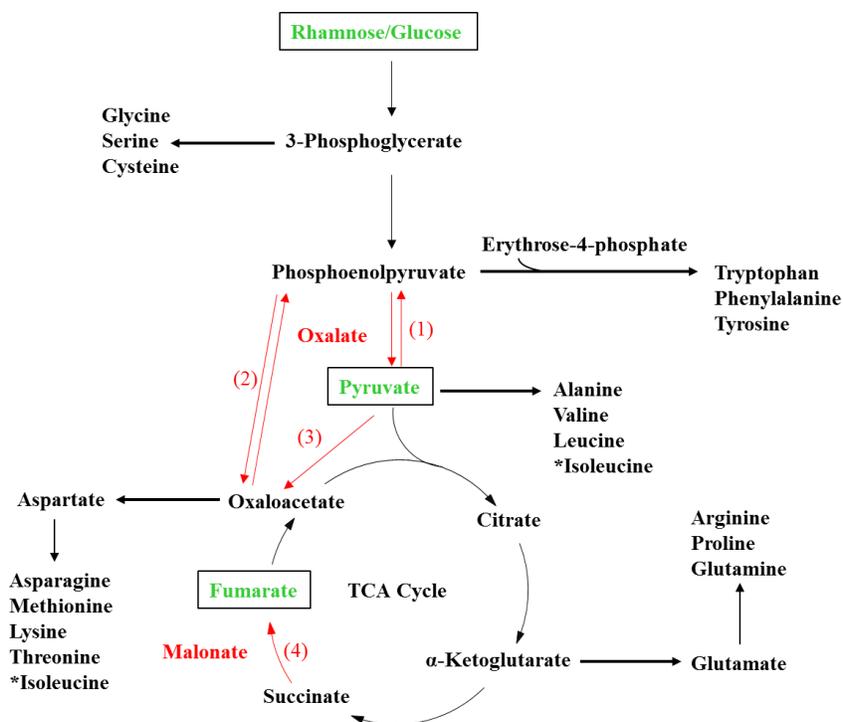


Figure 2.2 Overview of amino acid biosynthetic pathways. Carbon sources added to the growth media: rhamnose/glucose, pyruvate, and fumarate (green). Enzyme inhibitors: oxalate and malonate (red). Oxalate inhibits multiple enzymes involved in the interconversion of phosphoenolpyruvate, pyruvate, and oxaloacetate/malate: (1) phosphoenolpyruvate synthetase and pyruvate kinase, (2) phosphoenolpyruvate carboxykinase, and (3) pyruvate carboxylase. Malonate inhibits succinate dehydrogenase (4), preventing oxidation of succinate to fumarate.

We also postulated that the amino acids belonging to the oxaloacetate family in the TCA cycle (aspartate, asparagine, lysine, threonine, and methionine) could be stereospecifically labeled with ^1H at the $\text{H}^{\beta 2}$ position and ^2H at the $\text{H}^{\beta 3}$ position by supplying fumarate as an additional carbon source. Fumarate is converted into malate in a reaction catalyzed by fumarase, with one ^2H stereospecifically incorporated from D_2O (Figure 2.3)³³. Malate is then metabolized within the TCA cycle to oxaloacetate, which is converted into aspartate by transamination. Endogenous fumarate production by the TCA cycle can be suppressed by the addition of malonate to the growth media (Figure 2.2)^{34,35}.

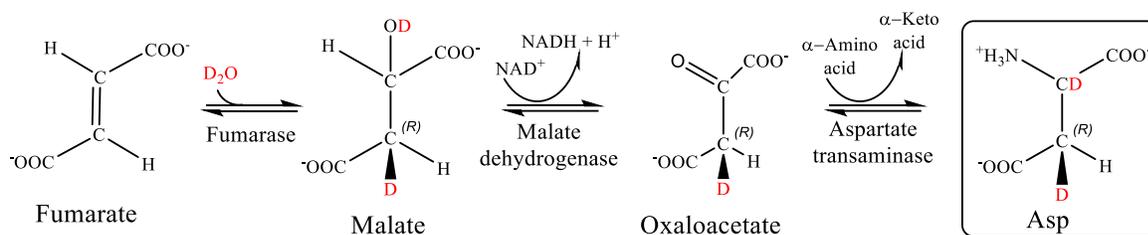


Figure 2.3. Stereoselective incorporation of ^2H from D_2O into the $\text{H}^{\beta 3}$ position in aspartate. Fumarate supplied in the growth media is converted into malate in a reaction catalyzed by fumarase, with one ^2H stereospecifically incorporated from D_2O . Malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate, which is converted to aspartate via a transamination reaction.

Production of mutant human Pin1 WW domain in D_2O -“FROMP” (fumarate, rhamnose, oxalate, malonate, and pyruvate) media

Cells were grown initially in 10 mL LB till an $A_{600\text{nm}} \sim 1$ was reached. The cells were diluted in 1 L M9/ H_2O medium containing 1g NH_4Cl and 3g ^{12}C -rhamnose. Doubling time was two hours, and the cells were allowed to grow till $A_{600\text{nm}} \sim 0.7$. The cells were then harvested and re-suspended in 0.9 L M9/ D_2O , containing 8 g Na_2HPO_4 , 2.2 g K_2HPO_4 , 1g NH_4Cl and 3g ^{12}C -rhamnose, 0.11 g MgSO_4 (anhydrous), 0.01 g CaCl_2 (anhydrous), 90 mg ampicillin, 100 mg thiamine, and 1 mg biotin that was sterile filtered. The bacterial cells were grown for one hour prior to the addition of 3 g sodium pyruvate, 3 g sodium fumarate, 1 g oxalic acid and 1 g malonic acid, induced at $A_{600\text{nm}} \sim 0.9$, and allowed to grow for six hours post-induction. Protein purification was the same as in the previous sections, yielding ~ 10 mg per liter of growth media. We will denote the NMR sample from this preparation as Pin1(FROMP). Peak intensities obtained from Pin1(FROMP) were compared to Pin1(unlabeled) and Pin1(glucose) using ^1H - ^{13}C HSQC spectra.

NMR Spectroscopy

All NMR samples were 500 μL in volume. The buffer conditions were: 100 mM KCl, 10 mM imidazole, 0.5 mM 2,2-dimethyl-2-silapentane-5-sulfonate- d_6 sodium salt (DSS- d_6) as an NMR chemical shift internal reference, and 0.01 % NaN_3 in 90% H_2O , 10% D_2O or 100 % D_2O , pH 6.8. Concentrations ranged from 0.5 to 2 mM protein.

All NMR experiments were conducted at 30 °C on a Varian Inova 500 MHz spectrometer. The spectrometer was equipped with triple resonance probes and Z-pulsed field gradients. All one-dimensional experiments were processed using VNMRJ (Varian Associates) and all two-dimensional and three-dimensional NMR experiments were processed using NMRPipe/NMRDraw software³⁶. The 2D and 3D spectra were analyzed further using NMRviewJ (One Moon Scientific)³⁷.

Backbone ¹H, ¹⁵N, ¹³C chemical shift assignments for the mutant human Pin1 WW domain were obtained by analyzing the 3D HNCACB and 3D CBCA(CO)NH experiments. Sidechain ¹H and ¹³C chemical shift assignments were obtained by analyzing the (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY experiments. Aromatic side-chain resonances were assigned using an aromatic 3D ¹³C-edited NOESY-HSQC (mixing time 100 ms). The 2D constant time ¹H, ¹³C-HSQC experiment on an NMR sample prepared with 10% ¹³C-glucose labeling was used to obtain stereospecific assignments of Leu and Val methyl groups in the protein (Pro-S methyl groups are in phase with methionine methyls, which are free from 1-bond ¹³C-¹³C J couplings). The chi-1 dihedral angle and the stereospecific assignment for H^β protons were obtained by analyzing the following experiments: 3D HNHB, 3D HN(CO)HB, and ¹H-TOCSY-¹⁵N-HSQC (28 ms mixing time). For these experiments, the intensity of the correlations between HN and H^{β2}/H^{β3} depends on ³J_{N-Hβ}, ³J_{CO-Hβ}, ³J_{Hα-Hβ} respectively, giving an internally redundant dataset for determining if H^{β2}/H^{β3} forms 60° (small ³J) or 180° (large ³J) dihedral angles with N, CO, and Hα, respectively, as determined by the three major χ_1 rotameric states: trans, gauche⁻, and gauche⁺.

Results and Discussion

Chemical shift assignment of Pin1 WW domain

The 2D ¹H-¹⁵N HSQC NMR spectrum of fully protonated Pin1 WW domain showed well dispersed amide proton signals, characteristic of a folded protein domain. A near-complete backbone and sidechain assignment was achieved for the structured region of the protein including all aromatic sidechains, providing probes of isotope labeling patterns in 19 out of the 20 naturally occurring amino acids (all except cysteine).

Quantitation of $^1\text{H}/^2\text{H}$ incorporation in Pin1 WW domain grown in D_2O - ^{12}C -glucose or D_2O -(FROMP) media

The 2D ^1H - ^{13}C HSQC spectrum of Pin1(unlabeled), shown in Figure 2.4a, serves as a reference against which deuterated samples can be compared, facilitating the testing of various unlabeled carbon sources for biosynthetic incorporation in *E. coli*. All HSQC spectra relied on natural abundance ^{13}C , allowing high resolution of the indirect ^{13}C dimension without constant time evolution. Peak intensities from Pin1(unlabeled) were compared to the peak intensities of Pin1(glucose) (Figure 2.4b) or Pin1(FROMP) (Figure 2.4c) to determine the level of $^1\text{H}/^2\text{H}$ incorporation into every amino acid site, normalized by a scaling factor. The scaling factor corrects for differences in concentration between the samples and was very close to 1. For aliphatic positions, the scaling factor was derived by comparing intensities of serine β -methylene hydrogens, since these were the most highly protonated positions in all samples studied (see Table 2.1). The scaling factor was calculated to ensure that the sum of isotopomers at the serine β position was the same for all three protein samples. It is also important to note that in quantitating each isotopomer, individual peaks corresponding to CH_3 and $\text{CH}_2(\text{D})$ have to be scaled down by factors of 3 and 2, respectively, when compared to $\text{CH}(\text{D})$, due to the signal contribution of every proton. For aromatic sidechain positions, the most highly protonated positions were the $\text{H}\delta$ ring protons of Phe/Tyr, and these appeared to be protonated $\sim 100\%$ in all samples. The aromatic region of Pin1(unlabeled) is shown in Figure 2.5a and compared that of Pin1(glucose) (Figure 2.5b) or Pin1(FROMP) (Figure 2.5c).

When *E. coli* is grown in D_2O , almost all alpha hydrogens are derived biosynthetically from solvent during transamination reactions, so these positions are highly deuterated in Pin1(glucose) (Figure 2.4b) or Pin1(FROMP) (Figure 2.4c). The result agrees well with the findings of Rosen *et al*¹¹ and Otten *et al*³⁸, who observed ($\sim 95\%$) deuteration at the H^α of all amino acids. A notable exception in our study, however, is the H^α positions of methionine residues, which retain a small degree of protonation (see Figures 2.4b and 2.4c), a distinction we are at a loss to explain.

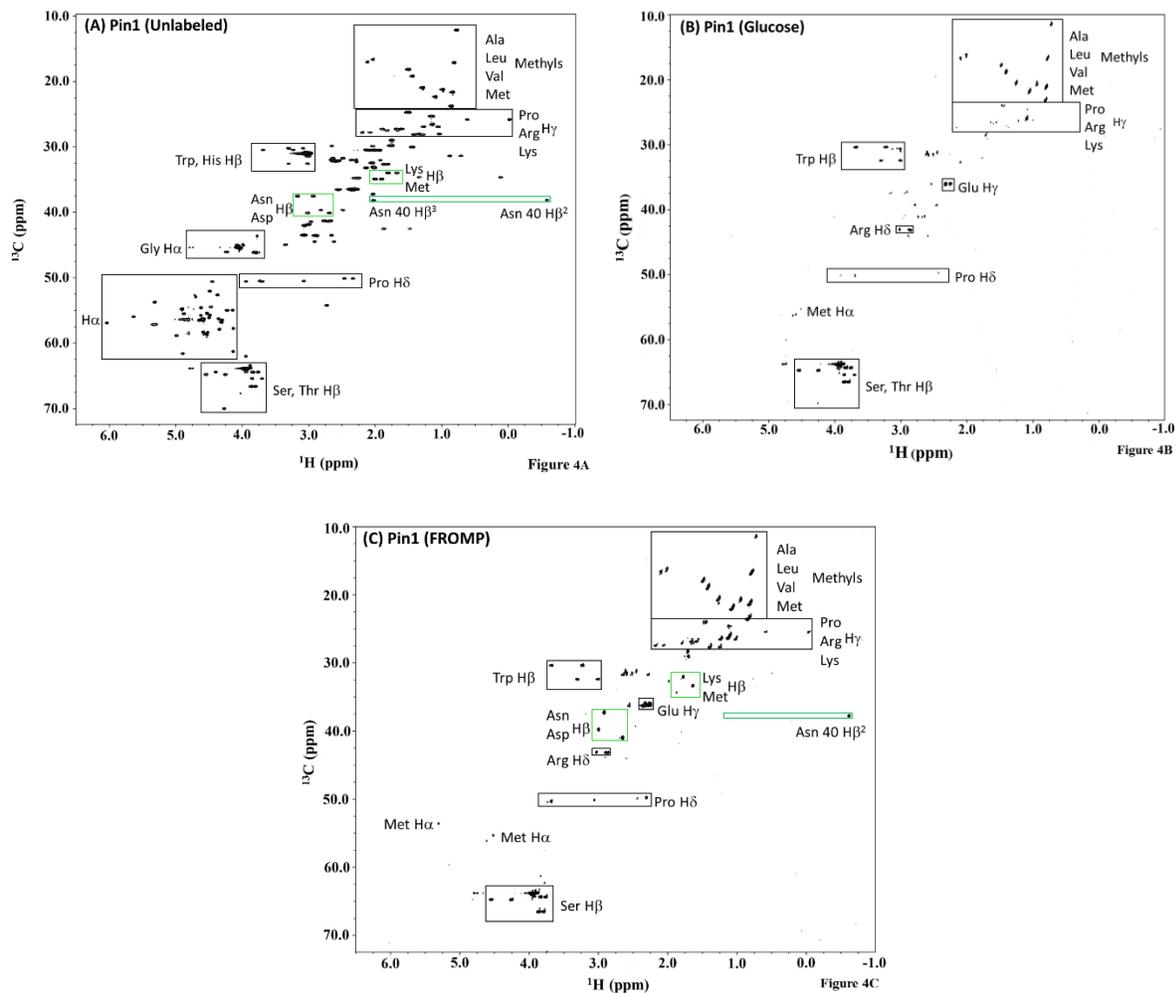


Figure 2.4. Natural abundance 2D ^1H - ^{13}C HSQC (aliphatic region) (A) Fully protonated Pin1 (unlabeled) (B) Deuterated Pin1 (glucose) (C) Deuterated Pin1 (FROMP). Highlighted in black are the fingerprint regions of human Pin1 for methyl-containing amino acids, H^α of all residues, H^β positions of Trp, His, Ser, Thr, H^γ positions of Pro, Arg, Lys, and H^δ positions of Pro. Highlighted in green are the H^β protons of Asp, Asn, Lys, and Met. Protein concentration was 2 mM, pH \sim 6.8 in each sample. The acquisition and processing parameters for each spectrum were identical for comparison.

Table 2.1. Average percent isotopomers in the amino acids of mutant Pin1 WW domain grown in glucose/FROMP and D₂O

Amino acid type	C β		C γ			C δ			C ϵ			
Val	CH: 0/0	CD: 100/100										
Ile	CH: 0/0	CD: 100/100		CH ₂ : 0/0	CHD: 27/0	CD ₂ : 73/100						
Leu	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH: 0/0	CD: 100/100							
Glu	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/35	CHD: 30/38	CD ₂ : 70/27						
Gln	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/35	CHD: 32/54	CD ₂ : 68/11						
Arg ^a	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/53	CHD: 43/45	CD ₂ : 57/2	CH ₂ : 0/0	CHD: 80/50	CD ₂ : 20/50			
Pro ^a	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/40	CHD: 30/33	CD ₂ : 70/27	CH ₂ : 0/0	CHD: 33/48	CD ₂ : 67/52			
Asp ^a	CH ₂ : 0/0	CHD: 0/156	CD ₂ : 100/0									
Asn	CH ₂ : 0/0	CHD: 20/160	CD ₂ : 80/0									
Met	CH ₂ : 0/0	CHD: 0/30	CD ₂ : 100/70	CH ₂ : 0/0	CHD: 35/45	CD ₂ : 65/55						
Lys ^a	CH ₂ : 0/0	CHD: 0/90	CD ₂ : 100/10	CH ₂ : 0/0	CHD: 38/35	CD ₂ : 62/65	CH ₂ : 0/0	CHD: 22/76	CD ₂ : 78/24	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100
Thr	CH: 27/0	CD: 73/100										
Ser ^b	CH ₂ : 90/75	CHD: 10/25	CD ₂ : 0/0									
Trp ^a	CH ₂ : 80/80	CHD: 20/20	CD ₂ : 0/0									
Phe	CH ₂ : 0/0	CHD: 40/35	CD ₂ : 60/65									
Tyr ^a	CH ₂ : 0/0	CHD: 30/30	CD ₂ : 70/70									
His	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100									

^aaverage of two residues, ^baverage of three residues.

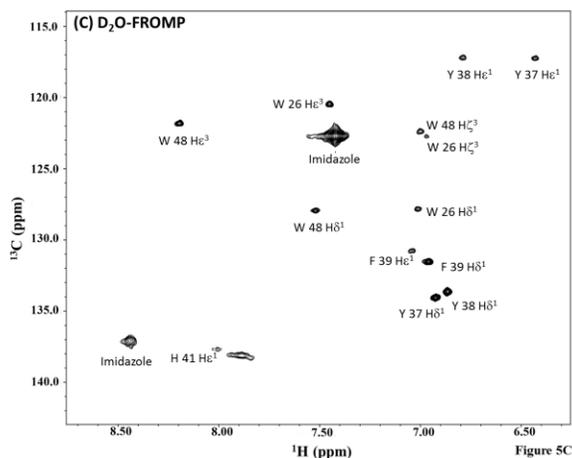
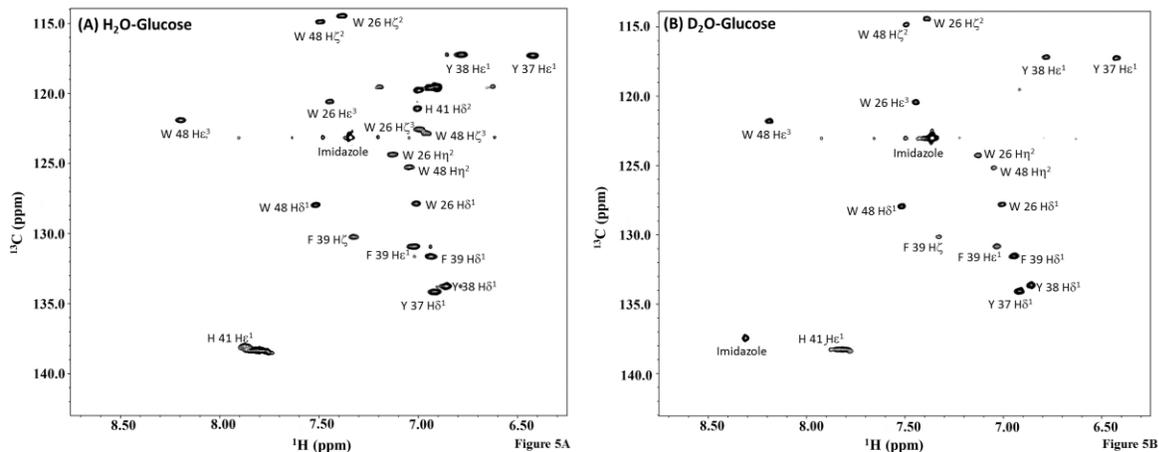


Figure 2.5. Natural abundance 2D ^1H - ^{13}C HSQC (aromatic region) (A) Fully protonated human Pin1 expressed in H_2O -glucose (B) Human Pin1 expressed in D_2O -glucose (C) Human Pin1 expressed in D_2O -FROMP. Protein concentration was 2 mM, pH \sim 6.8 in each sample. The acquisition and processing parameters for each spectrum were identical for comparison.

Pyruvate family (alanine, valine, leucine, isoleucine- γ 2)

Methyl groups are excellent probes of structure and dynamics in solution NMR studies because of their favourable relaxation properties and occurrence in protein hydrophobic cores and protein-protein interfaces^{39,40}. Experimental protocols for the production of highly deuterated proteins with selective ^1H - ^{13}C labeling at methyl groups of alanine, valine, leucine and isoleucine have been established in many studies.^{3,38} In theory, these protocols can be combined with our

FROMP-media to produce ^1H - ^{13}C methyl labeling at these amino acids. Thus, we will not discuss the biosynthetic pathways for these amino acids in detail, focusing more on other residue types for which isotope labeling strategies have not been as well established.

Isotopomeric distributions for methyl groups have been studied in detail using ^1H -glucose as carbon source by Otten *et al*³⁸ and Shekhtman *et al*⁴¹, as well as by Rosen *et al*¹¹ for ^1H -pyruvate. The results obtained for Pin1(FROMP) are similar to those obtained by Rosen *et al*¹¹ for ^1H -pyruvate, as expected, given that pyruvate is the source of Ala, Val, Leu, and Ile- γ 2 methyl groups in both studies. It is noteworthy that we obtain the same isotopomeric distributions for these amino acids despite our FROMP media also containing rhamnose and fumarate, which would be metabolized to phosphoenolpyruvate and oxaloacetate, respectively. There does not appear to be substantial pyruvate production from these metabolites supplied in the FROMP media, suggesting that oxalate in the media successfully suppresses metabolic flux through the PEP–pyruvate–oxaloacetate node²⁹. As pointed out by Rosen *et al*,¹¹ the high degree of scrambling observed in pyruvate may be due to the enzymatic activity of alanine aminotransferase, which catalyzes the interconversion of pyruvate and alanine, introducing solvent deuterons to the methyl group in the process^{42–44}.

The hydrogen atoms in the Ile- δ 1 and Thr- γ 2 methyl groups are not derived from pyruvate and display a different isotopomer distribution from the methyl groups of Val and Leu. The Ile- δ 1 methyl group is derived from Thr- γ 2, and these two methyl groups show almost identical isotopomer distributions (Table 2.2).

When ^1H -glucose is utilized as carbon source in D_2O media, much higher degrees of deuteration are obtained compared to ^1H -pyruvate, since all pyruvate for biosynthesis would have to be derived from phosphoenolpyruvate, oxaloacetate, or malate. The levels of methyl group deuteration shown in Table 2.2 are far greater than those obtained by Otten *et al*³⁸, due to a larger estimation of the population of “invisible” CD_3 isotopomers, ~50% versus ~0%. Aside from this discrepancy, both studies agree with a roughly 1:1 ratio of $\text{CH}_2\text{D}:\text{CHD}_2$ isotopomers.

Table 2.2. Average percent isotopomers in methyl-containing residues of Pin1 WW domain grown in glucose/FROMP and D₂O

	Ala (3)*-β	Val- γ2	Val- γ1	Leu- δ1	Leu- δ2	Ile-δ1	Ile-γ2	Met (2)*- ε	Thr-γ2
CH ₃	3/17	4/43	5/50	4/52	0/63	0/0	6/29	0/2	0/4
CH ₂ D	15/17	18/27	26/28	21/31	20/30	5/8	21/23	7/8	6/10
CHD ₂	21/24	18/22	26/26	24/26	23/17	26/31	18/23	7/7	31/37
CD ₃	61/42	60/8	43/0	51/0	57/0	69/61	55/25	86/83	63/49

* Average of three alanines and two methionines, all other values are derived from single residues.

Phosphoglycerate family (serine, cysteine, tryptophan, and glycine)

Serine is a major metabolic precursor in the biosynthesis of cysteine, tryptophan, and glycine. The human Pin1 WW domain possesses no cysteine residue. All alpha hydrogens exchange with solvent in a D₂O-based growth, so glycine is entirely deuterated. Serine and tryptophan have similarly high levels of protonation at the H^β position in Pin1(glucose), with 80-90% CH₂ isotopomer. These hydrogen atoms are derived from the 3-position of 3-phosphoglycerate in the glycolytic pathway (Figure 2.6 and Figure 2.2). It is somewhat surprising that the protonation is so markedly greater than 50%, since one might expect the 3-position of 3-phosphoglycerate to have an isotopomeric ratio CH₂:CHD of 1:1, with the two isotopomers derived from the glycolytic catabolism of a single glucose molecule. However, glycolysis yields 100% CH₂ isotopomer, because the enzyme, phosphohexose isomerase (which catalyzes the conversion of glucose-6-phosphate to fructose-6-phosphate), transfers a proton from the 2-position of glucose to the 1-position of fructose, instead of deriving a deuteron from solvent⁴⁵. Thus, glycolysis in D₂O yields two identical molecules of 3-¹H-3-phosphoglycerate from a single molecule of ¹H-glucose. Rosen *et al.* also observed very high levels of protonation at H^β for Ser, Trp, and Cys grown with glucose as a carbon source, but a much lesser degree of protonation when ¹H-pyruvate was used.¹¹ In contrast, a high proportion of Ser and Trp CH₂ isotopomer was maintained in Pin1(FROMP) Table 2.1, suggesting that most of the 3-phosphoglycerate used in the biosynthesis of serine was derived from the catabolism of rhamnose and not from gluconeogenic precursors derived from pyruvate/oxaloacetate. This result is likely due to the availability of rhamnose as a carbon source, as well as the presence of

oxalate in FROMP media as a tight inhibitor of PEP carboxykinase and PEP synthetase, limiting the production of phosphoenolpyruvate from oxaloacetate and pyruvate, respectively.

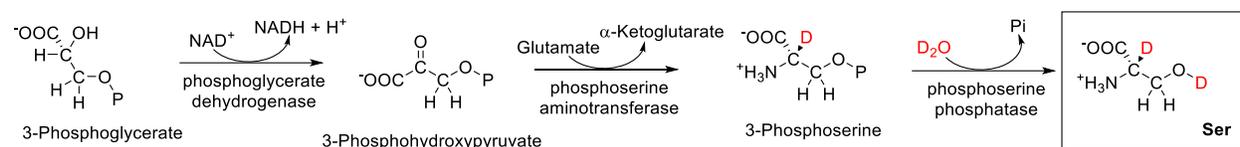


Figure 2.6. Biosynthesis of serine from 3-phosphoglycerate.

Oxaloacetate family (aspartate, asparagine, methionine, threonine, lysine, isoleucine)

The amino acids asparagine, methionine, lysine, threonine, and isoleucine (C^α , $\text{C}^{\gamma 1}$ and $\text{C}^{\delta 1}$) are derived from aspartate, formed in a transamination reaction involving oxaloacetate. Oxaloacetate is generated as a TCA cycle intermediate from the oxidation of (L)-malate, which in turn is generated by the addition of a water molecule across the double bond of fumarate. This reaction is catalyzed by the enzyme fumarase, and when the reaction takes place in D_2O , the result is (2S,3R)-3- ^2H -malate (Figure 2.3)³³. We hypothesized that using fumarate as a carbon source would result in stereospecific incorporation of ^1H at the $\text{H}^{\beta 2}$ position and ^2H at the $\text{H}^{\beta 3}$ position of aspartate. In fact, this is what was observed, with a strong ^1H signal at the $\text{H}^{\beta 2}$ position and no detectable ^1H signal at the $\text{H}^{\beta 3}$ position in Pin1(FROMP) Figure 2.4c, as opposed to virtually no detectable signal at either position in Pin1(glucose) Figure 2.4b. Based on signal intensities, one would estimate 160% ^1H incorporation into the $\text{H}^{\beta 2}$ position of Asp/Asn Pin1(FROMP) Table 2.1, which is clearly an overestimate. The signal is more intense than that observed in Pin1(unlabeled) because the transverse relaxation of the CHD group in Pin1(FROMP) is much slower than that of the CH_2 group in Pin1(unlabeled). In contrast, Asp16 is part of the mobile N-terminal tail of the protein, so its signals were twice as intense as those originating from other Asp/Asn residues. For Asp16, the observed signal intensity ratio was closer to 100% rather than 160%, because relaxation differences are less of a factor in determining signal intensity.

The high level of stereospecific ^1H incorporation at the $\text{H}^{\beta 2}$ position of Asp/Asn implies that the ^1H atom originating on fumarate remained attached through its enzymatic conversion to malate, oxaloacetate, aspartate, and asparagine. In contrast, when pyruvate is converted to

alanine by alanine aminotransferase, methyl protons are readily replaced by solvent deuterons,^{42–44} explaining why protonation of the Ala methyl is substantially lower than that of Leu and Val methyls, despite the fact that all are derived from pyruvate. Thus, aspartate aminotransferase can catalyze the conversion from oxaloacetate to aspartate without perturbing hydrogens at the β position, in contrast to alanine aminotransferase, which tends to exchange them with solvent.

A prerequisite for the stereospecific labeling of $^1\text{H}^{\beta 2}$ and $^2\text{H}^{\beta 3}$ in Asp/Asn is that the oxaloacetate precursor is derived entirely from the fumarate supplied in FROMP media. Oxaloacetate is produced from succinate through the activity of succinate dehydrogenase, which is inhibited by malonate. Oxaloacetate is also produced by anaplerotic reactions, most notably the carboxylation of phosphoenolpyruvate by PEP carboxylase, a reaction inhibited both by oxalate and TCA cycle intermediates^{46,47}. Hence, the FROMP media used in the current study was effective at preventing isotope scrambling related to the key PEP-oxaloacetate-pyruvate node²⁹.

Lys, Thr, and Met are also derived from aspartate. The biosynthesis of these amino acids requires reduction of the $\text{C}\gamma$ carboxyl of Asp, consuming two reducing equivalents from NADPH. The $\text{C}\gamma$ positions of Lys, Thr, and Met are partially protonated, with the fully deuterated isotopomer predominating (>50%) in both Pin1(glucose) and Pin1(FROMP) Table 2.1. This incorporation indicates that a small proportion of the hydride in NADPH is ^1H , rather than ^2H . In *E. coli*, NADP^+ is converted to NADPH through the activities of glucose-6-dehydrogenase and 6-phosphogluconate dehydrogenase in the pentose phosphate pathway, as well as isocitrate dehydrogenase in the TCA cycle⁴⁸. NADPH is required in the biosynthesis of several amino acids, so the $^1\text{H}:^2\text{H}$ ratio of NADPH will impact isotopic labeling at many different sites, resulting in isotopic mixtures. Improving labelling homogeneity will require either finding a way to control the $^1\text{H}:^2\text{H}$ ratio of NADPH or bypassing NADPH-dependent synthetic steps with downstream metabolites.

In the biosynthesis of lysine (Figure 2.7), pyruvate is added to the skeleton of aspartate- β -semialdehyde, accounting for the high level of protonation observed at the H^δ position in Pin1(FROMP). One intermediate in the biosynthetic pathway, L,L- α,ϵ -diaminopimelate, is a symmetric molecule (Figure 2.7), so that the β and δ positions are interchangeable, as are the

(deuterated) α and ϵ positions. Thus, the aspartate- β -semialdehyde component contributes equally to both β and δ positions, and the pyruvate component contributes equally to these positions as well. In Pin1(FROMP), the H^{β^2} position has an apparent protonation of 90% (although given the relaxation effects observed in Asp/Asn, the actual percentage may be closer to 60%), while the H^{β^3} position is entirely deuterated. Therefore, the stereospecificity of protonation introduced by the aspartate- β -semialdehyde precursor is the same as that introduced by the pyruvate precursor (coincidentally). Given the symmetry of the L,L- α,ϵ -diaminopimelate, the corresponding H^{δ^3} position must be equally protonated and the H^{δ^2} position deuterated. As is usually the case, the Lys sidechain δ hydrogens have chemical shifts that are too overlapped to make this distinction. Thus, it is not possible to make stereospecific assignments based on NMR spectra, yet it is possible to assign the stereochemistry based on symmetry considerations in the biosynthesis of lysine.

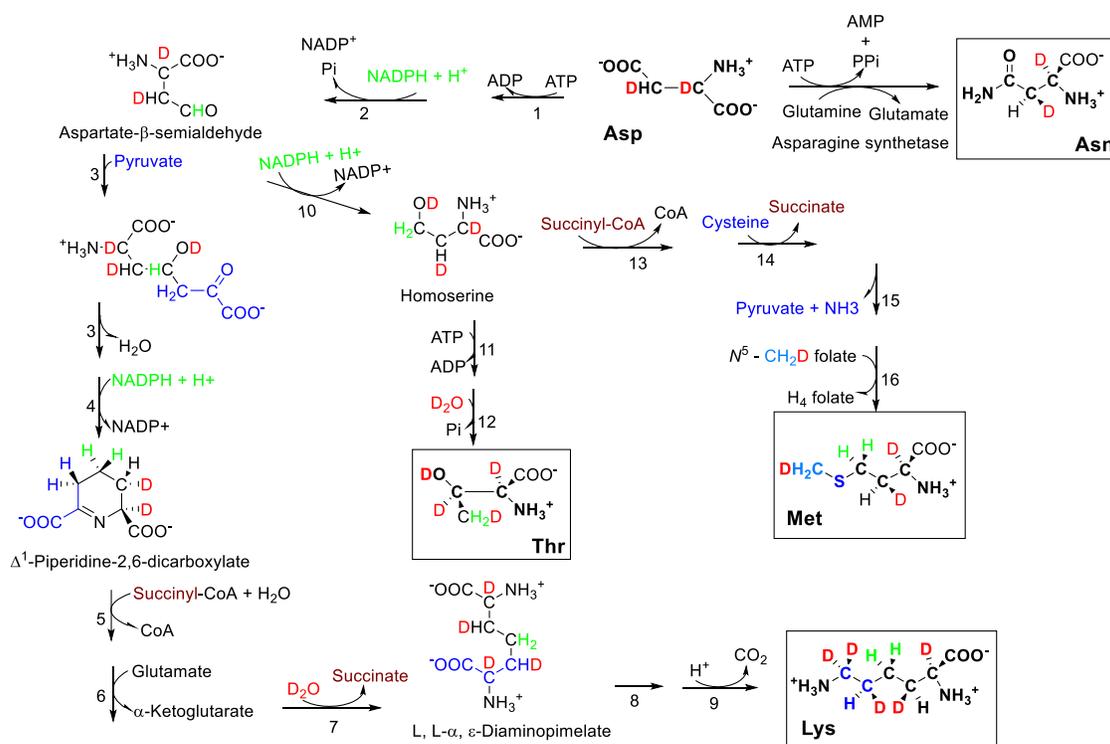


Figure 2.7. Pathways for the biosynthesis of asparagine, lysine, threonine, and methionine. Atoms derived from solvent (red), protons derived from NADPH (green), protons derived from pyruvate (blue), protons derived from methyltetrahydrofolate (light blue), sulfur derived from cysteine in methionine biosynthesis (blue). The pathway enzymes are: 1. Aspartokinase 2. Aspartate- β -semialdehyde dehydrogenase (entry point of a proton to H^γ of all amino acids

belonging to the oxaloacetate family) 3. Dihydropicolinate synthase 4. Δ^1 -piperidine-2,6-dicarboxylate dehydrogenase (entry point of a proton to the H $^\gamma$ of lysine) 5. *N*-succinyl-2-amino-6-ketopimelate synthase 6. Succinyl diaminopimelate aminotransferase (PLP enzyme) 7. Succinyl diaminopimelate desuccinylase 8. Diaminopimelate epimerase 9. Diaminopimelate decarboxylase 10. Homoserine dehydrogenase (entry point of proton to the H $^\gamma$ of threonine and methionine) 11. Homoserine kinase 12. Threonine synthase (PLP dependent enzyme) 13. Homoserine acyl transferase 14. Cystathionine- γ -synthase 15. Cystathionine- β -lyase 16. Methionine synthase.

Aspartate- β -semialdehyde is also converted into homoserine (Figure 2.7), a common branch point for the synthesis of threonine and methionine. Isomerization of homoserine to threonine occurs through a phosphorylated intermediate that displaces a hydrogen atom from C β . Thr H $^\beta$ retained around 27 % protonation in Pin1(glucose) Table 2.1, similar to the findings of Otten *et al*³⁸. However, in Pin1(FROMP), it is the H $^{\beta 3}$ from homoserine that is retained (unfortunately), so that the H $^\beta$ is completely deuterated at this position. Homoserine is also converted into homocysteine as a synthetic precursor to methionine. Homocysteine is then methylated at its sulfur atom by 5-methyltetrahydrofolate, producing methionine and tetrahydrofolate. 5-methyltetrahydrofolate can be regenerated either through the catabolism of serine or glycine. Given the very high level of deuteration observed at the methionine ϵ CH $_3$ group of Pin1(glucose) and Pin1(FROMP) (see Table 2.2), it would appear that glycine is the dominant source of single carbons in the folate cycle. The H $^{\beta 2}$ position of Met is substantially less protonated (30%) than the corresponding position in Asp/Asn/Lys, presumably as a result of one of the many enzymatic steps leading from aspartate- β -semialdehyde to methionine.

α -ketoglutarate family (glutamate, glutamine, proline, arginine)

α -ketoglutarate is a TCA cycle intermediate that gives rise to glutamate, glutamine, proline and arginine. As shown in Figure 2.8, α -ketoglutarate formed in the TCA cycle is deuterated at the H $^\beta$ positions because it originates from the oxaloacetate ketone carbon, which condenses with acetyl-CoA and subsequently acquires 2 H from D $_2$ O solvent. However, the C $^\gamma$ of α -ketoglutarate are derived from the methyl group of pyruvate, hence Glu, Gln, Arg, and Pro are highly protonated at H $^\gamma$ in Pin1(FROMP) and to a much lesser degree in Pin1(glucose) Table 2.1. The relative isotopomeric populations are consistent with those observed in the methyl groups of Val and Leu, which are also derived from pyruvate.

The H^δ hydrogens in proline are derived from NADPH or NADH. In the biosynthesis of proline, glutamate is phosphorylated, followed by the conversion of γ -glutamyl phosphate into glutamate γ -semialdehyde, resulting in the incorporation of a hydrogen at the H^δ of proline from NADH (Figure 2.8). Δ^1 -Pyrroline-5-carboxylate, the product of non-enzymatic cyclization of glutamate γ -semialdehyde, undergoes a reduction reaction with NADH or NADPH providing a second hydrogen at H^δ. There appeared to be some stereospecificity to the low protonation pattern at the proline H^δ position, but we were unable to resolve this further.

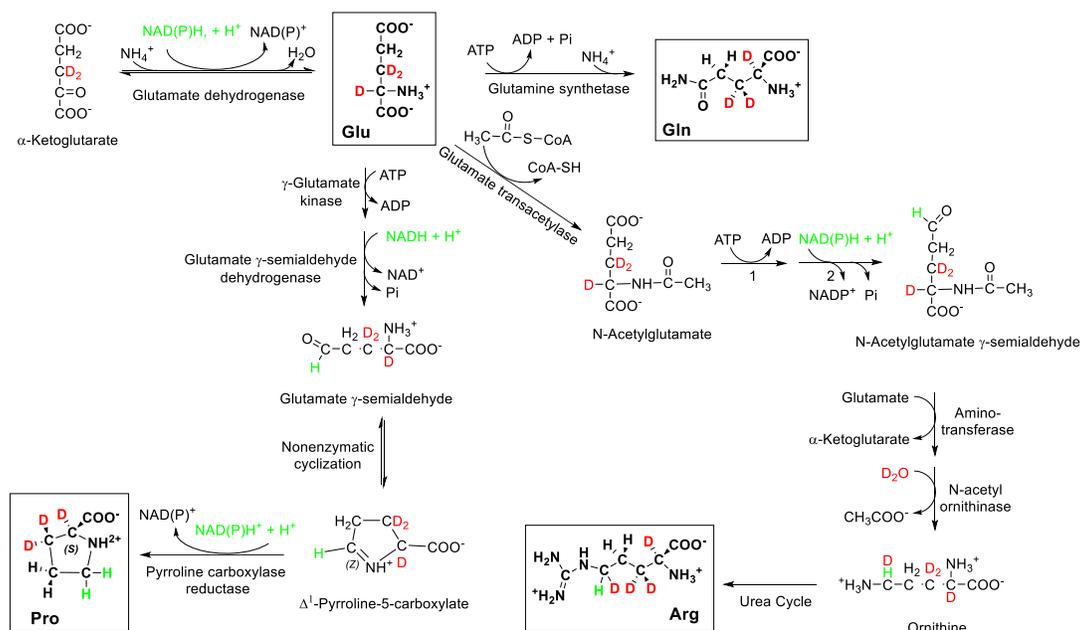


Figure 2.8. Biosynthesis of glutamate, glutamine, proline and arginine from α -ketoglutarate. Protons incorporated from NADPH are shown in green. NADPH supplies protons to the H^δ positions of proline and arginine. Atoms incorporated from solvent are shown in red. Pathway enzymes labeled 1 and 2 are *N*-acetylglutamate kinase and *N*-acetylglutamate dehydrogenase respectively.

In arginine, the reduction at the H^δ position requires the conversion of *N*-acetyl- γ -glutamyl phosphate into *N*-acetylglutamate γ -semialdehyde using NADH or NADPH as cofactors (Figure 2.8). The semialdehyde is then converted to *N*-acetylornithine through a transamination reaction that converts glutamate into α -ketoglutarate. Ornithine is subsequently converted into arginine via the urea cycle. Thus, the H^δ position of arginine acquires one H atom from NADH/NADPH and one solvent deuteron from a transamination reaction. We have no

explanation why the protonation level observed at H^δ of arginine was higher in Pin1(glucose) than Pin1(FROMP) Table 2.1. It is possible that there are cytosolic pools of NADPH with varying ratios of $^1H:^2H$ depending on the local enzymes that regenerate it.

Phosphoenolpyruvate + erythrose-4-phosphate family (tryptophan, phenylalanine, tyrosine)

Erythrose-4-phosphate and two molecules of phosphoenolpyruvate (PEP) are the precursors for chorismate, the common branch point leading to the synthesis of the aromatic amino acids, phenylalanine, tyrosine, and tryptophan.

The six-member ring in the tryptophan indole group is derived from chorismate. The 1- (aldehyde), 2-, 3-, and 4-positions of erythrose-4-phosphate become the C^{ζ^2} , C^{η^2} , C^{ζ^3} , and C^{ϵ^3} positions of Trp. H^{ζ^2} and H^{η^2} are partially protonated in Pin1(glucose) but entirely deuterated in Pin1(FROMP) (Figure 2.9). In contrast, the H^{ϵ^3} position in Trp is derived from the 4-position of erythrose-4-phosphate, and it is ~100% protonated in both Pin1(glucose) and Pin1(FROMP). The H^{ζ^3} position of Trp is derived from NADPH, so it would be expected to be partially protonated and partially deuterated in both Pin1(glucose) and Pin1(FROMP). However, the protonation level was higher in Pin1(FROMP) for unknown reasons. The five-member ring of the Trp indole group is derived from phosphoribose (Figure 2.10). The Trp H^{δ^1} position is derived from position 2 of phosphoribose and is highly protonated, although less so in Pin1(FROMP) than in Pin1(glucose) (Figure 2.9).

The labeling pattern observed at the H^δ and H^ϵ of both phenylalanine and tyrosine residues is similar in both Pin1(glucose) and Pin1(FROMP) Figure 2.9. The H^{δ^1} position of phenylalanine and tyrosine is derived from the methylene group of PEP, while the 4-position of erythrose-4-phosphate is the source of protons at H^{δ^2} , accounting for the high level of protonation observed at the H^δ site (the H^{δ^1} and H^{δ^2} positions are interchangeable) Figure 2.11. The H^{ϵ^1} and H^{ϵ^2} positions of Phe/Tyr have the same biosynthetic origin as the H^{ζ^3} and H^{ζ^2} positions of tryptophan, respectively. Although the H^{ϵ^1} and H^{ϵ^2} positions of Phe/Tyr cannot be distinguished from each other spectroscopically due to ring flips, one would expect levels of protonation similar to their corresponding sites in tryptophan. A higher degree of protonation at the H^ζ position of phenylalanine was observed for Pin1(glucose) versus Pin1(FROMP), which is

consistent with the corresponding $H^{\eta 2}$ position in tryptophan, both derived from the 2-position of erythrose-4-phosphate.

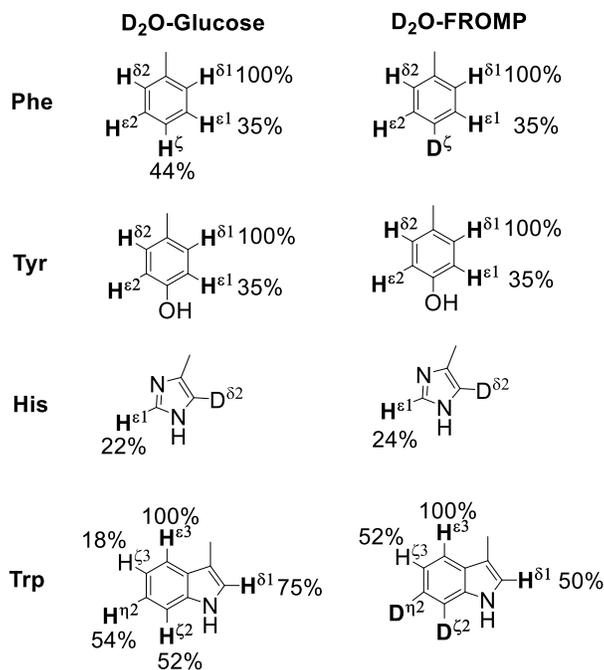


Figure 2.9. Average percent protonation observed at the aromatic sidechains of phenylalanine, tyrosine, tryptophan and histidine for Pin1(glucose) and Pin1(FROMP).

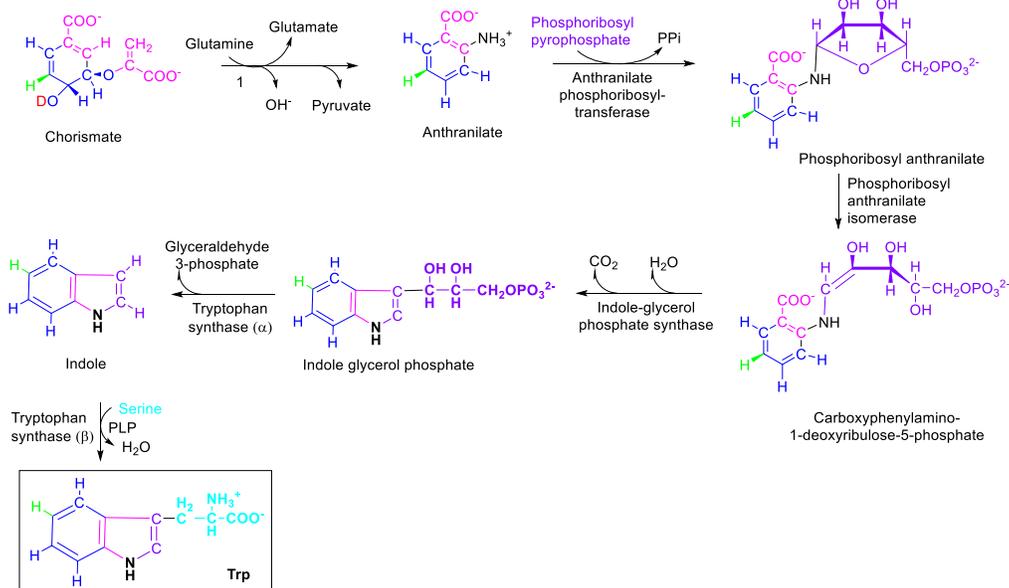


Figure 2.10. Biosynthesis of tryptophan from chorismate. Atoms derived from phosphoenolpyruvate, erythrose-4-phosphate, phosphoribosyl pyrophosphate, NADPH, serine

and solvent (D₂O) are shown in pink, blue, purple, green, light blue, and red, respectively. The pathway enzyme 1 is anthranilate synthase.

The H^β positions of phenylalanine and tyrosine are derived from a molecule of phosphoenolpyruvate, which condenses with shikimate to eventually form chorismate (Figure 2.11). It is unknown why the protonation levels at H^β for Phe and Tyr are so low (30-40%, non-stereospecific) compared with Ser and Trp, despite all being derived from phosphoenolpyruvate. It is possible that exchange with solvent occurs at the transamination step as it does with alanine, but this seems unlikely, given that the aromatic amino acid aminotransferase is similar to aspartate aminotransferase (which leaves H^β protonation intact, as we have noted) in protein structure, substrate specificity, and spectroscopic properties⁴⁹. There are a number of enzymes catalyzing reactions between the condensation of phosphoenolpyruvate with phosphoshikimate and the final transamination reaction at which solvent exchange could occur: 3-enoylpyruvateshikimate-5-phosphate synthase, chorismate synthase, and chorismate mutase (Figure 2.11). Of these, chorismate mutase catalyzes the formation of prephenate from chorismate via a Claisen rearrangement, establishing the C^β-C^γ bond, so it is quite possible that the solvent exchange occurs at this step.

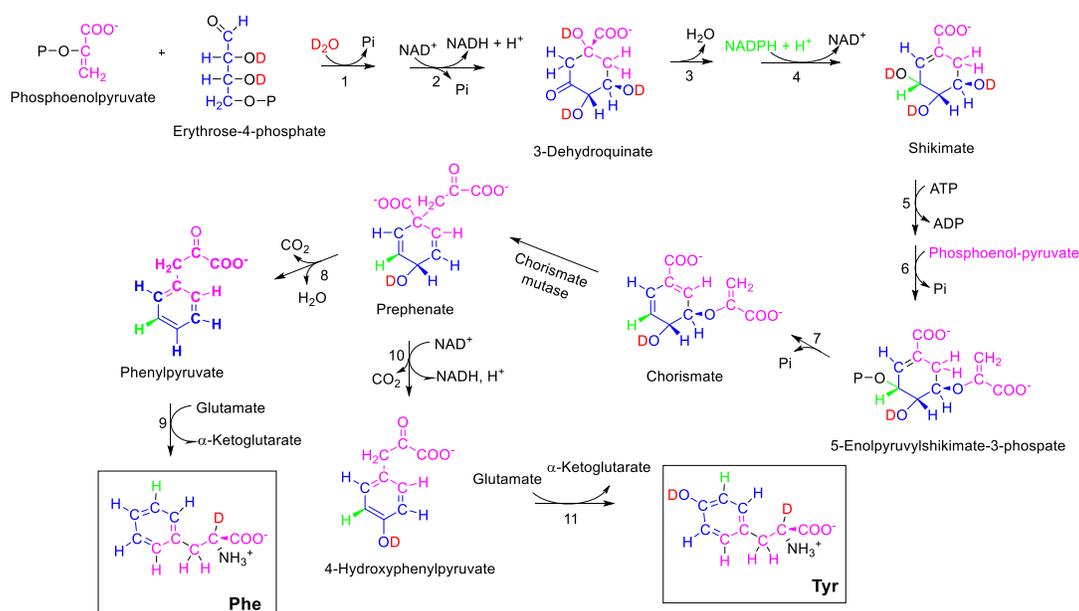


Figure 2.11. Biosynthesis of phenylalanine and tyrosine. Atoms derived from phosphoenolpyruvate, erythrose-4-phosphate, NADPH and solvent (D₂O) are shown in pink, blue, green and red, respectively. The pathway enzymes are: 1. 2-keto-3-deoxy-D-

arabinoheptulosonate-7-phosphate synthase 2. dehydroquinate synthase 3. 3-dehydroquinate dehydratase 4. shikimate dehydrogenase 5. shikimate kinase 6. 3-enoylpyruvateshikimate-5-phosphate synthase 7. chorismate synthase 8. prephenate dehydratase 9. aminotransferase 10. prephenate dehydrogenase 11. aminotransferase.

Ribose phosphate family (histidine)

Histidine is the only amino acid belonging to this family, and it is derived from three precursors. Phosphoribosyl pyrophosphate (PRPP) contributes five carbon atoms, the purine ring of ATP contributes a carbon and a nitrogen, and glutamine contributes the second nitrogen in the imidazole ring (Figure 2.12). The labeling pattern observed in histidine was consistent for both Pin1(glucose) and Pin1(FROMP), producing complete deuteration at H^β and $H^{\delta 2}$. A small degree of protonation at $H^{\epsilon 1}$ (~20%, Figure 2.9) derives from the six-membered ring of adenine, which is contributed by N^{10} -formyl-tetrahydrofolate.

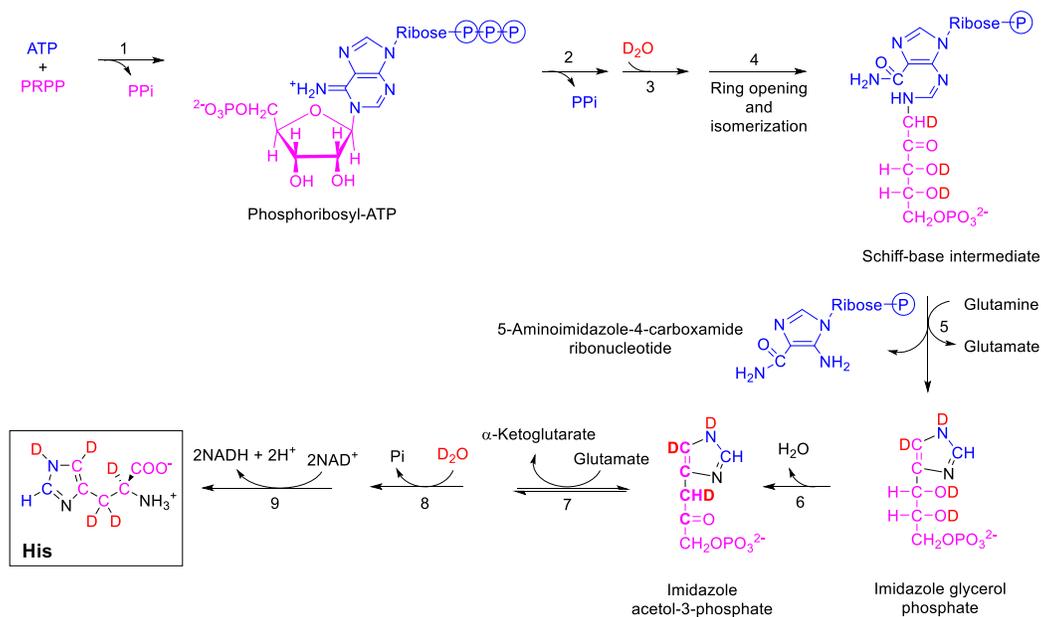


Figure 2.12. Outline of the biosynthesis of histidine. Atoms derived from PRPP and ATP are shaded in pink and blue, respectively. Deuterium atoms derived from solvent are shown in red. The pathway enzymes are: 1. ATP phosphoribosyl transferase 2. pyrophosphohydrolase 3. phosphoribosyl-AMP cyclohydrolase 4. phosphoribosyl-formimino-5-aminoimidazole-4-carboxamide ribonucleotide isomerase 5. glutamine amidotransferase 6. imidazole glycerol-3-phosphate dehydratase 7. L-histidinol phosphate aminotransferase 8. histidinol phosphate phosphatase 9. histidinol dehydrogenase.

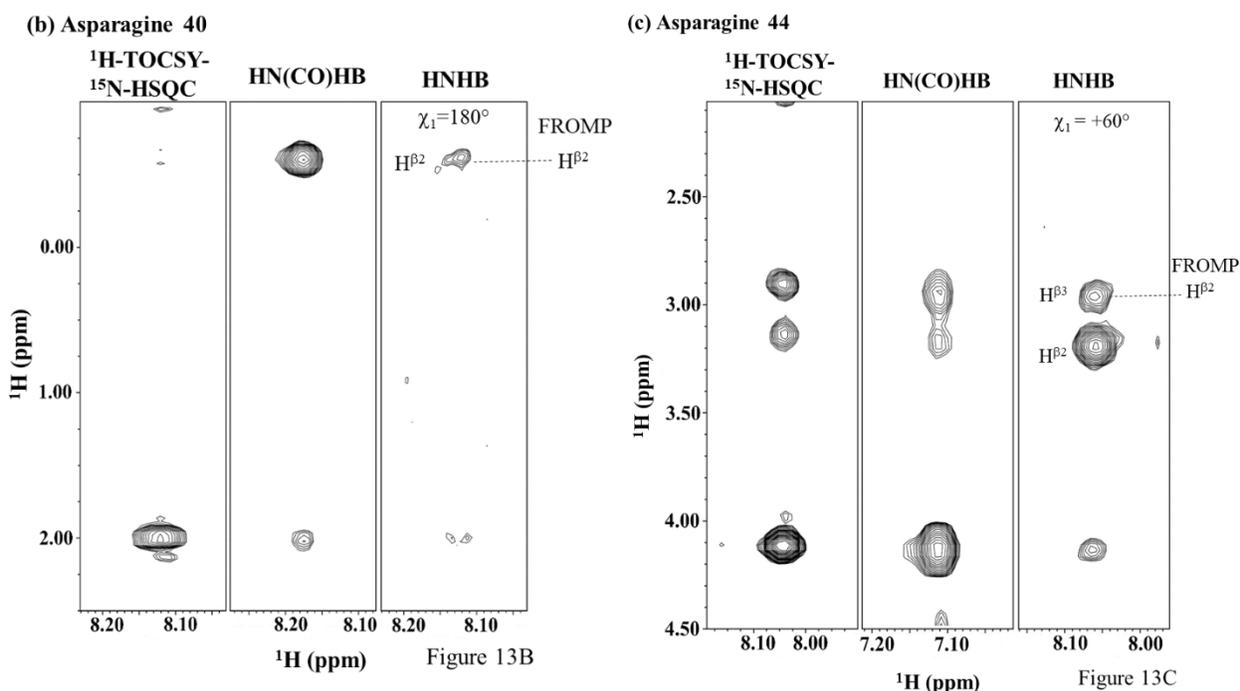
Application of FROMP labeling to χ_1 dihedral angle determination

The precision and accuracy of NMR-derived protein structures could be significantly improved with stereospecific assignments of β -methylene protons and χ_1 dihedral angle restraints⁵⁰. The utility of three-bond J coupling for determining backbone and side chain dihedral angles has long been established,^{51–53} and 3D experiments have been developed to measure these. One such experiment is the 3D HNHB, which correlates amide proton and nitrogen resonances with intra-residue H^β resonances, providing semi-quantitative information on the size of ${}^3J(\text{NH}^\beta)$ Figure 2.13b and 13c⁵¹. An analogous experiment is the HN(CO)HB, which transfers magnetization via the 3J -coupling between carbonyl carbon and H^β protons⁵⁴. ${}^3J_{\alpha/\beta}$ can also be estimated using relative peak intensities in a 3D ${}^1\text{H}$ -TOCSY- ${}^{15}\text{N}$ -HSQC⁵⁰. Together, these three experiments provide redundant information about the χ_1 dihedral angle, along with stereospecific assignment of the β -methylene protons (Figure 2.13a).

Figures 2.13b and 2.13c show strip plots of the HNHB, HN(CO)HB, and ${}^1\text{H}$ -TOCSY- ${}^{15}\text{N}$ -HSQC spectra for human Pin1 Asn40 and Asn44, respectively. Asn40 is found in the structured region of the second β -strand of Pin1 (Figure 2.1), and the crystal structure (1ZCN) shows that it adopts a “trans” (180°) χ_1 dihedral angle. All of the experiments in Figure 2.13b are consistent with the “trans” conformation, also allowing the stereospecific assignment of the upfield resonance to $H^{\beta 2}$. The application of FROMP labeling technique confirmed this assignment (Figure 2.4c).

Asn44 is found in the loop region connecting the second and third β -strands of human Pin1 WW domain (Figure 2.1). The χ_1 dihedral angle of this residue is gauche⁺ (-60°) from the crystal structure. However, the conformation predicted using 3J -coupling measurements was gauche⁻ ($+60^\circ$) (Figure 2.13c), with the upfield resonance assigned as $H^{\beta 3}$. However, stereospecific labeling with FROMP media unambiguously indicated the upfield resonance as $H^{\beta 2}$. These incongruous results are most readily explained by internal motions, more specifically, averaging between two χ_1 dihedral angle conformers, trans and gauche⁺. This is more in keeping with the gauche⁺ conformation observed in the crystal structure. Consistent with motional averaging, the difference between strong and weak signals in the 3J -coupling experiments was less pronounced in Asn44 than in Asn40. Thus, the semi-quantitative assessment of 3J -couplings was unable to distinguish between a pure gauche⁻ conformation and rotameric averaging between

gauche⁺ and trans conformations, requiring the correct stereospecific chemical shift assignments



to make the distinction.

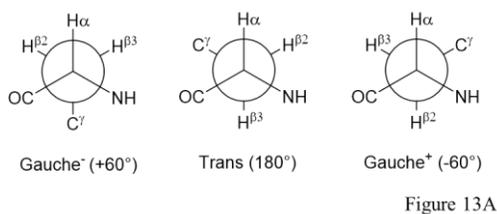


Figure 2.13. Application of FROMP labeling technique for stereospecific assignment of β -methylene protons. (A) Rotamers showing gauche⁻ (+60°), trans (180°), and gauche⁺ (-60°) χ_1 dihedral angles. Strip plots from 3D HNHB, HN(CO)HB, and ¹H-TOCSY-¹⁵N-HSQC spectra for (B) Asn 40 and (C) Asn 44 of the mutant Pin1 WW domain. 3D HNHB and ¹H-TOCSY-¹⁵N-HSQC strip plots are taken at the amide frequency of Asn 40 and Asn 44, while HN(CO)HB strip plots are taken at the amide frequency of His 41 and Ala 45. The intensity of each correlation is related to the size of each 3-bond J-coupling.

Asp18 and Lys21 are found in the flexible N-terminal region of Pin1, and both residues are absent in the crystal structure. The H ^{β} signals from these residues were too overlapped in 3D-

NMR experiments to determine a χ_1 dihedral angle. However, using FROMP labeling, we could still obtain stereospecific assignments for these residues.

The residue Asp33 was absent in the crystal structure because it was mutated in our Pin1 construct²¹. We could not detect Asp33 H $^\beta$ signals in two of the three χ_1 dihedral angle experiments due to the exchange broadening of Asp33 NH. However, our isotope labelling technique could still obtain stereospecific assignments for this residue.

Lys28 and Met30 are found in the first β -strand of Pin1. Both residues adopt trans (180°) χ_1 dihedral angles in the crystal structure. However, the conformation predicted through 3J -coupling measurement was gauche⁻(+60°), with the assignment of the downfield H $^\beta$ signals to H $^{\beta 2}$ for both residues. However, stereo-selective labeling using FROMP media assigned the upfield signals to H $^{\beta 2}$, meaning that the χ_1 dihedral angle assignment was incorrect, similar to Asn44. The χ_1 dihedral angles of Lys28 and Met30 are better described as a rotameric average between gauche⁺ and trans conformations, which is again more consistent with the X-ray crystal structure.

In high resolution X-ray crystal structures, a small percentage of residues can be seen to adopt multiple χ_1 dihedral angles. In NMR studies of residual dipolar couplings in multiple alignment media, up to 50% of residues adopt multiple χ_1 dihedral angles^{55,56}. Our limited data on Asp/Asn/Lys/Met residues seem to support the notion that multiple side chain conformations are common in solution. Our analysis of Lys28, Met30, and Asn44 suggests that using 3J -coupling measurement experiments alone to determine stereospecific chemical shift assignments and χ_1 dihedral angles is unreliable when rotameric averaging is present. Therefore, it is anticipated that stereo-selective labeling will improve the accuracy of NMR-based χ_1 dihedral angle determinations.

Application of FROMP labeling technique for structure determination of large proteins

Structure determination of larger proteins relies mainly on deuteration to circumvent the effects associated with transverse relaxation causing decreased sensitivity and resolution^{4,8}. However, deuteration reduces the number of protons in protein eliminating long-range NOESY information from sidechains, except for selectively protonated side-chain moieties¹¹. Almost all

previously developed methods have focused on the production of highly deuterated proteins with selective ^1H - ^{13}C labeling at methyl groups of alanine, valine, leucine and isoleucine^{3,38} which are readily observable in large systems by solution NMR. Structures of larger proteins are then solved entirely based on NH-NH, NH-methyl and methyl-methyl NOEs^{6,8}. The difficulty in determining accurate protein structures due to limited side chain information halts the routine application of solution NMR to larger protein systems⁵⁷.

Our labeling strategy provides a complimentary approach that provides additional information about Asp, Asn, Lys, and Met (H β s) amino acid sidechains, which have been less studied in large systems because of their lack of methyl groups. Our approach provides highly specific (^1H - ^{12}C) labeling in an otherwise highly deuterated background for 9 out of the 20 amino acids. When combined with methyl-specific protonation of aliphatic amino acids, it will now be possible to obtain NOE-based distance restraints for most protein residues. We anticipate that our approach can eventually be extended to produce stereo-selective isotope labeling in all 20 amino acids, facilitating NMR-based structure determinations of high molecular weight protein systems.

Conclusions

In this study, we present the WW domain of human Pin1 as a model system for assessing isotope labeling via the amino acid biosynthetic pathways of *E. coli* bacteria. Our proposed FROMP media (fumarate, rhamnose, oxalate, malonate, and pyruvate) for *E. coli* is optimal for the stereospecific deuteration of aspartate, asparagine, and lysine, but improvements are clearly needed for all other amino acids. The development of easy and cost-effective strategies for the homogeneous and stereospecific incorporation of deuterons into proteins will be important not only for structural studies of high molecular weight proteins systems, but also for the accurate determination of χ_1 side chain dihedral angles in any protein system.

Acknowledgements

This work was supported by a Discovery Grant, RGPIN-2015- 06664, from NSERC (Natural Sciences and Engineering Research Council of Canada), startup funds from the Department of Medicine and Faculty of Medicine & Dentistry at the University of Alberta, and private funding

from the Hwang Professional Corporation. We would like to acknowledge the anonymous NSERC reviewer who recommended WW domains as a potential model system, as well as Monica Li, for her assistance with acquiring NMR spectra.

References

- (1) PDB Statistics <https://www.rcsb.org/stats>.
- (2) Clore, G. M.; Gronenborn, A. M. NMR Structure Determination of Proteins and Protein Complexes Larger than 20 KDa. *Curr. Opin. Chem. Biol.* **1998**, *2* (5), 564–570. [https://doi.org/10.1016/S1367-5931\(98\)80084-7](https://doi.org/10.1016/S1367-5931(98)80084-7).
- (3) Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, *1* (2), 749–754. <https://doi.org/10.1038/nprot.2006.101>.
- (4) LeMaster, D. M.; Richards, F. M. NMR Sequential Assignment of Escherichia Coli Thioredoxin Utilizing Random Fractional Deuteriation. *Biochemistry* **1988**, *27* (1), 142–150. <https://doi.org/10.1021/bi00401a022>.
- (5) Farmer, B. T.; Venters, R. A. NMR of Perdeuterated Large Proteins. **2002**, *16*, 75–120. https://doi.org/10.1007/0-306-47083-7_3.
- (6) Gardner, K. H.; Kay, L. E. The Use of ²H, ¹³C, ¹⁵N Multidimensional NMR to Study the Structure and Dynamics of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357–406. <https://doi.org/10.1146/annurev.biophys.27.1.357>.
- (7) Goto, N. K.; Kay, L. E. New Developments in Isotope Labeling Strategies for Protein Solution NMR Spectroscopy. *Curr. Opin. Struct. Biol.* **2000**, *10* (5), 585–592. [https://doi.org/10.1016/S0959-440X\(00\)00135-4](https://doi.org/10.1016/S0959-440X(00)00135-4).
- (8) Gardner, K. H.; Rosen, M. K.; Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR. *Biochemistry* **1997**, *36* (6), 1389–1401. <https://doi.org/10.1021/bi9624806>.
- (9) LeMaster, D. M. Chiral β and Random Fractional Deuteration for the Determination of Protein Sidechain Conformation by NMR. *FEBS Lett.* **1987**, *223* (1), 191–196. [https://doi.org/10.1016/0014-5793\(87\)80534-3](https://doi.org/10.1016/0014-5793(87)80534-3).
- (10) Gardner, K. H.; Kay, L. E. Production and Incorporation of ¹⁵N, ¹³C, ²H (1H- Δ 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **1997**, *119* (32), 7599–7600. https://doi.org/10.1021/JA9706514/SUPPL_FILE/JA7599.PDF.
- (11) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *Journal of Molecular Biology.* 1996. <https://doi.org/10.1006/jmbi.1996.0603>.
- (12) 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 (δ 1). *J. Biomol. NMR* **1999**, No. 13, 369–374. <https://doi.org/10.1023/a:1008393201236>.
- (13) Ayala, I.; Sounier, R.; Usé, N.; Gans, P.; Boisbouvier, J. An Efficient Protocol for the Complete Incorporation of Methyl-Protonated Alanine in Perdeuterated Protein. *J. Biomol. NMR* **2009**, *43* (2), 111–119. <https://doi.org/10.1007/s10858-008-9294-7>.
- (14) Tugarinov, V.; Kay, L. E. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *J. Biomol. NMR* **2004**, *28* (2), 165–172.

- <https://doi.org/10.1023/B:JNMR.0000013824.93994.1f>.
- (15) Fischer, M.; Kloiber, K.; Häusler, J.; Ledolter, K.; Konrat, R.; Schmid, W. Synthesis of a ¹³C-Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. *ChemBioChem* **2007**, *8* (6), 610–612. <https://doi.org/10.1002/cbic.200600551>.
 - (16) Velyvis, A.; Ruschak, A. M.; Kay, L. E. An Economical Method for Production of ²H,¹³CH₃-Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 KDa Proteasome. *PLoS One* **2012**, *7* (9), 1–8. <https://doi.org/10.1371/journal.pone.0043725>.
 - (17) Gans, P.; Hamelin, O.; Sounier, R.; Ayala, I.; Durá, M. A.; Amero, C. D.; Noirclerc-Savoie, M.; Franzetti, B.; Plevin, M. J.; Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angew. Chemie - Int. Ed.* **2010**, *49* (11), 1958–1962. <https://doi.org/10.1002/anie.200905660>.
 - (18) Vuister, G. W.; Kim, S. J.; Wu, C.; Bax, A. 2D and 3D NMR Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labeling. *J. Am. Chem. Soc.* **1994**, *116* (20), 9206–9210. <https://doi.org/10.1021/ja00099a041>.
 - (19) Aghazadeh, B.; Zhu, K.; Kubiseski, T. J.; Liu, G. A.; Pawson, T.; Zheng, Y.; Rosen, M. K. Structure and Mutagenesis of the Dbl Homology Domain. *Nat. Struct. Biol.* **1998**, *5* (12), 1098–1107. <https://doi.org/10.1038/4209>.
 - (20) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal Isotope Labelling for NMR Protein Structure Determinations. *Nature* **2006**, *440* (7080), 52–57. <https://doi.org/10.1038/nature04525>.
 - (21) Jäger, 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. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (28), 10648–10653. <https://doi.org/10.1073/pnas.0600511103>.
 - (22) Verdecia, M. A.; Bowman, M. E.; Ping, K.; Hunter, T.; Noel, J. P. Letters Recognition by Group IV WW Domains. *Nat. Struct. Biol.* **2000**, *7* (8), 639–643.
 - (23) Kowalski, J. A.; Liu, K.; Kelly, J. W. NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the x-Ray Crystal Structures of Pin1. *Biopolymers* **2002**, *63* (2), 111–121. <https://doi.org/10.1002/bip.10020>.
 - (24) Wintjens, R.; Wieruszkeski, J. M.; Drobecq, H.; Rousselot-Pailley, P.; Buée, L.; Lippens, G.; Landrieu, I. ¹H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* **2001**, *276* (27), 25150–25156. <https://doi.org/10.1074/jbc.M010327200>.
 - (25) Cai, F.; Li, M. X.; Pineda-Sanabria, S. E.; Gelozia, S.; Lindert, S.; West, F.; Sykes, B. D.; Hwang, P. M. Structures Reveal Details of Small Molecule Binding to Cardiac Troponin. *J. Mol. Cell. Cardiol.* **2016**, *101*, 134–144. <https://doi.org/10.1016/j.yjmcc.2016.10.016>.
 - (26) Zahran, S.; Pan, J. S.; Liu, P. B.; Hwang, P. M. Combining a PagP Fusion Protein System with Nickel Ion-Catalyzed Cleavage to Produce Intrinsically Disordered Proteins in *E. Coli*. *Protein Expr. Purif.* **2015**, *116*, 133–138. <https://doi.org/10.1016/j.pep.2015.08.018>.
 - (27) Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wüthrich, K. Stereospecific Nuclear Magnetic Resonance Assignments of the Methyl Groups of Valine and Leucine in the DNA-Binding Domain of the 434 Repressor by Biosynthetically Directed Fractional ¹³C

- Labeling. *Biochemistry* **1989**, 28 (19), 7510–7516. <https://doi.org/10.1021/bi00445a003>.
- (28) 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-Labeling and Production of Perdeuterated Proteins in H₂O. *J. Biomol. NMR* **2011**, 50 (1), 35–42. <https://doi.org/10.1007/s10858-011-9477-5>.
- (29) Sauer, U.; Eikmanns, B. J. The PEP-Pyruvate-Oxaloacetate Node as the Switch Point for Carbon Flux Distribution in Bacteria. *FEMS Microbiol. Rev.* **2005**, 29 (4), 765–794. <https://doi.org/10.1016/j.femsre.2004.11.002>.
- (30) Stiffin, R. M.; Sullivan, S. M.; Carlson, G. M.; Holyoak, T. Differential Inhibition of Cytosolic PEPCK by Substrate Analogues. Kinetic and Structural Characterization of Inhibitor Recognition. *Biochemistry* **2008**, 47 (7), 2099–2109. <https://doi.org/10.1021/bi7020662>.
- (31) Yang, Z.; Floyd, D. L.; Loeber, G.; Tong, L. Structure of a Closed Form of Human Malic Enzyme and Implications for Catalytic Mechanism. *Nat. Struct. Biol.* **2000**, 7 (3), 251–257. <https://doi.org/10.1038/73378>.
- (32) Narindrasorasak, S.; Bridger, W. A. Probes of the Structure of Phosphoenolpyruvate Synthetase: Effects of a Transition State Analogue on Enzyme Conformation. *Can. J. Biochem.* **1978**, 56 (8), 816–819. <https://doi.org/10.1139/o78-124>.
- (33) Kasperek, G. J.; Pratt, R. F. The Fumarase Reaction. A Nuclear Magnetic Resonance Experiment for Biological Chemistry Students. *J. Chem. Educ.* **1977**, 54 (8), 515. <https://doi.org/10.1021/ed054p515>.
- (34) THORN, M. B. Inhibition by Malonate of Succinic Dehydrogenase in Heart-Muscle Preparations. *Biochem. J.* **1953**, 54 (4), 540–547. <https://doi.org/10.1042/bj0540540>.
- (35) PARDEE, A. B.; POTTER, V. R. Malonate Inhibition of Oxidations in the Krebs Tricarboxylic Acid Cycle. *J. Biol. Chem.* **1949**, 178 (1), 241–250. [https://doi.org/10.1016/s0021-9258\(18\)56954-4](https://doi.org/10.1016/s0021-9258(18)56954-4).
- (36) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, 6 (3), 277–293. <https://doi.org/10.1007/BF00197809>.
- (37) Johnson, B. A.; Blevins, R. A. NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* **1994**, 4 (5), 603–614. <https://doi.org/10.1007/BF00404272>.
- (38) Otten, R.; Chu, B.; Krewulak, K. D.; Vogel, H. J.; Mulder, F. A. A. Comprehensive and Cost-Effective NMR Spectroscopy of Methyl Groups in Large Proteins. *J. Am. Chem. Soc.* **2010**, 132 (9), 2952–2960. <https://doi.org/10.1021/ja907706a>.
- (39) Tugarinov, V.; Kay, L. E. Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins. *ChemBioChem* **2005**, 6 (9), 1567–1577. <https://doi.org/10.1002/cbic.200500110>.
- (40) Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. Methyl-Specific Isotopic Labeling: A Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* **2015**, 32, 113–122. <https://doi.org/10.1016/j.sbi.2015.03.009>.
- (41) Shekhtman, A.; Ghose, R.; Goger, M.; Cowburn, D. NMR Structure Determination and Investigation Using a Reduced Proton (REDPRO) Labeling Strategy for Proteins. *FEBS Lett.* **2002**, 524 (1–3), 177–182. [https://doi.org/10.1016/S0014-5793\(02\)03051-X](https://doi.org/10.1016/S0014-5793(02)03051-X).
- (42) Malaisse, W. J.; Malaisse-Lagae, F.; Liemans, V.; Ottinger, R.; Willem, R.

- Phosphoglucoisomerase-Catalyzed Interconversion of Hexose Phosphates: Isotopic Discrimination between Hydrogen and Deuterium. *Mol. Cell. Biochem.* **1990**, *93* (2), 153–165. <https://doi.org/10.1007/BF00226187>.
- (43) Golichowski, A.; Harruff, R. C.; Jenkins, W. T. The Effects of PH on the Rates of Isotope Exchange Catalyzed by Alanine Aminotransferase. *Arch. Biochem. Biophys.* **1977**, *178* (2), 459–467. [https://doi.org/10.1016/0003-9861\(77\)90216-8](https://doi.org/10.1016/0003-9861(77)90216-8).
- (44) Cooper, A. J. L. Proton Magnetic Resonance Studies of Glutamate Alanine Transaminase Catalyzed Deuterium Exchange. Evidence for Proton Conservation during Prototropic Transfer from the α Carbon of L Alanine to the C4 Position of Pyridoxal 5' Phosphate. *J. Biol. Chem.* **1976**, *251* (4), 1088–1096. [https://doi.org/10.1016/s0021-9258\(17\)33805-x](https://doi.org/10.1016/s0021-9258(17)33805-x).
- (45) Babu, U. M.; Johnston, R. B. Nuclear Magnetic Resonance Studies of D₂O-Substrate Exchange Reactions Catalyzed by Glutamic Pyruvic and Glutamic Oxaloacetic Transaminases. *Biochemistry* **1976**, *15* (25), 5671–5678. <https://doi.org/10.1021/bi00670a037>.
- (46) Mildvan, A. S.; Scrutton, M. C.; Utter, M. F. Pyruvate Carboxylase. *J. Biol. Chem.* **1966**, *241* (15), 3488–3498. [https://doi.org/10.1016/s0021-9258\(18\)99859-5](https://doi.org/10.1016/s0021-9258(18)99859-5).
- (47) Zeczycki, T. N.; St. Maurice, M.; Attwood, P. V. Inhibitors of Pyruvate Carboxylase. *Open Enzym. Inhib. J.* **2010**, *3* (1), 8–26. <https://doi.org/10.2174/1874940201003010008>.
- (48) Spaans, S. K.; Weusthuis, R. A.; van der Oost, J.; Kengen, S. W. M. NADPH-Generating Systems in Bacteria and Archaea. *Front. Microbiol.* **2015**, *6* (JUL), 1–27. <https://doi.org/10.3389/fmicb.2015.00742>.
- (49) Hayashi, H.; Inoue, K.; Nagata, T.; Kuramitsu, S.; Kagamiyama, H. Escherichia Coli Aromatic Amino Acid Aminotransferase: Characterization and Comparison with Aspartate Aminotransferase. *Biochemistry* **1993**, *32* (45), 12229–12239. <https://doi.org/10.1021/bi00096a036>.
- (50) Clore, G. M.; Bax, A.; Gronenborn, A. M. Stereospecific Assignment of B-Methylene Protons in Larger Proteins Using 3D ¹⁵N-Separated Hartmann-Hahn and ¹³C-Separated Rotating Frame Overhauser Spectroscopy. *J. Biomol. NMR* **1991**, *1*, 13–22. <https://doi.org/10.1007/BF01874566>.
- (51) Archer, S. J.; Ikura, M.; Torchia, D. A.; Bax, A. An Alternative 3D NMR Technique for Correlating Backbone ¹⁵N with Side Chain H β Resonances in Larger Proteins. *J. Magn. Reson.* **1991**, *95* (3), 636–641. [https://doi.org/10.1016/0022-2364\(91\)90182-S](https://doi.org/10.1016/0022-2364(91)90182-S).
- (52) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. Measurement of Homo- and Heteronuclear J Couplings from Quantitative J Correlation. *Methods Enzymol.* **1994**, *239* (C), 79–105. [https://doi.org/10.1016/S0076-6879\(94\)39004-5](https://doi.org/10.1016/S0076-6879(94)39004-5).
- (53) Montelione, G. T.; Winkler, M. E.; Rauenbuehler, P.; Wagner, G. Accurate Measurements of Long-Range Heteronuclear Coupling Constants from Homonuclear 2D NMR Spectra of Isotope-Enriched Proteins. *J. Magn. Reson.* **1989**, *82* (1), 198–204. [https://doi.org/10.1016/0022-2364\(89\)90183-2](https://doi.org/10.1016/0022-2364(89)90183-2).
- (54) Grzesiek, S.; Ikura, M.; Marius Clore, G.; Gronenborn, A. M.; Bax, A. A 3D Triple-Resonance NMR Technique for Qualitative Measurement of Carbonyl-H β J Couplings in Isotopically Enriched Proteins. *J. Magn. Reson.* **1992**, *96* (1), 215–221. [https://doi.org/10.1016/0022-2364\(92\)90307-S](https://doi.org/10.1016/0022-2364(92)90307-S).
- (55) Mittermaier, A.; Kay, L. E. X1 Torsion Angle Dynamics in Proteins, From Dipolar

- Couplings. *J. Am. Chem. Soc.* **2001**, *123* (28), 6892–6903. <https://doi.org/10.1021/ja010595d>.
- (56) Li, F.; Grishaev, A.; Ying, J.; Bax, A. Side Chain Conformational Distributions of a Small Protein Derived from Model-Free Analysis of a Large Set of Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, *137* (46), 14798–14811. <https://doi.org/10.1021/jacs.5b10072>.
- (57) Wagner, G. Prospects for NMR of Large Proteins. *J. Biomol. NMR* **1993**, *3* (4), 375–385. <https://doi.org/10.1007/BF00176005>.

Chapter 3

Cost-effective selective deuteration of aromatic amino acid residues produces long-lived solution ^1H -NMR magnetization in proteins

A version of this chapter has been submitted to *Angew. Chem. Int.* to be published as: Danmaliki, G.I., Yu, S., Braun, S., Zhao, Y.Y., Moore, J., Fahlman, R.P., West, F.G., Hwang, P.M. Cost-effective selective deuteration of aromatic amino acid residues produces long-lived solution ^1H -NMR magnetization in proteins. PMH directed the research. SY synthesized the aromatic precursors under FGW supervision. GID expressed and purified the protein constructs, tested many metabolic precursors and inhibitors, acquired, and analyzed the NMR data under PMH supervision. SB and YYZ analyzed the protein samples by mass spectrometry under JM and RPF supervision. GID wrote the first draft of the manuscript under PMH supervision.

Summary

Solution NMR studies of large proteins are hampered by rapid signal decay due to short-range dipolar ^1H - ^1H and ^1H - ^{13}C interactions. These are attenuated by rapid rotation in methyl groups and by deuteration (^2H), so selective ^1H , ^{13}C -isotope labeling of methyl groups in otherwise perdeuterated proteins, combined with methyl transverse relaxation optimized spectroscopy (methyl-TROSY), is now standard for solution NMR of large protein systems >25 kDa. For non-methyl positions, long-lived magnetization can be introduced as isolated ^1H - ^{12}C groups. We have developed a cost-effective chemical synthesis for producing selectively deuterated phenylpyruvate and hydroxyphenylpyruvate. Feeding these amino acid precursors to *E. coli* in D_2O , along with selectively deuterated anthranilate and unlabeled histidine, results in isolated and long-lived ^1H magnetization in the aromatic rings of Phe (HD, HZ), Tyr (HD), Trp (HH2, HE3) and His (HD2 and HE1). In addition, we have refined our previously developed method for stereoselective deuteration of Asp, Asn, and Lys amino acid residues to utilize unlabeled glucose and fumarate as carbon sources. Using additional additives to limit isotope scrambling, we can produce isolated ^1H - ^{12}C groups in Phe, Tyr, Trp, His, Asp, Asn, Lys, Arg, and Pro in a perdeuterated background, compatible with standard ^1H - ^{13}C labeling of methyl groups in Ala, Ile, Leu, Val, Thr, Met. We are thus able to create long-lived ^1H magnetization in most amino acid residues for NMR studies of larger protein systems.

Introduction

The development of multinuclear multi-dimensional solution NMR methods for studying uniformly ^{15}N , ^{13}C -labeled proteins has enabled detailed structural and dynamic characterization of proteins up to 25 kDa in size.^{1,2} However, for protein systems >25 kDa, solution NMR studies are hampered by rapid signal decay caused by dipolar relaxation of spatially close ^1H - ^1H and ^1H - ^{13}C spin pairs.^{3,4} One strategy used to minimize dipolar relaxation is deuteration, where ^1H atoms in the protein are replaced with ^2H , providing sensitivity and resolution gains in various NMR experiments.⁵ Complete deuteration can be accomplished by bacterially expressing the protein in D_2O using deuterated glucose as a carbon source, but the problem with this approach is that it reduces the number of ^1H positions available as probes of structure and dynamics.

Several strategies have been developed to re-introduce protons at selected sites in proteins. Methyl groups are ideal molecular probes for solution NMR since they populate and define protein hydrophobic cores. Rapid spinning about the methyl symmetry axis attenuates intra-methyl dipolar relaxation mechanisms by about an order of magnitude, giving methyl groups the most intense and longest-lived NMR signals in a uniformly ^{13}C -labeled protein.⁶⁻⁸ The introduction of ^1H - ^{13}C -methyl groups into perdeuterated proteins, combined with solvent exchange of backbone and side-chain amide protons, permits the observation of ^1HN - ^1HN , ^1HN - $^{13}\text{CH}_3$ and $^{13}\text{CH}_3$ - $^{13}\text{CH}_3$ nuclear Overhauser enhancements (NOEs) to facilitate chemical shift assignments and structure determination.^{9,10} Protocols for methyl-selective protonation of alanine, leucine, isoleucine, valine, threonine, and methionine in a perdeuterated background have already been developed, and in combination with the methyl-TROSY pulse sequence, these have significantly extended the range of proteins accessible to multi-dimensional NMR.¹¹⁻¹⁴ However, a relative inability to observe the 14 non-methyl-containing amino acid side chains is a significant limitation of solution NMR of high molecular weight protein systems.

Aromatic amino acids are also highly represented in protein hydrophobic core and ligand binding interfaces, serving as valuable probes of protein dynamics and interactions.¹⁵⁻¹⁷ The unambiguous identification of NOEs involving aromatic rings is essential for accurate structure determination.¹⁸⁻²⁰ The application of ^1H , ^{13}C site-selective labeling of aromatic rings using SAIL amino acids in cell-free protein synthesis and α -ketoacid precursors in *E. coli* have been utilized in numerous structural and dynamic studies of proteins.^{15,17-19,21} These amino acids and

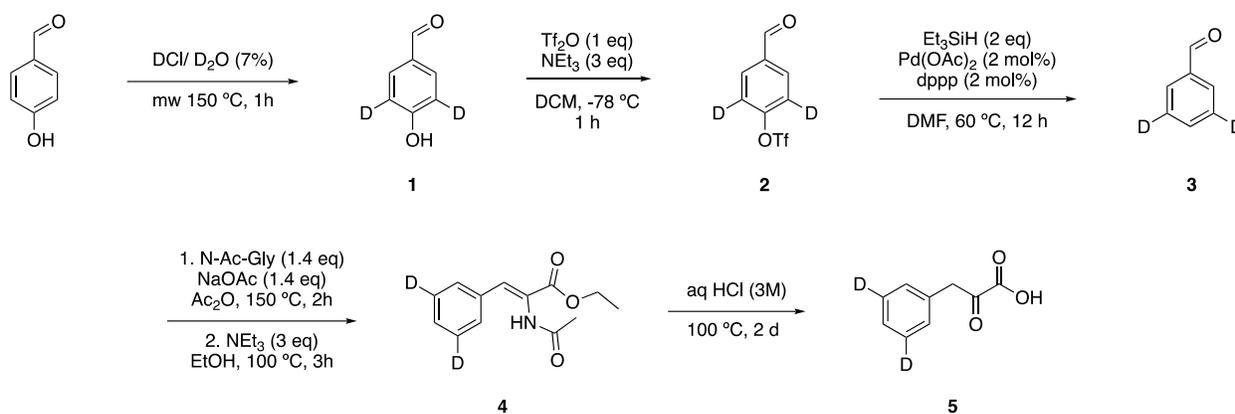
metabolic precursors, featuring alternating [$^1\text{H}, ^{13}\text{C}$]-[$^2\text{H}, ^{12}\text{C}$]-[$^1\text{H}, ^{13}\text{C}$] ring patterns, embed ^1H - ^{13}C spin systems in a deuterated background, eliminating large ^{13}C - ^{13}C and ^1H - ^1H scalar couplings and greatly simplifying chemical shift assignment. However, a significant drawback of this labeling strategy is the cost associated with synthesizing ^{13}C -labeled aromatic precursors and the rapid dipolar relaxation that still dominates any isolated ^1H - ^{13}C spin system, preventing the routine application of this labeling strategy to large protein systems.

Here, we develop a strategy for the cost-effective introduction of isolated ^1H - ^{12}C groups into the aromatic Phe, Tyr, Trp, and His amino acid residues in a manner compatible with existing methyl-labeling techniques and with minimal isotope scrambling. We combine this with previously developed methods for stereospecific deuteration of Asp, Asn, and Lys amino acid residues.²² The isolated ^1H - ^{12}C groups give rise to narrow ^1H -NMR signals to supplement ^1H - ^{13}C -methyl and ^1H - ^{15}N amide groups for NMR studies of large proteins.

Experimental

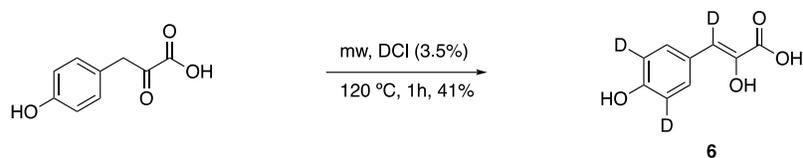
Synthesis of selectively deuterated phenylpyruvate, hydroxyphenylpyruvate and anthranilate precursor compounds for aromatic residue labeling

Lichtenecker et al. have demonstrated acid-catalyzed microwave-assisted deuteration of anthranilic acid, followed by efficient incorporation into protein tryptophan residues in *E. coli*.²³ Unlabeled anthranilate is readily available and can be deuterated in this manner, providing a cost-effective and efficient way to incorporate isolated ^1H - ^{12}C groups into tryptophan in a perdeuterated background. Here we present a cost-effective synthetic strategy for producing selectively deuterium-labeled phenylpyruvic acid and 4-hydroxyphenylpyruvic acid as precursors for Phe and Tyr, respectively.



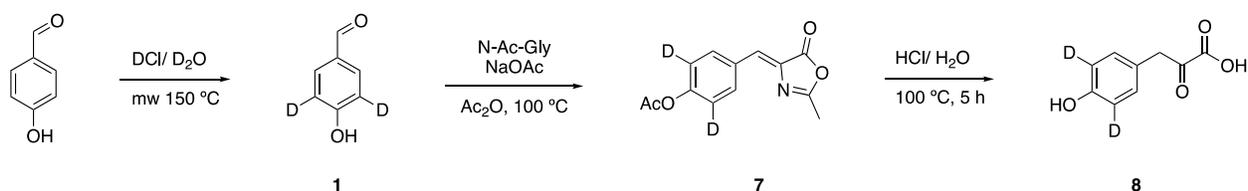
Scheme 1. Preparation of deuterium labeled phenylpyruvic acid at the gram scale

Phenylpyruvic acid is prepared from 4-hydroxybenzaldehyde. The first step introduces deuterium by Bronsted acid-catalyzed H/D exchange under microwave irradiation (Scheme 1). The presence of the hydroxyl and aldehyde groups ensures selective and complete deuterium incorporation at the positions *ortho* to the hydroxyl. This acid-catalyzed reaction involves the formation of an unstable intermediate, which can be readily converted to the aldehyde product after stirring in dichloromethane (DCM). We suspect the intermediate to be an acetal oligomer of 4-hydroxybenzaldehyde. Acid-catalyzed deuterium labeling reactions are unprecedented in the literature for substrates with an aldehyde group. Our method's advantage is preserving the aldehyde group, allowing its further transformation. This method dramatically enhances the synthetic efficiency compared with previous methods that require additional steps to produce aldehyde after deuterium incorporation.¹⁹ To achieve reductive cleavage of the phenolic C-O bond, we first convert the hydroxyl group to a better leaving group, followed by C-O bond activation. Treating the deuterium-labeled compound **1** with triflic anhydride transforms the hydroxy into a triflate leaving group. Pd-catalyzed hydrogenolysis of the triflate **2** is realized using silane while the aldehyde remains intact. Knoevenagel condensation between aldehyde **3** and N-acetylglycine yields the oxazolone product. For the preparation of phenylpyruvic acid, direct hydrolysis gives unsatisfactory yield and purity as both oxazolone and phenylpyruvic acid decompose on silica. The problem is solved by introducing an additional ethanolysis step which produces the more stable compound **4** that allows purification by chromatography. Final hydrolysis affords the desired product **5** with high purity of >95%. Following the same procedure, the gram-scale preparation affords a final yield of 34% over six steps.



Equation 1. Preparation of deuterium-labeled 4-hydroxyphenylpyruvic acid at gram scale

With the presence of a hydroxyl group, deuterium labeling of 4-hydroxyphenylpyruvic acid can be accomplished in just one step by microwave-assisted acid-catalyzed H/D exchange (Equation 1). The yield is 41% on the gram scale with a purity of 95% or higher. An alternative multistep synthesis of 4-hydroxyphenylpyruvic acid was also developed. The synthetic route involves 1, deuterium labeling of 4-hydroxybenzaldehyde; 2, Knoevenagel condensation with N-acetylglycine; 3, hydrolysis under acidic conditions (Scheme 2).



Scheme 2. Preparation of deuterium-labeled 4-hydroxyphenylpyruvic acid

Human Pin1 WW domain as an ideal system for monitoring precursor incorporation

To monitor precursor incorporation into protein, we used a model system, an engineered human Pin1 WW domain, the smallest folded protein domain containing all 20 amino acids except for cysteine²² (see Figure 3.S1). The small size of the protein and its high expression levels (100 mg/L M9 minimal media) facilitate the production of NMR samples and the collection of well-resolved 2D natural abundance (1%) ¹³C-HSQC spectra. For comparison, we prepared Pin1 uniformly labeled with ¹⁵N in H₂O to serve as a reference for comparing samples expressed in D₂O, facilitating the testing of various biosynthetic labeling methodologies. We also used mass spectrometry (MS) of acid-hydrolyzed protein to quantitate isotope incorporation and compared the results to our NMR data.

Protocol for expressing Pin1 with metabolic precursors in D₂O

Standard procedures were used to transform *E. coli* cells with a polyhistidine-tagged Pin1 protein construct (52-amino acid residues). Cells were grown using unlabeled carbon and nitrogen sources (10 g glucose, 2-10 g fumarate, 1 g ribose, and 1 g NH₄Cl). One hour prior to induction, 200 mg of selectively deuterated 4-hydroxyphenylpyruvate, 165 mg of deuterated phenylpyruvate, 30 mg deuterated anthranilate and 100 mg unlabeled histidine were added to 1 kg deuterated M9 minimal media for labeling aromatic residues, along with other unlabeled amino acids or their precursors (70 mg ketobutyrate, 125 mg isoketovalerate, 250 mg methionine, 100 mg threonine, 750 mg alanine, and 150 mg cysteine) and metabolic inhibitors (150 mg glyphosate, 1 g oxalate, 1 g malonate, and 5 mg cycloserine) used to limit isotope scrambling (see Table S1 for the complete list of precursors and small molecule inhibitors used to express Pin1 in D₂O).

Results

Incorporation of selectively deuterated phenylpyruvate, hydroxyphenylpyruvate, and anthranilate precursors into Phe, Tyr, and Trp.

Phenylalanine and Tyrosine

Phe and Tyr residues are produced from the chemically synthesized phenylpyruvate and hydroxyphenylpyruvate precursors specifically deuterated at the C ϵ positions (Figure 3.1). *In vivo*, an aromatic transaminase not only converts these precursors into phenylalanine and tyrosine, but also deuterates the C α and C β positions in D₂O. Protein production using deuterated aromatic precursors simplifies the NMR spectrum (Figure 3.2). The H ϵ positions of Phe and Tyr are highly deuterated and could not be detected above the noise level, suggesting deuteration levels >95%. In contrast, the H δ and H ζ positions of Phe and the H δ positions of tyrosine are highly protonated with peak intensities >130% compared to uniformly labeled Pin1 due to the elimination of ¹H-¹H 3-bond J couplings and improved relaxation properties (Figure 3.1). Given these changes in the NMR spectrum, it is difficult to obtain an exact value for the efficiency of incorporation, so we turned to data obtained from MS analysis of hydrolyzed Pin1, showing uniform deuteration at five positions (H α , H β_2 , H β_3 , H ϵ_1 , and H ϵ_2 , as determined by NMR, Table S2), as expected based on the deuterated precursor, suggesting >99% phenylpyruvate precursor incorporation into Phe. Unfortunately, acid-catalyzed hydrolysis

destroyed Tyr, preventing mass spectrometry profile analysis, but a comparison of NMR to Phe suggests ~99% labeling for Tyr as well.

Tryptophan

Anthranilate, a precursor for tryptophan biosynthesis, was selectively deuterated through the microwave-assisted acid-catalyzed solvent exchange in D₂O (Figure 3.1), as described previously²³. The six-member ring of the tryptophan indole group is derived exclusively from anthranilate, so the H ϵ 3 and H η 2 positions of the tryptophan aromatic rings are highly protonated, while the H ζ 2 and H ζ 3 positions are highly deuterated long-lived aromatic signals (Figures 3.1 and 3.2), with levels of incorporation comparable to Phe. The susceptibility of tryptophan to digestion upon acid hydrolysis prevents quantitation by mass spectrometry (Table S2).

In the pathway leading to tryptophan synthesis, some carbons derived from ribose-5-phosphate are incorporated into the five-member ring of the tryptophan indole (Figure 3.S2). Thus, supplementing ribose in our growth media improved the protonation at the Trp-H δ ₁ site from 40% to ~80% (Figure 3.1).

Supplementation of the bacterial growth media with ribose also stimulates endogenous synthesis of Phe/Tyr/Trp, which competes with isotope labeling via our aromatic precursors, manifesting as residual protonation of ~10% observed at the Trp H ζ 2 and Tyr H ϵ . The biosynthesis of aromatic amino acids in plants and many microbes relies on the enzymatic activity of 5-enolpyruvylshikimate-3-phosphate synthase (EPSP), a key enzyme in the shikimate pathway commonly targeted for drug and herbicide development (Figure 3.S3).²⁴ Previous studies have shown the successful inhibition of EPSP with glyphosate, the active ingredient in Roundup herbicide.²⁵ The addition of glyphosate into the *E. coli* growth media inhibited the endogenous aromatic amino acid biosynthesis without affecting cell viability and restored the labeling pattern of the aromatic residues to that observed in the absence of ribose. Thus, glyphosate as a metabolic inhibitor is not necessary unless ribose is supplied in the bacterial growth media.

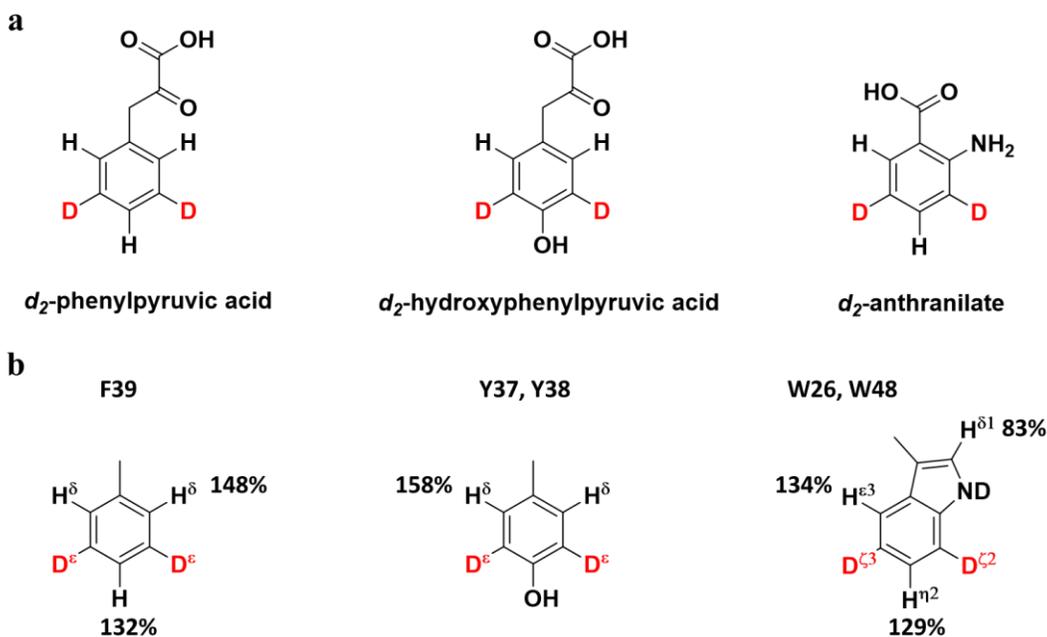


Figure 3.1. (a) α -ketoacid precursors used for labeling Phe, Tyr, and Trp aromatic rings. (b) Relative NMR signal (compared to fully protonated sample) observed in Pin1 aromatic residues.

Incorporation of isolated ^1H - ^{12}C groups into the aromatic ring of histidine

The unique acid/base properties of histidine make it a versatile catalytic residue in many enzymes. Selectively ^{13}C -labeled imidazolepyruvate has been used previously as an effective precursor for labeling the histidine aromatic ring.²⁶ The imidazole ring of histidine naturally contains isolated ^1H - ^{12}C groups at the δ and ϵ positions, derived biosynthetically from phosphoribose and ATP precursors, respectively (Figure 3.S4). The N-H site in the imidazole ring exchanges with solvent when protein is dissolved in D_2O , allowing the isolated protons within the ring to be free from scalar or dipolar ^1H - ^1H interactions in D_2O solvent. (From our observation, the imidazole C-H sites also exchange with solvent, but at a much slower rate over weeks.) We supplied *E. coli* cells with unlabeled histidine. Based on the HSQC data, both H^δ and H^ϵ positions of the histidine side chain had very high protonation (Table S2), suggesting near-100% incorporation with negligible endogenous biosynthesis. MS data also suggest a predominant undeuterated species (D_0). Apparent protonation levels $>100\%$ in the imidazole ring may be due to the control sample sitting in D_2O for a longer period, causing deuteration of the imidazole C-H sites.

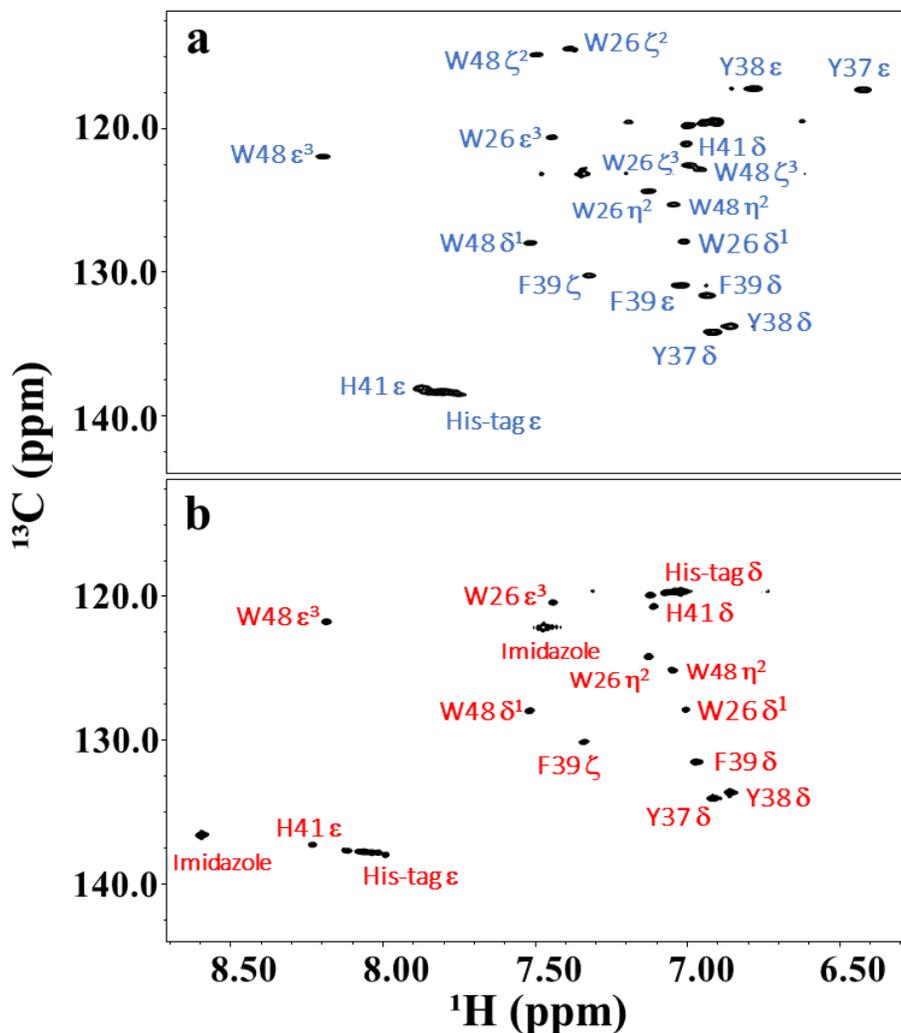


Figure 3.2: 2D ^1H , ^{13}C -HSQC spectra of Pin1 aromatic region. a) Natural abundance ^1H - ^{13}C 2D spectrum of uniformly ^{15}N -labeled Pin1. Cross peaks from this sample were set as the reference for quantitation of isotope incorporation in deuterated samples (without accounting for spectral changes caused by deuteration). b) Corresponding region of Pin1 labeled with selectively deuterated phenylpyruvate, hydroxyphenylpyruvate, anthranilate and unlabeled histidine in D_2O . Spectra were taken with the same acquisition parameters, and the intensities were normalized to the Met H ϵ peak to correct differences in concentration between samples.

Measurement of linewidth

2D ^1H - ^1H NOESY spectra were recorded to measure the linewidth of protein NMR signals introduced by our labeling scheme (Figure 3.3). The NMR signals of selectively deuterated aromatic signals are close to 10-fold more intense than an unlabeled Pin1 sample

(Figure 3.3c). These sensitivity gains were not appreciated in the natural abundance ^1H - ^{13}C 2D spectra because of the dominance of the short range ^1H - ^{13}C dipole-dipole in signal decay, but the 10-fold improvement in both sensitivity and resolution is expected based on deuteration in the absence of ^{13}C label. The narrowest linewidths in the entire protein belong to the $\text{H}\epsilon$ methyl group of methionine and the $\text{H}\delta$ of Phe and Tyr residues (Table S3). This is likely due to the high degree of mobility in these residues.²⁷ Rapid rotation about the χ_2 -dihedral angle in Phe and Tyr results in ring flips that can also reduce transverse relaxation rates.²⁸ Thus, the selective deuteration pattern introduced by our labeling scheme gives rise to NMR signals comparable to the methyl groups of alanine, threonine, isoleucine, leucine, and valine, indicating that this labeling scheme will result in viable long-lived NMR signals in high molecular weight protein systems.

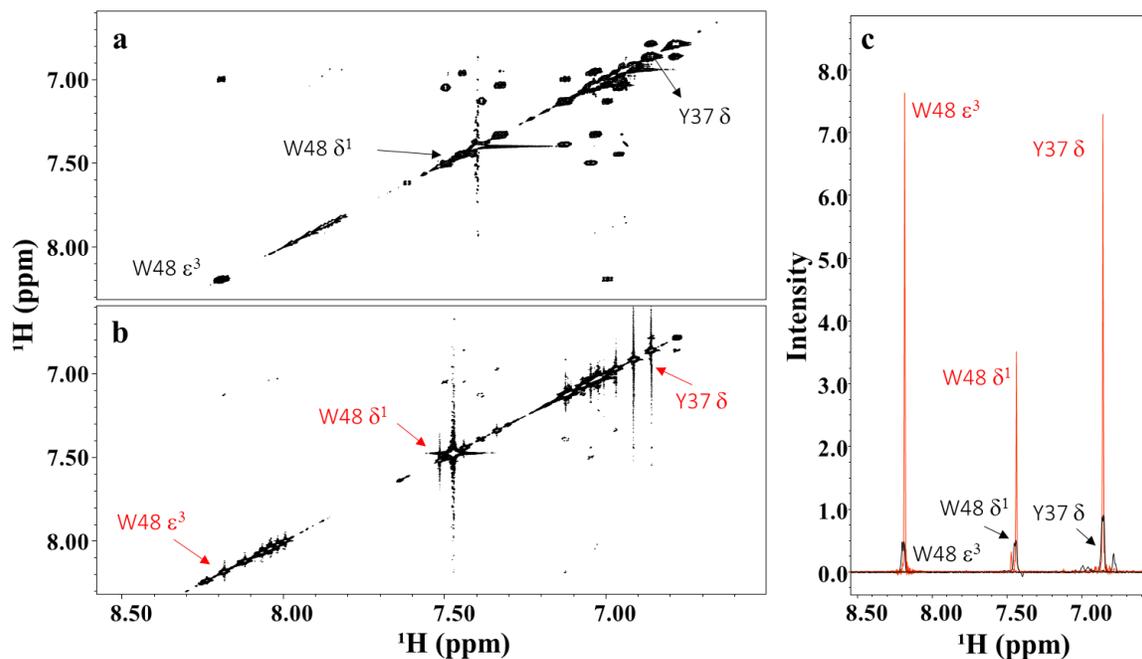


Figure 3.3: 2D ^1H - ^1H NOESY Spectra of Pin1 in D_2O . a, Unlabeled Pin1. b, Selectively deuterated Pin1 with aromatic precursors. c, corresponding intensities of the residues highlighted in a and b.

Cycloserine prevents alanine-pyruvate interconversion.

Pyruvate and alanine interconvert through the enzymatic activity of alanine aminotransferase, which catalyzes the final step of alanine biosynthesis from pyruvate. This enzyme also exchanges the hydrogen atoms at the alanine β -position with solvent deuterons,

significantly diluting the methyl signal intensity into multiple isotopomers. Studies have shown inhibition of alanine aminotransferase activity by incubating cells with L-cycloserine.²⁹ Consequently, we added L-cycloserine to our deuterated media to limit the scrambling of isotope labels at the Ala β -methyl group, and this increased the intensities by about 50% (Figure 3.4). Importantly, cycloserine did not noticeably affect cell growth or protein yields by inhibiting the enzymatic activity of other aminotransferases.²⁹ Our mass spectrometry analysis shows that ~20% of Ala is fully deuterated, suggesting that even in the presence of cycloserine, there is considerable biosynthetic production of deuterated Ala from pyruvate.

Combinatorial labeling of ILV methyl groups has been used to study many biological systems.^{10,30,31} Most of these studies have supplied *E. coli* 2-ketobutyrate and 2-ketoisovalerate precursors for labeling I- δ^1 and the prochiral methyl groups of leucine and valine, respectively. Supplementing our media with these precursors labeled Ile- δ^1 , Leu- δ^1 and 2 , and Val- γ^1 and 2 with intensities of 75%, 95-113%, and 97-98%, respectively, relative to non-deuterated protein (Table S2). This corresponds to about 85% isotopic label incorporation for Val assessed by mass spectrometry. Ile and Leu could not be separated by liquid chromatography for mass spectrometric analysis, as they have identical molecular weights, but based on the NMR data, isotopic labeling for Leu is on par with that for Val, whereas Ile labeling is less effective, about 65%. Another consequence of alanine aminotransferase activity is that protonated pyruvate from alanine becomes incorporated into the isoleucine- γ_2 methyl group. The use of cycloserine mitigates this, but there is still incorporation of protons into the Ile- γ_2 methyl group, probably due to enzymatic conversion of Ala to pyruvate, as well as endogenous pyruvate produced from unlabeled glucose (Table S2). The issue of residual protonation in the Ile- γ_2 methyl group has been noted previously by Rosen⁶ and can be ameliorated using 2-hydroxy-2-ethyl-3-ketobutyrate precursor.^{32,33}

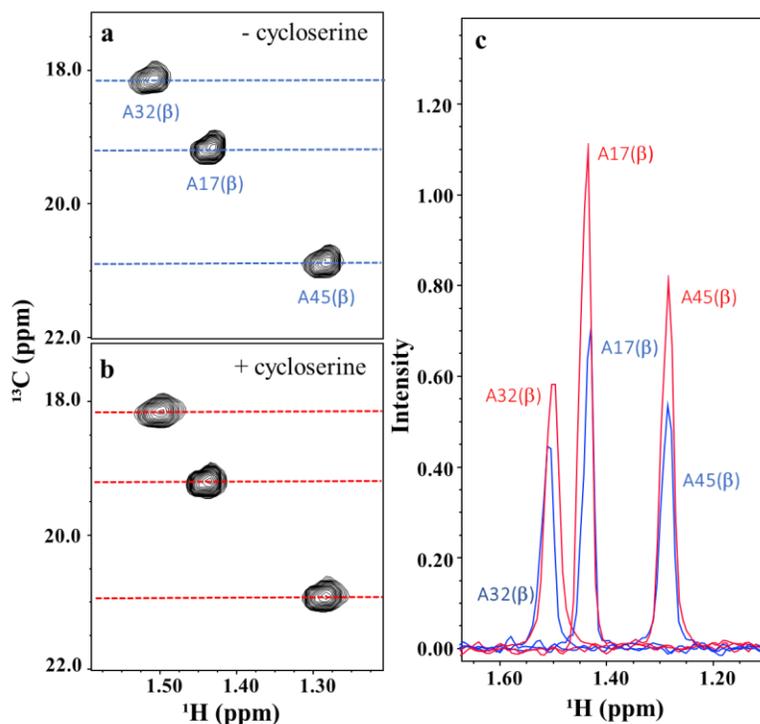


Figure 3.4: Cycloserine prevents alanine-pyruvate isotope scrambling. a, Natural abundance ^1H - ^{13}C HSQC showing alanines of selectively-labelled Pin1 grown without cycloserine. b, the corresponding region of Pin1 grown with cycloserine. c, cross-sections of the peaks highlighted in a, and b.

The addition of unlabeled cysteine and methionine to *E. coli* growth media decreases isotope scrambling at threonine- γ^2

Threonine residues are localized predominantly to protein surfaces (unlike other methyl-containing amino acids), especially at protein-nucleic acid interfaces.³⁴ These residues participate in hydrogen bonding, non-polar interactions, and signalling via phosphorylation.^{14,34} Specific labeling of the threonine methyl group using metabolic precursors is challenging due to the biosynthetic diversion of Thr isotopic label to other amino acids through complex, interconnected biosynthetic pathways. Velyvis and coworkers devised a biosynthetic strategy for producing deuterated ^1H , ^{13}C -methyl-labeled Thr, which was incorporated >90% into protein expressed by *E. coli* when supplied at 100 mg/L to the growth media.¹⁴ In this study, the supplemented threonine was metabolized to α -ketobutyrate (isoleucine precursor), but its incorporation into Ile could be minimized by supplementing α -ketobutyrate in the growth media as we have done.

In our growth media supplemented with threonine, we initially observed CH₃, CH₂D, CHD₂ (and likely CD₃) isotopomers (Figure 3.5a), consistent with competing endogenous biosynthesis of Thr via the Asp/Asn/Met/Lys/Thr biosynthetic pathway. (In contrast, endogenous Met biosynthesis seemed to be less of an issue in Met labeling.) Aspartate, derived from oxaloacetate in the TCA cycle, gives rise to threonine, methionine, and lysine, with metabolic branch points occurring at aspartate- β -semialdehyde (an intermediate in all three pathways) and homoserine (precursor for the formation of threonine and methionine) (Figure 3.S5). The biosynthesis of Thr, Lys and Met requires reduction of the C γ carboxyl group of Asp, consuming two hydride atoms from 2 NADPH molecules. Previous studies have shown inhibition of homoserine dehydrogenase (an essential enzyme in the biosynthetic pathway for threonine, methionine, and isoleucine) by Cys and Met.³⁵ We reasoned that Cys could be added to limit endogenous Thr biosynthesis and isotope scrambling, along with the Thr and Met already supplied to the growth media, as natural inhibitors of homoserine dehydrogenase. Indeed, adding Cys and Met to *E. coli* helped suppress endogenous Thr biosynthesis, improving isotopic labeling of Thr (Figure 3.5). Deuteration analysis by MS confirmed minimal scrambling at the methyl group of threonine, with ~100% of the mass profiles divided into three species, depending on protonation/deuteration at H α and H β (Figure 3.S5). There may be an endogenous transaminase that is catalyzing the deuteration of Thr at these positions.

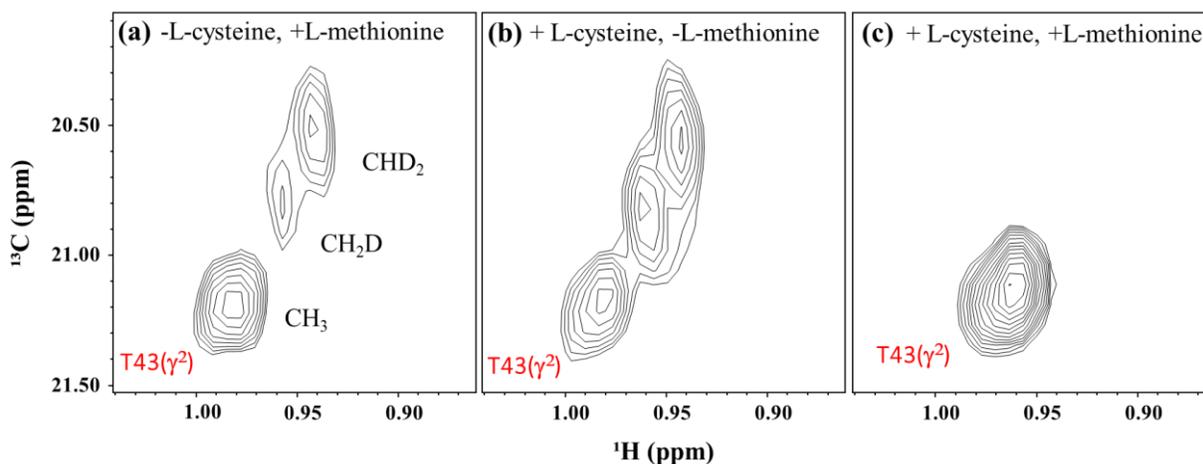


Figure 3.5: Prevention of threonine- γ_2 isotope scrambling by supplementing unlabeled methionine and cysteine in deuterated *E. coli* growth media. a, Natural abundance ¹H-¹³C HSQC spectra of selectively-labeled Pin1 grown with methionine but lacking cysteine. b, the

corresponding region of Pin1 grown with cysteine but without methionine c, the corresponding part of Pin1 produced with methionine and cysteine. The CH₃, CH₂D and CHD₂ isotopomers of T43(γ^2) are shown in panel a. The spectra were taken under identical experimental conditions and scaled for equal noise levels.

Methionine

Unlabeled methionine supplied in our D₂O medium was highly incorporated into protein, showing high levels of ¹H-¹²C protonation (~100%) at the ϵ -methyl group and all other sites except the H α position, which became highly deuterated (Table S2).

Stereoselective ¹H-¹²C-labeling of beta methylene proton H β^2 in Asp, Asn, Lys, and other amino acid residues

Our group previously introduced a method of stereoselective protonation at the beta methylene proton H β^2 in aspartate, asparagine, and lysine amino acid residues.²² These amino acids are derived from oxaloacetate, a key intermediate in the TCA cycle. Stereoselective deuteration at the beta methylene proton H β^3 is achieved by supplementing *E. coli* cells with unlabeled fumarate as a carbon source in D₂O. The enzyme fumarase converts fumarate into malate, with one ²H stereospecifically incorporated from D₂O. Malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate within the TCA cycle, which is converted to aspartate by a transamination reaction, preserving the stereochemistry at the beta methylene proton (Figure 3.S6). We add oxalate to the growth media to prevent isotope scrambling, inhibiting multiple enzymes involved in the interconversion between phosphoenolpyruvate, pyruvate, and oxaloacetate at the juncture of key metabolic pathways: glycolysis, TCA cycle and gluconeogenesis.³⁶⁻³⁸ Endogenous fumarate production by the TCA cycle is also inhibited by adding malonate to the growth media, a competitive succinate dehydrogenase inhibitor.³⁹

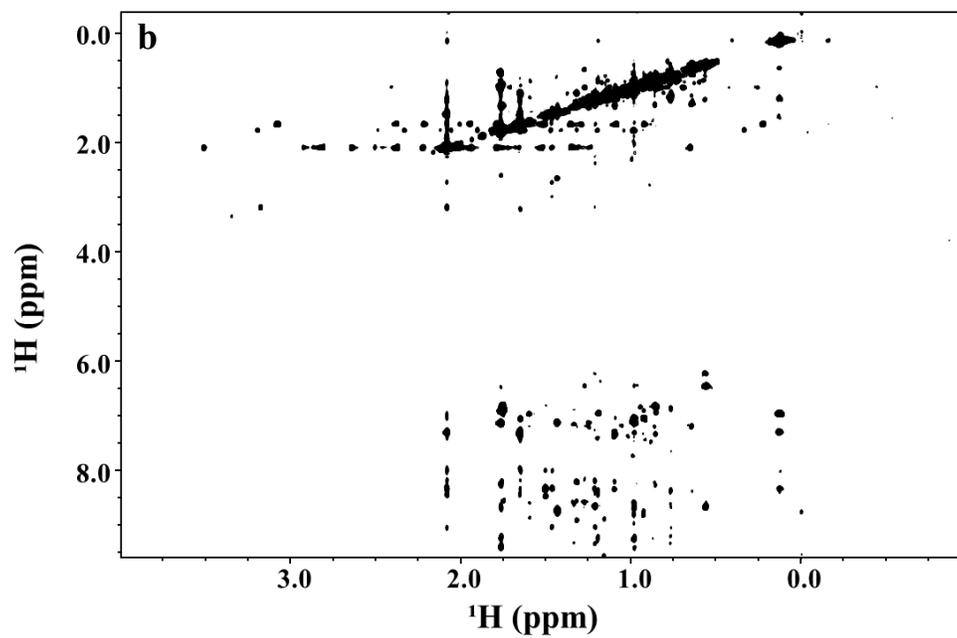
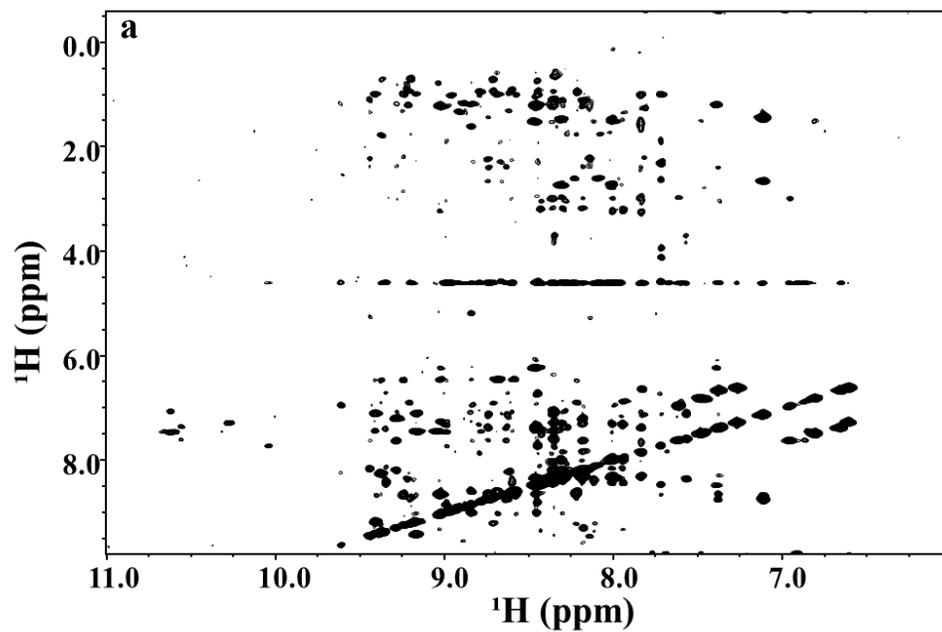
In our original scheme for stereospecifically deuterating Asp/Asn/Lys, we used rhamnose, pyruvate, and fumarate as carbon sources for the major metabolic pathways. However, such a strategy is untenable for efficiently labelling all 20 amino acids. Isotope scrambling occurs rapidly in pyruvate through the action of multiple alanine aminotransferase enzymes, and pyruvate is readily incorporated into Ala, Ile, Leu, and Val. Another major issue with our previous strategy is that rhamnose is not readily imported and metabolized in *E. coli*

BL21 strain, severely reducing protein production, a major problem when using expensive isotope labeling reagents. We, therefore, replaced rhamnose and pyruvate with glucose, the preferred carbon source of *E. coli*. Using our updated bacterial growth media containing glucose and different amino acid precursors decreased the efficiency of stereospecific H β labeling in Asp/Asn/Lys. We could offset this somewhat by increasing the sodium fumarate concentration in our media to 10 g/L. We obtained high degrees of stereospecific labeling for Asp and Asn, as measured by NMR, with signal intensities of 121% and 107%, respectively, compared to unlabeled Asp and Asn (Table S2). However, it is difficult to estimate the exact degree of labeling efficiency, because mass spectrometry of Asp is inaccurate due to re-protonation of the residue (exchange of proton for deuteron) that occurs during acid hydrolysis, as noted previously⁶. Lys sidechain labeling is a complex mixture with labeling at H β 2 about half of that observed for Asp/Asn H β 2, and Lys H β 2 > H δ > H γ .

γ H of Lys and Arg and δ H of Pro are contributed by NADPH, so there are low levels of protonation at these sites, with the hydride from NADPH partially derived from unlabeled glucose via the pentose phosphate pathway. There is also a high degree of protonation at Ser β H, derived from glucose, as previously observed,²² as well as Trp β H (derived from Ser biosynthetically).

Application of isolated ^1H - ^{12}C groups for structure determination of protein system > 30 kDa

Our isotope labeling strategy specifically introduces targeted ^1H - ^{12}C groups in Asp, Asn, Lys, Arg, Pro, Phe, Tyr, Trp, and His amino acid positions to complement ^1H - ^{13}C methyl groups in Ala, Thr, Met, Ile, Val, and Leu. We show that these groups can be connected to ^1H - ^{15}N amide positions using a multi-dimensional through-space NOESY experiment using the outer membrane protein PagP in DPC detergent micelles (~50 kDa) (Figure 3.6).



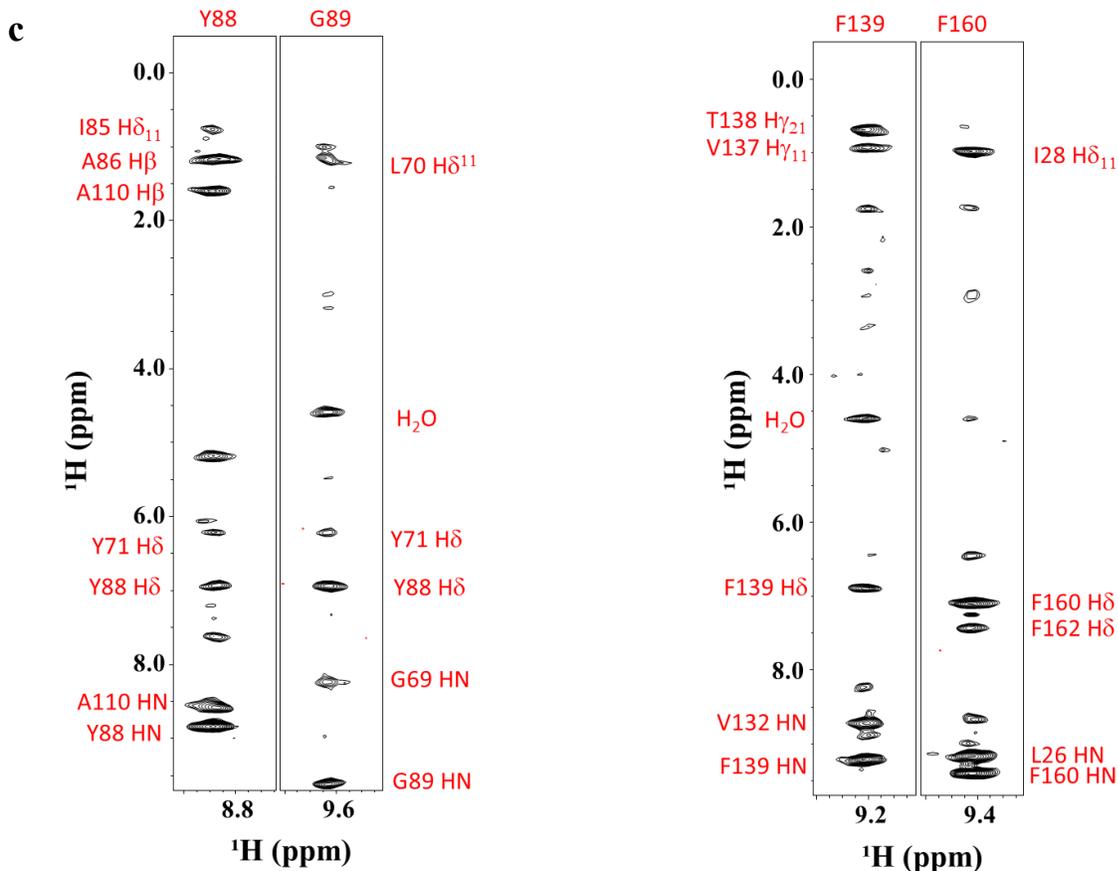


Figure 3.6. 3D NOESY spectra of PagP in DPC detergent micelles (50 kDa). a. 3D NOESY- $(^1\text{H}-^{15}\text{N})$ -HSQC, b. 3D NOESY- $(^1\text{H}-^{13}\text{C})$ -HSQC, c. strip plots taken from 3D NOESY- $(^1\text{H}-^{15}\text{N})$ -HSQC

Conclusion

We have developed an isotope labeling strategy in *E. coli* using inexpensive metabolic precursors in D_2O , complementing the now-standard method of ^1H , ^{13}C -methyl labeling for NMR studies of proteins >30 kDa. These precursors have shown a high degree of selectivity and incorporation levels, accompanied by a low cost of synthesis, adding additional ^1H signals for Phe, Tyr, and Trp residues that can be transferred via through-space NOEs to nearby methyl or amide protons for chemical shift assignment or structure determination of larger protein systems. In addition to these amino acids, it is also possible to obtain additional isolated ^1H signals in His, Asp, Asn, Lys, Arg, and Pro. The isolated ^1H - ^{12}C groups in a perdeuterated

background give rise to narrow ^1H -NMR signals with improved relaxation properties that rival methyl groups.

Acknowledgement

This work was supported by a Discovery Grant, RGPIN-2022-04105, from NSERC (Natural Sciences and Engineering Research Council of Canada), startup funds from the Department of Medicine and Faculty of Medicine & Dentistry at the University of Alberta, and private funding from the Hwang Professional Corporation.

References

- (1) Oschkinat, H.; Griesinger, C.; Kraulis, P. J.; Sørensen, O. W.; Ernst, R. R.; Gronenborn, A. M.; Clore, G. M. Three-Dimensional NMR Spectroscopy of a Protein in Solution. *Nature* **1988**, *332* (6162), 374–376. <https://doi.org/10.1038/332374a0>.
- (2) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.* **1990**, *89* (3), 496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-5](https://doi.org/10.1016/0022-2364(90)90333-5).
- (3) 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.* **2004**, *126* (34), 10560–10570. <https://doi.org/10.1021/ja047904v>.
- (4) Yu, H. Extending the Size Limit of Protein Nuclear Magnetic Resonance. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (2), 332–334. <https://doi.org/10.1073/pnas.96.2.332>.
- (5) Sattler, M.; Fesik, S. W. Use of Deuterium Labeling in NMR: Overcoming a Sizeable Problem. *Structure* **1996**, *4* (11), 1245–1249. [https://doi.org/10.1016/s0969-2126\(96\)00133-5](https://doi.org/10.1016/s0969-2126(96)00133-5).
- (6) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* **1996**, *263* (5), 627–636. <https://doi.org/10.1006/jmbi.1996.0603>.
- (7) Otten, R.; Chu, B.; Krewulak, K. D.; Vogel, H. J.; Mulder, F. A. A. Comprehensive and Cost-Effective NMR Spectroscopy of Methyl Groups in Large Proteins. *J. Am. Chem. Soc.* **2010**, *132* (9), 2952–2960. <https://doi.org/10.1021/ja907706a>.
- (8) Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: Explanation and Experimental Verification. *Magn. Reson. Chem.* **2003**, *41* (10), 843–852. <https://doi.org/10.1002/mrc.1256>.
- (9) Sprangers, R.; Velyvis, A.; Kay, L. E. Solution NMR of Supramolecular Complexes: Providing New Insights into Function. *Nat. Methods* **2007**, *4* (9), 697–703. <https://doi.org/10.1038/nmeth1080>.
- (10) Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, *1* (2), 749–754. <https://doi.org/10.1038/nprot.2006.101>.
- (11) 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 (δ 1). *J. Biomol. NMR* **1999**, No. 13, 369–374. <https://doi.org/10.1023/a:1008393201236>.
- (12) Ayala, I.; Sounier, R.; Usé, N.; Gans, P.; Boisbouvier, J. An Efficient Protocol for the Complete Incorporation of Methyl-Protonated Alanine in Perdeuterated Protein. *J. Biomol. NMR* **2009**, *43* (2), 111–119. <https://doi.org/10.1007/s10858-008-9294-7>.
 - (13) Fischer, M.; Kloiber, K.; Häusler, J.; Ledolter, K.; Konrat, R.; Schmid, W. Synthesis of a ^{13}C -Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. *ChemBioChem* **2007**, *8* (6), 610–612. <https://doi.org/10.1002/cbic.200600551>.
 - (14) Velyvis, A.; Ruschak, A. M.; Kay, L. E. An Economical Method for Production of $2\text{H},^{13}\text{C}$ -Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 KDa Proteasome. *PLoS One* **2012**, *7* (9), 1–8. <https://doi.org/10.1371/journal.pone.0043725>.
 - (15) Takeda, M.; Ono, A. M.; Terauchi, T.; Kainosho, M. Application of SAIL Phenylalanine and Tyrosine with Alternative Isotope-Labeling Patterns for Protein Structure Determination. *J. Biomol. NMR* **2010**, *46* (1), 45–49. <https://doi.org/10.1007/s10858-009-9360-9>.
 - (16) Weininger, U. Site-Selective ^{13}C Labeling of Histidine and Tryptophan Using Ribose. *J. Biomol. NMR* **2017**, *69* (1), 23–30. <https://doi.org/10.1007/s10858-017-0130-9>.
 - (17) Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late Metabolic Precursors for Selective Aromatic Residue Labeling. *J. Biomol. NMR* **2018**, *71* (3), 129–140. <https://doi.org/10.1007/s10858-018-0188-z>.
 - (18) Miyanoiri, Y.; Takeda, M.; Jee, J.; Ono, A. M.; Okuma, K.; Terauchi, T.; Kainosho, M. Alternative SAIL-Trp for Robust Aromatic Signal Assignment and Determination of the X2 Conformation by Intra-Residue NOEs. *J. Biomol. NMR* **2011**, *51* (4), 425–435. <https://doi.org/10.1007/s10858-011-9568-3>.
 - (19) Lichtenecker, R. J. Synthesis of Aromatic $^{13}\text{C}/^2\text{H}$ - α -Ketoacid Precursors to Be Used in Selective Phenylalanine and Tyrosine Protein Labelling. *Org. Biomol. Chem.* **2014**, *12* (38), 7551–7560. <https://doi.org/10.1039/c4ob01129e>.
 - (20) Torizawa, T.; Ono, A. M.; Terauchi, T.; Kainosho, M. NMR Assignment Methods for the Aromatic Ring Resonances of Phenylalanine and Tyrosine Residues in Proteins. *J. Am. Chem. Soc.* **2005**, *127* (36), 12620–12626. <https://doi.org/10.1021/ja051386m>.
 - (21) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal Isotope Labelling for NMR Protein Structure Determinations. *Nature* **2006**, *440* (7080), 52–57. <https://doi.org/10.1038/nature04525>.
 - (22) Danmaliki, G. I.; Liu, P. B.; Hwang, P. M. Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for Escherichia Coli in D_2O . *Biochemistry* **2017**, *56* (45). <https://doi.org/10.1021/acs.biochem.7b00991>.
 - (23) Schörghuber, J.; Geist, L.; Bisaccia, M.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Anthranilic Acid, the New Player in the Ensemble of Aromatic Residue Labeling Precursor Compounds. *J. Biomol. NMR* **2017**, *69* (1), 13–22. <https://doi.org/10.1007/s10858-017-0129-2>.
 - (24) Schönbrunn, E.; Eschenburg, S.; Shuttleworth, W. A.; Schloss, J. V.; Amrhein, N.; Evans,

- J. N. S.; Kabsch, W. Interaction of the Herbicide Glyphosate with Its Target Enzyme 5-Enolpyruvylshikimate 3-Phosphate Synthase in Atomic Detail. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (4), 1376–1380. <https://doi.org/10.1073/pnas.98.4.1376>.
- (25) Amrhein, N.; Deus, B.; Gehrke, P.; Steinrücken, H. C. The Site of the Inhibition of the Shikimate Pathway by Glyphosate. *Plant Physiol.* **1980**, *66* (5), 830–834. <https://doi.org/10.1104/pp.66.5.830>.
- (26) Schörghuber, J.; Geist, L.; Platzer, G.; Konrat, R.; Lichtenecker, R. J. Highly Selective Stable Isotope Labeling of Histidine Residues by Using a Novel Precursor in E. Coli-Based Overexpression Systems. *ChemBioChem* **2017**, *18* (15), 1487–1491. <https://doi.org/10.1002/cbic.201700192>.
- (27) Danmaliki, G. I.; Hwang, P. M. Proton TOCSY NMR Relaxation Rates Quantitate Protein Side Chain Mobility in the Pin1 WW Domain. *J. Biomol. NMR* **2022**, *76* (4), 121–135. <https://doi.org/10.1007/s10858-022-00400-5>.
- (28) Gauto, D. F.; Macek, P.; Barducci, A.; Fraga, H.; Hessel, A.; Terauchi, T.; Gajan, D.; Miyanoiri, Y.; Boisbouvier, J.; Lichtenecker, R.; Kainosho, M.; Schanda, P. Aromatic Ring Dynamics, Thermal Activation, and Transient Conformations of a 468 KDa Enzyme by Specific ¹H-¹³C Labeling and Fast Magic-Angle Spinning NMR. *J. Am. Chem. Soc.* **2019**, *141* (28), 11183–11195. <https://doi.org/10.1021/jacs.9b04219>.
- (29) Cornell, N. W.; Zuurendonk, P. F.; Kerich, M. J.; Straight, C. B. Selective Inhibition of Alanine Aminotransferase and Aspartate Aminotransferase in Rat Hepatocytes. *Biochem. J.* **1984**, *220* (3), 707–716. <https://doi.org/10.1042/bj2200707>.
- (30) Tugarinov, V.; Kay, L. E. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **2003**, *125* (45), 13868–13878. <https://doi.org/10.1021/ja030345s>.
- (31) 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.* **2014**, *136* (32), 11308–11310. <https://doi.org/10.1021/ja504791j>.
- (32) Ruschak, A. M.; Velyvis, A.; Kay, L. E. A Simple Strategy for ¹³C,¹H Labeling at the Ile- Γ 2 Methyl Position in Highly Deuterated Proteins. *J. Biomol. NMR* **2010**, *48* (3), 129–135. <https://doi.org/10.1007/s10858-010-9449-1>.
- (33) 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.* **2012**, *48* (10), 1434–1436. <https://doi.org/10.1039/c1cc12932e>.
- (34) Sinha, K.; Jen-Jacobson, L.; Rule, G. S. Specific Labeling of Threonine Methyl Groups for NMR Studies of Protein-Nucleic Acid Complexes. *Biochemistry* **2011**, *50* (47), 10189–10191. <https://doi.org/10.1021/bi201496d>.
- (35) Ogata, K.; Yajima, Y.; Nakamura, S.; Kaneko, R.; Goto, M.; Ohshima, T.; Yoshimune, K. Inhibition of Homoserine Dehydrogenase by Formation of a Cysteine-NAD Covalent Complex. *Sci. Rep.* **2018**, *8* (1), 1–2. <https://doi.org/10.1038/s41598-018-24063-1>.
- (36) Stiffin, R. M.; Sullivan, S. M.; Carlson, G. M.; Holyoak, T. Differential Inhibition of Cytosolic PEPCK by Substrate Analogues. Kinetic and Structural Characterization of Inhibitor Recognition. *Biochemistry* **2008**, *47* (7), 2099–2109. <https://doi.org/10.1021/bi7020662>.
- (37) Yang, Z.; Floyd, D. L.; Loeber, G.; Tong, L. Structure of a Closed Form of Human Malic

- Enzyme and Implications for Catalytic Mechanism. *Nat. Struct. Biol.* **2000**, *7* (3), 251–257. <https://doi.org/10.1038/73378>.
- (38) Narindrasorasak, S.; Bridger, W. A. Probes of the Structure of Phosphoenolpyruvate Synthetase: Effects of a Transition State Analogue on Enzyme Conformation. *Can. J. Biochem.* **1978**, *56* (8), 816–819. <https://doi.org/10.1139/o78-124>.
- (39) Valls-Lacalle, L.; Barba, I.; Miró-Casas, E.; Ruiz-Meana, M.; Rodríguez-Sinovas, A.; García-Dorado, D. Selective Inhibition of Succinate Dehydrogenase in Reperfused Myocardium with Intracoronary Malonate Reduces Infarct Size. *Sci. Rep.* **2018**, *8* (1), 1–10. <https://doi.org/10.1038/s41598-018-20866-4>.

Supplementary Information

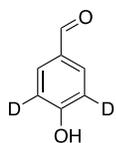
Experimental Procedures

Synthesis of selectively deuterated phenylpyruvate and 4-hydroxyphenylpyruvate

Chemicals were purchased from Millipore Sigma and used upon receiving without further purification. Microwave reactions were conducted in the Biotage Initiator Microwave System. For each microwave reaction, new consumables were required, including vials, stir bars and caps with septum from the Biotage microwave reaction kit. Three vial sizes are available: 0.5-2 mL, 2-5 mL, and 10-20 mL. Microwave reactions were performed at 0.5 mL scale first before gradually increasing to 2 mL and 10 mL.

Organic reactions were performed in the fume hood under nitrogen atmosphere unless otherwise stated. The glassware and stir bar were dried in the oven at 120 °C overnight before cooling down to room temperature under nitrogen.

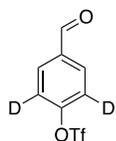
4-hydroxybenzaldehyde (10 mmol, 1.22 g) were added into the vial equipped with a stir bar. 8 mL D₂O and 2 mL 35% DCl/ D₂O solution were then added to the vial by syringes. The mixed solution is 7% w/w DCl/ D₂O. The vial was placed in the rack of Biotage Microwave reactor. The reaction condition was set to heat at 150 °C for one hour. After the microwave reaction completed, 10 mL DCM was injected into the vial. The reaction was stirred at room temperature overnight. The product was extracted with DCM (20 mL x 3) and the organic layer was dried over sodium sulfate. The reaction afforded **1** with a 95% yield after solvent evaporation. The deuterium incorporation level at the ortho position was 95%.



1

3,5-dideuterium-4-hydroxybenzaldehyde **1** (95%); $^1\text{H NMR}$ (CDCl_3 , 500 MHz) δ 9.87 (s, 1H), 7.82 (s, 2H).

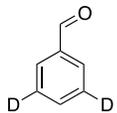
Deuterium labelled 4-hydroxybenzaldehyde **1** (10 mmol, 1.25 g) was weighed into a round bottom flask equipped with a stir bar. The flask was sealed with a septum and placed under nitrogen atmosphere. 20 mL anhydrous DCM was added into the flask by syringe. NEt_3 (30 mmol, 4.2 mL) distilled from CaH_2 was then added into the flask by syringe. The reaction flask was immersed into a dry ice/ acetone bath at -78°C . Tf_2O (10 mmol, 1.7 mL) was added dropwise into the reaction mixture. Upon addition, the reaction was maintained at -78°C for one hour while monitored by TLC. Once completed, 20 mL water was added into the reaction mixture. The organic layer was first extracted with DCM (20 mL x 3), then dried over sodium sulfate. After solvent evaporation, the crude product **2** was purified by chromatography (5-10% ethyl acetate/ hexane).



2

3,5-dideuterium-4-triflate-benzaldehyde **2**: oily (93%); $^1\text{H NMR}$ (CDCl_3 , 500 MHz) δ 10.05 (s, 1H), 8.01 (s, 2H).

Product **3** was obtained from Pd-catalyzed reduction according to a literature procedure.^[1]

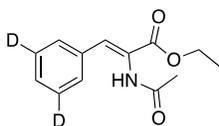


3

3,5-dideuterium-benzaldehyde **3**: clear liquid (80%); $^1\text{H NMR}$ (CDCl_3 , 500 MHz) δ 10.03 (s, 1H), 7.89 (s, 2H), 7.64 (s, 1H).

Oxazolone was prepared according to a literature procedure.^[2]

The reaction mixture from oxazolone preparation was concentrated by solvent evaporation. 10 mL anhydrous ethanol and 1 mL NEt_3 were added to the reaction flask. The reaction mixture was refluxed for three hours under nitrogen atmosphere. Upon completion, solvent was removed by evaporation. The crude product was purified by chromatography using 20-50% ethyl acetate/hexane.

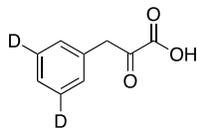


4

Compound **4**: (60% yield over two steps); $^1\text{H NMR}$ (CDCl_3 , 500 MHz) δ (ppm) 7.46 (s, 2H), 7.38 (s, 1H), 7.33 (s, 1H), 6.99 (s, 1H), 4.31 (q, $J = 7.0$ Hz, 2H), 2.14 (s, 3H), 1.36 (t, $J = 7.0$ Hz, 3H); $^{13}\text{C NMR}$ (CDCl_3 , 125 MHz) δ (ppm) 169.0, 165.3, 133.8, 132.0, 129.5, 129.1, 128.2, 124.7, 61.8, 23.3, 14.2; **HRMS (EI)**: Calcd. For $\text{C}_{13}\text{H}_{13}\text{D}_2\text{NO}_3$: 235.1177; Found: 235.1179; **IR (film)**: 3246, 3050, 2983, 2936, 2908, 2266, 17200, 1667, 1643, 1514, 1443, 1415, 1373, 1287, 1248, 1197, 1124, 1208, 920, 863, 812 cm^{-1} .

Compound **4** (5 mmol, 1.176 g) was added to a round bottom flask equipped with a stir bar. 10 mL of aqueous HCl (3 M) was then added to the flask. The reaction was heated to reflux for 2 days while monitored by TLC. The reaction mixture was cooled down and stored in fridge to

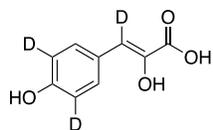
allow precipitation. After rinsed with cold water (3 mL x 3), the solid product **5** was obtained by filtration.



5

3,5-dideuterium-phenylpyruvic acid **5** (81%); $^1\text{H NMR}$ (CD_3OD , 500 MHz) δ 7.76 (s, 2H), 7.21 (s, 1H), 6.48 (s, 1H), 4.84 (s, 1H), 3.30 (s, 1H); $^{13}\text{C NMR}$ (CD_3OD , 125 MHz) δ (ppm) 168.4, 142.2, 136.4, 130.8, 129.3, 128.2, 111.6; **HRMS (EI)**: Calcd. For $\text{C}_9\text{H}_6\text{D}_2\text{O}_3$: 166.0599; Found: 166.0598; **IR (film)**: 3476, 3456, 3095, 2619, 1704, 1689, 1663, 1624, 1456, 1419, 1376, 1292, 1247, 1193, 1114, 917, 898, 884 cm^{-1} .

Add 4-hydroxyphenylpyruvic acid (10 mmol, 1.80 g) into a microwave vial (size 10-20 mL) equipped with a stir bar. 9 mL D_2O and 1 mL 35% $\text{DCl}/\text{D}_2\text{O}$ solution was then added into the reaction. The mixed solution is 3.5% w/w $\text{DCl}/\text{D}_2\text{O}$. The vial was placed on the rack in the Biotage Microwave reactor which was set to heat at 120 $^\circ\text{C}$ for one hour. Upon completion, the solid product was obtained by filtration and then rinsed with cold water (2 mL x 3). The reaction afforded compound **6** with a 41% yield. The deuterium incorporation level at the ortho positions was 95%.



6

3,5-dideuterium-4-hydroxyphenylpyruvic acid **6** (41%); $^1\text{H NMR}$ (CD_3OD , 500 MHz) δ 7.62 (s, 2H); $^{13}\text{C NMR}$ (CD_3OD , 125 MHz) δ (ppm) 168.7, 158.1, 139.9, 132.4, 132.3, 127.9, 116.0; **HRMS (EI)**: Calcd. For $\text{C}_9\text{H}_5\text{D}_3\text{O}_4$: 183.0611; Found: 183.0611; **IR (film)**: 3450, 3400, 3093,

2657, 2555, 1683, 1639, 1598, 1479, 1462, 1416, 1370, 1325, 1303, 1241, 1161, 916, 894, 863 cm⁻¹.

Protocol for bacterial expression of Pin1 protein in 1 L H₂O-M9 minimal media

Standard procedures were used to transform BL21(DE3) *E. coli* bacteria cells with an IPTG-inducible histidine-tagged Pin1 expression plasmid. Four to six transformed colonies were inoculated into 20 mL starter culture of LB rich medium containing 100 mg/L ampicillin. Cells were grown in a shaker at 37 °C to an optical density (OD) of about 0.7-0.9 at 600 nm. 10 mL from the starter culture was diluted into sterile filtered 1 L M9 minimal media, containing 9 g anhydrous Na₂HPO₄, 2.5 g K₂HPO₄, 10 g anhydrous D-glucose ¹²C/¹³C, 1 g NH₄Cl/¹⁵NH₄SO₄, 0.09 g ampicillin, 0.22 g MgSO₄, 0.01g CaCl₂, 0.10 g thiamine, and 0.01 g biotin and allowed to grow further until OD_{600nm} ~0.8. Cells were induced with 200 mg IPTG and grown for 6 h post induction. The cells were centrifuged at 5000 rpm (4420 g) for 20 mins and harvested. The cells were lysed, purified by Ni-NTA affinity chromatography followed by dialysis and lyophilization as previously described.^[3] The protein identity and purity were confirmed by gel electrophoresis.

Protocol for bacterial expression of Pin1 protein with metabolic precursors and small molecule inhibitors in 1 L D₂O-M9 minimal media

The transformation and starter cultures were the same as described in the previous section. 10 mL from the starter culture of LB rich medium was diluted into 1 L sterile M9-H₂O medium containing unlabeled glucose and nitrogen sources, selected precursor compounds for labeling desired amino acids and small molecule inhibitors to prevent scrambling of isotope labels as shown in Table 3.S1. Cells were grown to an OD_{600nm} of about 0.3-0.4 (doubling time of H₂O M9 about is 1 h). The cells were spun down at 5000 rpm for 20 min at room temperature and resuspended into 0.9 kg D₂O-M9 minimal media containing the mixture shown in Table 3.S1. The cells were grown for at least one doubling time (~3 h), aiming for to an OD_{600nm} of about 0.6-0.9. One hour before induction with IPTG, 150 mg selectively deuterated phenylpyruvate, 200 mg deuterated hydroxyphenylpyruvate, 30 mg deuterated anthranilate were added to the media, along with other methyl-containing amino acids or their metabolic precursors (70 mg deuterated ketobutyrate, 125 mg isoketovalerate, 250 mg methionine, 100 mg threonine and 750 mg alanine) and glyphosate, oxalate, malonate, cysteine and cycloserine

metabolic inhibitors used to limit isotope scrambling as shown in Table 3.S1. Note that ketobutyric acid was added 20 minutes before induction to limit isotope scrambling in the methyl groups of isoleucine. Post-IPTG induction, the cells were grown for an additional 6 hours at 37 °C and then harvested and lysed. The protein was purified using nickel affinity chromatography.

Table 3.S1. M9 Minimal Media Composition for Stereoselective Labeling of Proteins

1 L H₂O-M9 Minimal Media	D₂O-M9 (0.9 L = 1 kg) Minimal Media
9 g anhydrous Na ₂ HPO ₄	8g Na ₂ HPO ₄
2.5 g K ₂ HPO ₄	2.2g K ₂ HPO ₄
10 g D-Glucose (anhydrous)	10 g D-Glucose (1D, 98%)
3 g Sodium Fumarate	10 g Sodium Fumarate
1 g Ribose	1 g Ribose
1 g NH ₄ Cl	1 g NH ₄ Cl
90 mg Ampicillin	90 mg Ampicillin
0.22 g MgSO ₄ (anhydrous)	0.22 g MgSO ₄ (anhydrous)
0.01 g CaCl ₂ (anhydrous)	0.01 g CaCl ₂ (anhydrous)
1 g Oxalic acid dehydrate	1 g Oxalic acid dehydrate
1 g Malonic acid	1 g Malonic acid
100 mg Adenine	100 mg Adenine
100 mg Thiamine (optional)	100 mg Thiamine (optional)
1 mg Biotin (optional)	1 mg Biotin (optional)
target pH is ~7.4	target pH is ~7.4

Precursor compounds added 1 h prior to induction

50 mL, 22 mM deuterated isoketovalerate

750 mg Alanine

5 mg L-cycloserine

250 mg Methionine

100 mg Threonine
150 mg L-cysteine
150 mg Phenylpyruvate
200 mg Hydroxyphenylpyruvate
30 mg Anthranilate
150 mg Glyphosate
100 mg Histidine

Precursor compound added 20 min prior to induction

50 mL, 14 mM deuterated ketobutyric acid

Protocol for isoketovalerate and ketobutyrate deuteration

125 mg isoketovalerate was dissolved in 50 mL D₂O and incubated at 37 °C overnight at pH ~12.

70 mg of ketobutyrate was dissolved in 50 mL D₂O and incubated at 37 °C overnight at a pH ~10.

NMR spectroscopy

The NMR samples were prepared in 5 mm NMR tubes to a final volume of 500 µL. The protein samples were dissolved in 100 mM KCl, 10 mM imidazole, 0.01% NaN₃ in 90% H₂O, 10% D₂O, or 100% D₂O using 0.5 mM 2,2-dimethyl-2-silapentane-5-sulfonate-d₆ sodium salt (DSS-d₆) as an NMR chemical shift internal reference. The 2 mM protein samples were maintained at pH ~6.8. All NMR experiments were carried out at 30 °C on a Varian Inova 500 MHz spectrometer equipped with a triple-resonance pulsed field gradient probe head. 1D experiments were processed on VNMRJ (Varian Inc.); 2D experiments were processed using the NMRDraw/NMRPipe software and analyzed using NMRViewJ (One moon scientific).^[4,5]

Amino acid digestion and Mass Spectrometry Analysis

Amino acid hydrolysis:

Samples were dried under vacuum in Kimax, 10 x 75mm, glass tubes. 100 uL 6M HCl was added and sealed under vacuum. Samples were hydrolyzed for 24 hours at 110 °C, and hydrolysates were dried under a vacuum.

HPLC/MS:

Sample solutions were analyzed using an Aria MX HPLC system (Thermo Fisher Scientific) coupled with an Orbitrap Elite Mass spectrometer (Thermo Fisher Scientific). An Agilent Poroshell 120 EC-C18 column 100 mm x 2.1 mm, 2.7 µm particle size (Agilent Technologies, Mississauga, ON) was employed for LC separations. The column temperature was controlled at 25 °C. The mobile phase A had acetonitrile with 0.1% formic acid. Phase B had 95:5 water: acetonitrile (V/V) with 0.1% formic acid. The gradient was as follows: 0-2 min, 5% A; 2-20 min, linear gradient to 60% A; 20-25min, linear gradient to 90% A; 25-26 min, back to 5% A and hold for 7 min. The flow rate of the mobile phase was 200 µl/min, and the cycle time was 33 min/injection. The orbitrap mass spectrometer was operated under electrospray positive ion mode. The ionization voltage was set at +3.5 kV. Nitrogen was used as sheath, aux, and sweep gas with arbitrary units of 25, 20, and 3, respectively. The ion source and capillary temperature were at 300 °C and 325 °C, respectively. Mass calibration and tuning were done by infusing LTQ Velos ESI Positive Ion Calibration Solution (Thermo Fisher Scientific) before performing HPLC/MS analysis.

Mass spectrometry data analysis:

The Xcalibur software v. 2.2 (Thermo Fisher Scientific) was used for data acquisition and analysis. The acquisition was carried out in full scan mode with a mass range from 50 to 800 amu with resolving power set to a nominal value of 120,000 at full-width half-maximum at m/z 400. The acquisition method was set using a chromatogram from the unlabeled sample and applied to the labelled sample to scan specific masses, yielding expected retention times in both instances. Isotopic masses were calculated for each of the 19 unlabelled amino acids, and expected masses were computed for deuterium(s) labelled sample. The masses for each amino acid were calibrated against a corresponding expected retention time chosen as an internal standard. Data were extracted to yield areas and retention times related to unlabelled and heavy

labels of each amino acid. The main method parameters were as follows: mass tolerance, 10.0 ppm; mass precision, 4 decimals.

Results

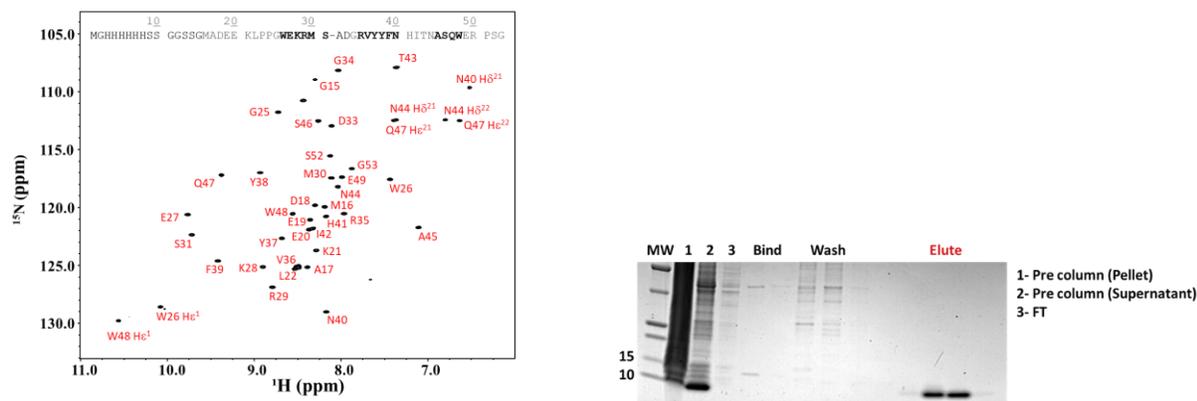


Figure 3.S1. Two dimensional ^1H - ^{15}N HSQC spectrum of Pin1, along with its amino acid sequence. SDS-Page gel stained with Coomassie brilliant blue, showing the production and purification of poly histidine-tagged Pin1.

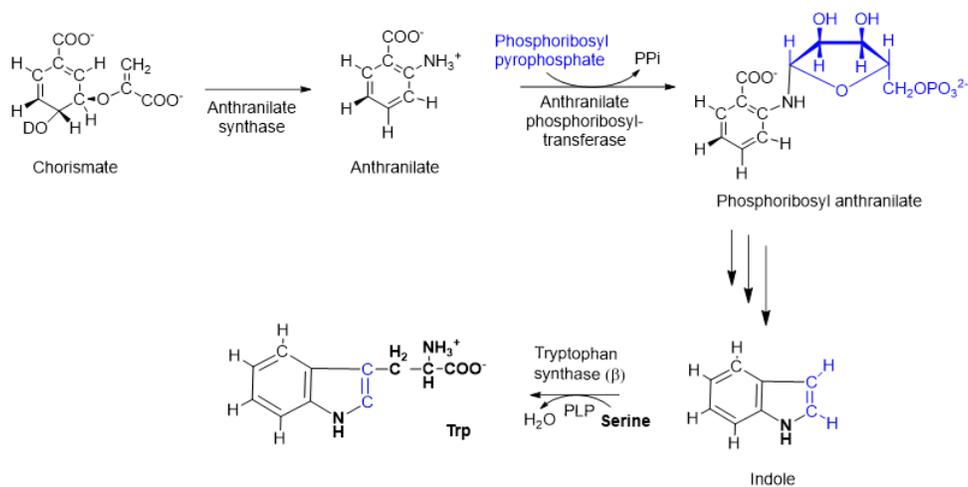


Figure 3.S2. Biosynthesis of tryptophan from chorismate. Atom derived from phosphoribose are highlighted in blue.

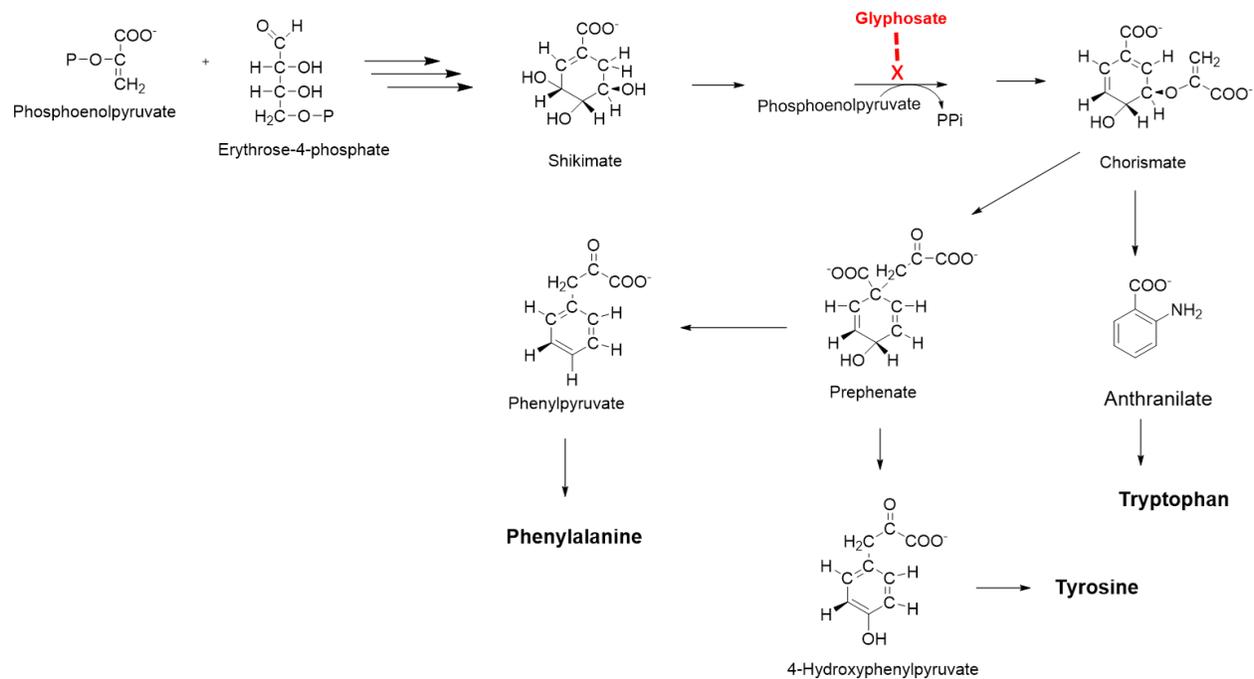


Figure 3.S3. Inhibition of aromatic amino acid biosynthesis. Glyphosate inhibits 5-enolpyruvylshimate-3-phosphate synthase (EPSP) activity, preventing the endogenous production of phenylalanine, tyrosine, and tryptophan.

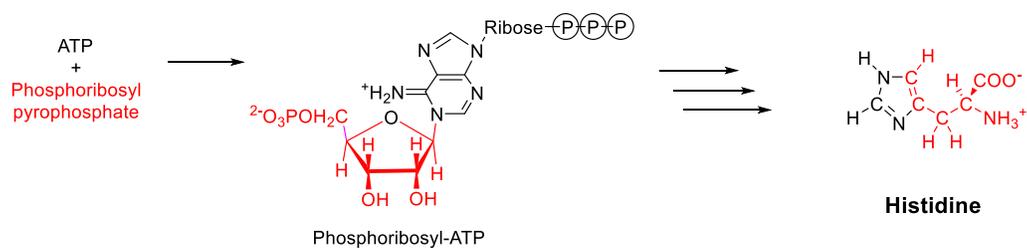


Figure 3.S4. Biosynthesis of histidine starting from ATP and PRPP Precursors.

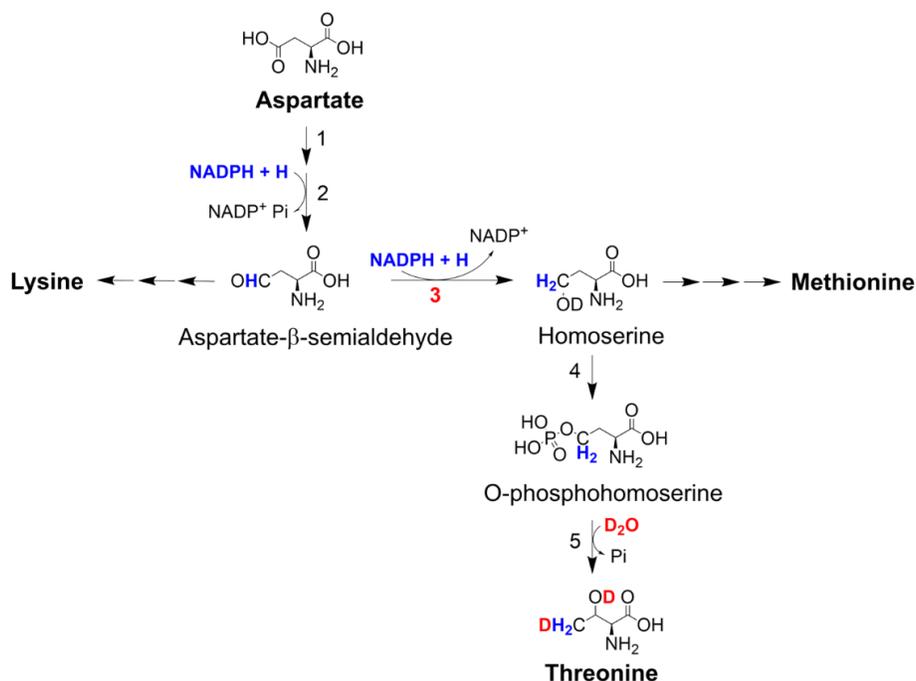


Figure 3.S5. Threonine, methionine, and lysine biosynthesis starting from aspartate. Atoms derived from NADPH (blue), ²H atoms derived from solvent (red). (1) Aspartokinase (2) aspartate-β-semialdehyde dehydrogenase (3) homoserine dehydrogenase (4) homoserine kinase (5) threonine synthase.

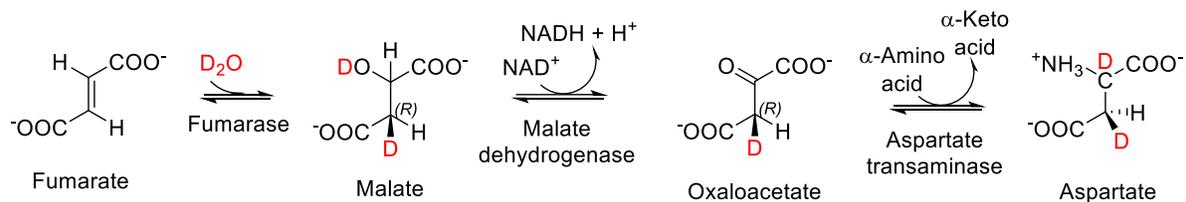


Figure 3.S6. Stereoselective incorporation of ²H from D₂O into the Hβ³ position in aspartate. Fumarase converts unlabeled fumarate supplied in the growth media into malate, with one ²H stereospecifically incorporated from D₂O. Malate is then metabolized within the TCA cycle to oxaloacetate which is converted into aspartate by transamination.

Table 3.S2: Isotope incorporation into Pin1 Amino Acids Monitored by Solution NMR and Mass Spectrometry

Deuteration Analysis of Pin1 Amino Acids by Mass Spectrometry

Amino acid	Atom	Average Peak Intensities (NMR)		Intensity Ratio (SDP/ULP) x 100	Normalized Intensities ((SDP/ULP) x 100)/Peak Intensity of Met H ϵ_1	Deuteration Analysis of Pin1 Amino Acids by Mass Spectrometry			
		Uniformly labeled Pin1 (ULP)	Selectively deuterated Pin1 (SDP)			Amino acid	Specie	Area	Percent deuteration
Gly (2)	H α_1	158669	deuterated			Gly	D0	18273	2
	H α_2	167272	deuterated				D2	715359	98
							Sum	733632	
Ala (3)	H α	182081	157240	86	68	Ala	D0	518396	39
	H β_1	768639	1067240	139	110		D1	385933	29
							D2	47908	4
							D3	106442	8
							D4	259795	20
							Sum	1318474	
Val	H α	Overlap	deuterated			Val	D0	110072	
	H β	161427	deuterated				D2	7953956	83
	H γ_{21}	744840	905119	122	97		D5	527523	5
	H γ_{11}	923552	1146920	124	98		D8	1138821	12
							Sum	9620300	
Leu	H α	48173	deuterated			Leu	Identical retention times with isoleucine		
	H β_2	56138	deuterated						
	H β_3	78996	deuterated						
	H γ	94222	deuterated						
	H δ_{11}	637868	766575	120	95				

Hδ ₂₂	656397	936386	142	113
------------------	--------	--------	-----	-----

Ile

Hα	135195	deuterated		
Hβ	106910	deuterated		
Hγ ₁₂	122074	deuterated		
Hγ ₁₃	91071	deuterated		
Hδ ₁₁	817524	780146	95	75
		CH ₃ , CH ₂ D and CHD ₂		
Hγ ₂₁	631031	Isotopomers	6, 17, and 35	5, 13, and 28

Ile

Identical retention times with leucine

Met (2)

Hα	200586	deuterated		
Hβ ₂	151618	156247	103	82
Hβ ₃	149486	153324	103	82
Hγ ₂	181798	216626	119	94
Hγ ₃	205129	234515	114	90
Hε ₁	1853910	2330125	126	100

Met

D0	577552	10
D1	4933257	90
Sum	5510809	

Thr

Hα	238676	159589	67	53
Hβ	138831	80146	57	45
Hγ ₂₁	776555	1006920	130	103

Thr

D0	679688	16
D1	2171521	53
D2	1185435	29
D3	35671	1
D4	33965	1

				D5	15811	0
				Sum	4122091	
Tyr (2)				Tyr	Digested	
	H α	138360	deuterated			
	H β_2	64623	deuterated			
	H β_3	65660	deuterated			
	H δ_1	298977	596315	199	158	
	H ϵ_1	359718	deuterated			
	H ϵ_2	359718	deuterated			
	H δ_2	298977	596315	199	158	
Phe				Phe		
	H α	86661	deuterated		D0	52673
	H β_2	81314	deuterated		D1	2550
	H β_3	70013	deuterated		D3	190076
	H δ_1	295987	551683	186	D5	45638572
	H ϵ_1	247442	deuterated		Sum	45883871
	H ζ	136071	226668	167		
	H ϵ_2	247442	deuterated			
	H δ_2	295987	551683	186		
Trp (2)				Trp	Digested	
	H α	112715	deuterated			
	H β_2	60616	65188	108		86
	H β_3	71482	61531	86		68
	H δ_1	259967	273954	105		83
	H ζ_2	186879	deuterated			
	H η_2	143305	233073	163		129
	H ζ_3	133287	deuterated			

	H ϵ_3	159286	268975	169	134				
His	H α	128853	164933	128	102	His	D0	5079008	4
	H β_2	73896	86041	116	92		D1	105305730	91
	H β_3	79283	90339	114	90		D2	5336855	5
	H δ_2	193902	302775	156	124		D3	70922	0
	H ϵ_1	170228	227295	134	106		D4	54441	0
							D5	<u>25696</u>	0
							Sum	115872652	
Asp (2)	H α	219528	deuterated			Asp	D0	374397	13
	H β_2	153806	234947	153	121		D1	1862323	65
	H β_3	151777	deuterated				D2	576004	20
							D3	<u>61325</u>	2
							Sum	2874049	
Asn (2)	H α	116366	deuterated			Asn	Converted to aspartate by acid hydrolysis		
	H β_2	115957	156167	135	107				
	H β_3	99332	deuterated						
Lys	H α	191603	deuterated			Lys	D0	148156	5
	H β_2	118767	90643	76	60		D5	106567	4
	H β_3	110930	deuterated				D6	487320	17
	H γ_2	108689	36598	34	27		D7	894943	32
	H γ_3	108689	36598	34	27		D8	885570	31
	H δ_2	169321	98476	58	46		D9	<u>292839</u>	10
	H δ_3	169321	deuterated				Sum	2815395	

H δ_2	83328	46257	56	44
H δ_3	97432	56289	58	46

Gln

H α	overlap	deuterated
H β_2	80714	deuterated
H β_3	80714	deuterated
H γ_2	97799	deuterated
H γ_3	86058	deuterated

Gln Converted to aspartate by acid hydrolysis

* Peak intensities taken from a different selectively labelled sample

(2) Average of two residues

(3) Average of three residues

Table 3.S3: Linewidth measurement for the stereoselectively deuterated aromatic amino acids, methyl groups, and methylene proton of aspartate and asparagine.

NMR standard and amino acid residue	Linewidths (Hz)
DSS-d ₆	1.8
M H ϵ	2.0
Y H δ	2.0
F H δ	2.0
F H ζ	2.8
W H ϵ	2.8
W H η	3.0
W H δ	3.0
N H β	3.0
D H β	3.0
V H γ ^{1and2}	3.0
L H δ ^{1and2}	3.0
T H γ	3.0
A H β	3.0
I H δ	3.0

References

- [1] H. Kotsuki, Hiyoshizo; Datta, Probal Kanti; Hayakawa, Hiroyuki; Suenaga, *Synth.* **1995**, *11*, 1348–1350.
- [2] B. S. Jursic, S. Sagiraju, D. K. Ancalade, T. Clark, E. D. Stevens, *Synth. Commun.* **2007**, *37*, 1709–1714.
- [3] G. I. Danmaliki, P. B. Liu, P. M. Hwang, *Biochemistry* **2017**, *56*, DOI 10.1021/acs.biochem.7b00991.
- [4] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.
- [5] B. A. Johnson, R. A. Blevins, *J. Biomol. NMR* **1994**, *4*, 603–614.

Chapter 4

Proton TOCSY NMR relaxation rates quantitate protein side chain mobility in the Pin1 WW domain

A version of this chapter has been published as Danmaliki, G.I., Hwang, P.M. Proton TOCSY NMR relaxation rates quantitate protein side chain mobility in the Pin1 WW domain. *J Biomol NMR* 76, 121–135 (2022). PMH directed the research, designed the NMR pulse sequence, and acquired the NMR data. GID purified the protein constructs, analyzed the NMR experiments, and wrote the first draft of the manuscript.

Summary

Protein side chain dynamics play a vital role in many biological processes but differentiating mobile from rigid side chains remains a technical challenge in structural biology. Solution NMR spectroscopy is ideally suited for this but suffers from limited signal-to-noise, signal overlap, and a need for fractional ^{13}C or ^2H labeling. Here we introduce a simple strategy in which ^1H relaxation rates are measured during a ^1H TOCSY sequence like DIPSI-2, which can be appended to the beginning of any multi-dimensional NMR sequence that begins on ^1H . The TOCSY RF field compels all ^1H atoms to behave similarly under the influence of strong coupling and rotating frame cross-relaxation, so that differences in relaxation rates are due primarily to side chain mobility. We apply the scheme to a thermostable mutant Pin1 WW domain and demonstrate that the observed ^1H relaxation rates correlate well with two independent NMR measures of side-chain dynamics, cross-correlated ^{13}C relaxation rates in $^{13}\text{C}^\beta\text{H}_2$ methylene groups and maximum observable ^3J couplings sensitive to the χ_1 side chain dihedral angle ($^3\text{J}_{\text{H}\alpha,\text{H}\beta}$, $^3\text{J}_{\text{N},\text{H}\beta}$, and $^3\text{J}_{\text{CO},\text{H}\beta}$). The most restricted side chains belong to Trp26 and Asn40, which are closely packed to constitute the folding center of the WW domain. None of the other conserved aromatic residues is as immobile as the first tryptophan side chain of the WW domain. The proposed ^1H relaxation methodology should make it relatively easy to measure side chain dynamics on uniformly ^{15}N - or ^{13}C -labeled proteins, so long as chemical shift assignments are obtainable.

Introduction

The power of nuclear magnetic resonance (NMR) spectroscopy lies in its ability to describe protein structure and how it changes as a function of time in solution at physiologic temperatures. Proteins are most mobile in their side chains, with a relatively rigid backbone except in the case of long loops and tails. In contrast, there is a high degree of variability between individual amino acid residues with respect to rotation along the C α -C β bond defined by the χ_1 -dihedral angle. Steric considerations heavily favor three major rotamers (gauche+, gauche-, and trans), each pointing the side chain towards one of three corners of a tetrahedron (Figure 4.1a). More than half of the side chains in a protein are mobile with respect to the χ_1 dihedral angle^{1,2}, but which ones are mobile or rigid may not be evident from a structural model derived from conventional X-ray crystallography, cryo-EM, or NMR data. A transition between χ_1 rotamers can dramatically remodel the surface of a protein, potentially uncovering or obliterating a binding site, so it is important to know which side chains are mobile versus rigid.

Traditionally, NMR has seen widespread application of ¹⁵N relaxation, because of its uniformity throughout the protein backbone and ease of magnetically isolating the ¹⁵N nucleus. Discerning side-chain mobility by NMR is technically more challenging because of the diverse and interconnected ¹H and ¹³C spin networks present in the 18 different amino acids with a structurally distinct χ_1 dihedral angle (not glycine or alanine). Previous attempts to examine side-chain relaxation have focused on ²H or ¹³C relaxation. ²H relaxation is appealing because the local quadrupolar relaxation mechanism dominates all others and can be used to map spectral densities that reflect the magnitude and timescale of dynamic fluctuations^{3,4}. The main drawbacks are the need for fractional deuteration and multiple magnetization transfer steps from ¹H to ¹³C to ²H and back again. ¹³C relaxation studies typically use fractional ²H and ¹³C labeling to produce magnetically isolated ¹³C-¹H or ¹³C-¹H₃ groups. Because of the development of labeling methods that produce near-100% ¹³C-¹H₃ labeled methyl groups in an otherwise ¹²C, ²H background, methyl group relaxation has become dominant in ¹³C relaxation studies⁵. Of course, the drawback of this strategy is the lack of information about non-methyl-containing amino acids.

An alternative approach to studying side-chain dynamics is cross-correlated relaxation in ¹³C¹H₂-methylene groups, related to the ratio of signal intensities within the ¹³C¹H₂ multiplet

(triplet)^{6,7}. The beauty of this approach is that it can be applied to a uniformly ¹³C,¹⁵N-labeled sample without deuteration and easily incorporated into conventional 3-D pulse sequences that provide the needed spectral resolution to resolve most signals in the protein. Yang and Kay have demonstrated a reasonable correlation between relaxation parameters obtained using this method and the more rigorous ²H relaxation techniques⁶. The drawback of this approach is the prolonged spin evolution time needed to separate out ¹³C¹H₂ multiplet components, substantially decreasing signal-to-noise relative to the original pulse sequences.

Herein we propose that the measurement of ¹H relaxation rates, while complicated by remote ¹H-¹H interactions, provides the most facile and versatile means for probing side chain dynamics in uniformly ¹⁵N- and/or ¹³C-labeled protein. ¹H relaxation is complex because of multiple ¹H-¹H interactions, mediated via through-bond J couplings and through-space dipolar interactions. Both mechanisms are dependent on the chemical shift frequency difference between interacting ¹H spins. J coupling evolution can proceed via "weak" or "strong" coupling, depending on the magnitude of the J coupling relative to the chemical shift frequency difference. As well, dipolar cross-relaxation for transverse magnetization, rotating-frame Overhauser enhancement (ROE), increases in contribution as the chemical shift frequency difference between interacting sites decreases. Because of these effects, it is preferable to apply a radiofrequency (RF) field so that all ¹H nuclei in the protein behave similarly with respect to these phenomena, thus evolving with strong coupling and maximum ROE contribution. Moreover, since the purpose of the RF field is to achieve uniformity throughout the protein, it is preferable to use a mixing scheme that compensates for off-resonance effects, like DIPSI-2, so that all ¹H nuclei in the protein evolve similarly regardless of their position in the spectrum (Figure 4.1b).

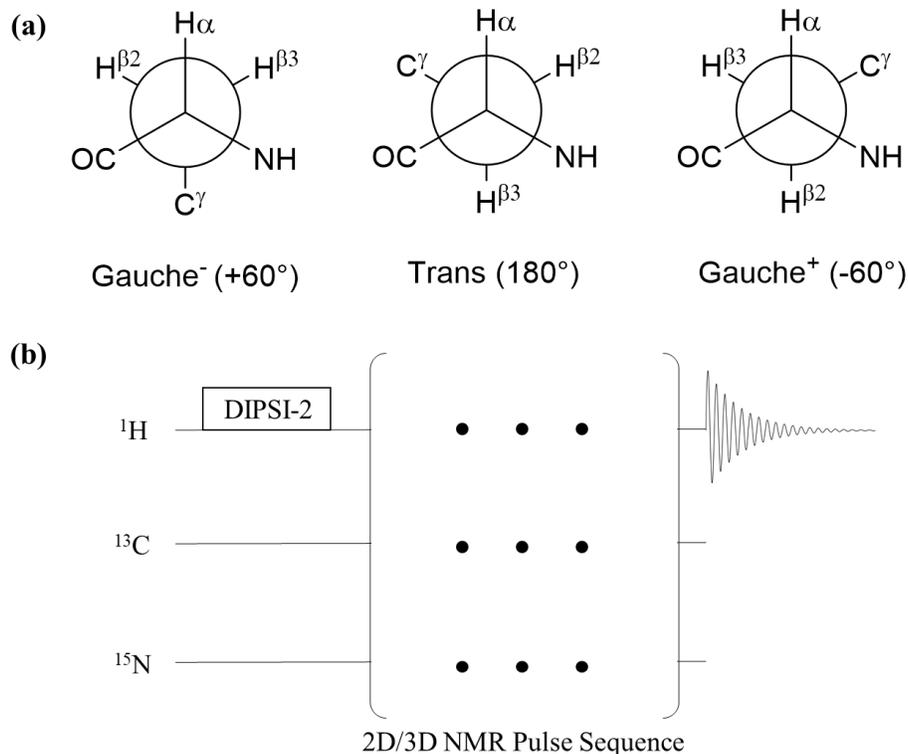


Figure 4.1 (a) The three major rotamers favored by the χ_1 dihedral angle. (b) Pulse scheme used to measure ^1H relaxation rates, using any multi-dimensional sequence preceded by a ^1H TOCSY sequence such as DIPSI-2

In the DIPSI-2 mixing scheme, magnetization starts along the z-axis and is rotated repeatedly clockwise or counter-clockwise around the x-axis by applied RF fields. Then the z-component of the magnetization relax according to

$$\frac{d\Delta I_{z,1}(t)}{dt} = -\rho_{z,1}\Delta I_{z,1}(t) - \sigma_{1,2}^{NOE}\Delta I_{z,2}(t) \quad [1]$$

Where $\Delta I_{z,1}(t) = I_{z,1}(t) - I_{z,1}^0$, and $I_{z,1}^0$ is the equilibrium magnetization of spin-1 along the +z-axis. $\rho_{z,1}$ is the longitudinal auto-relaxation rate for spin-1 and $\sigma_{1,2}^{NOE}$ is the longitudinal cross-relaxation rate for spin-1 as it depends on spin-2. In the slow tumbling limit, where $\omega_0\tau_c \gg 1$,

$$\sigma_{1,2}^{NOE} = -\rho_{z,1} = \frac{-\hbar^2\mu_0^2\gamma_H^4\tau_c}{160\pi^2r_{H_1,H_2}^6} \quad [2]$$

Thus, when spins 1 and 2 have the same direction and magnitude, the longitudinal auto-relaxation term $-\rho_{z,1}\Delta I_{z,1}(t)$ cancels out the cross-relaxation term, $\sigma_{1,2}^{NOE}\Delta I_{z,2}(t)$, which is to say

that longitudinal relaxation effects are very small when all the ^1H nuclei (including H_2O) start off along the +z-axis and evolve similarly under the influence of the DIPSI-2 mixing sequence.

For the transverse components of the magnetization,

$$\frac{dI_{y,1}(t)}{dt} = -\rho_{y,1}I_{y,1}(t) - \sigma_{1,2}^{ROE}I_{y,2}(t) \quad [3]$$

In contrast to the situation for the longitudinal component magnetization, for the transverse magnetization component, the auto- and cross-relaxation terms have the same (rather than opposite) signs. $\sigma_{1,2}^{ROE}$ has a sign opposite to $\sigma_{1,2}^{NOE}$ and double the magnitude:

$$\sigma_{1,2}^{ROE} = \frac{4\rho_{y,1}}{5} = \frac{\hbar^2\mu_0^2\gamma_H^4\tau_c}{80\pi^2r_{H1,H2}^6} \quad [4] \text{ (Bothner-By et al. 1984)}$$

Thus, the effect of cross-relaxation is to approximately double the observed transverse relaxation rate relative to auto-relaxation (rather than canceling out to zero as for longitudinal relaxation)(Bothner-By et al. 1984). During a DIPSI-2 mixing sequence, magnetization is rotated around the x-axis, wherein transverse auto- and cross-relaxation (ROE) dominate the relaxation of ^1H spins.

Within a methylene CH_2 group, the geminal dipolar interaction is at least 8 times stronger than any other ^1H - ^1H interactions due to the r^{-6} distance dependence (1.8 Å distance between geminal protons versus >2.55 Å for vicinal protons). Thus, due to cross-relaxation, geminal proton pairs within a single methylene group would be expected to have very similar relaxation rates. While "mixing" also occurs with respect to more remote proton pairs, the impact of more remote protons is minor compared to the local interactions within a single CH_2 group.

As the spins evolve under the influence of the applied RF field, strong J-coupling also causes magnetization transfer between interacting spins:

$$I_{z,1}(t) \rightarrow \frac{1}{2}(I_{z,1}(1 + \cos 2\pi Jt) + I_{z,2}(1 - \cos 2\pi Jt)) + \frac{1}{2}(2I_{y,1}I_{x,2} - 2I_{x,1}I_{y,2}) \sin 2\pi Jt \quad [5]^9$$

$$I_{z,2}(t) \rightarrow \frac{1}{2}(I_{z,2}(1 + \cos 2\pi Jt) + I_{z,1}(1 - \cos 2\pi Jt)) + \frac{1}{2}(2I_{y,2}I_{x,1} - 2I_{x,2}I_{y,1}) \sin 2\pi Jt \quad [6]^9$$

Similar equations can be written for magnetization components along the x- and y-axes. It will be noted that if $I_{z,1}$ and $I_{z,2}$ are both present at their equilibrium populations at time zero and both spins have the same relaxation rates, then there is no net transfer of magnetization.

However, if the two spins have different relaxation rates, then net magnetization transfer will occur as the magnetization of the two spins begin to diverge in magnitude. As we have pointed out above, the geminal protons in a CH₂ group will have nearly identical transverse relaxation rates due to cross-relaxation, so the majority of magnetization transfer due to strong coupling will be between vicinal protons. The maximum ¹H-¹H ³J coupling observed in proteins is about 14 Hz, so in this case, according to Equations [5] and [6], magnetization would be fully interchanged between two strongly coupled spins by time $1 / (2J) = 36$ ms. Thus, we recommend measuring relaxation rates much shorter than this, and ideally less than 18 ms, when the strong coupling-mediated transfer reaches its maximal rate (if relaxation effects are ignored). Moreover, we recommend only measuring initial relaxation rates (for instance, so that the most rapidly relaxing ¹H signals do not decay to less than half of their starting signal) to further minimize magnetization transfer, so that shorter relaxation delays well below 18 ms should be employed as the size of the protein system examined increases. Thus, the confounding effect of strong coupling becomes less as larger systems are used, though we demonstrate that this methodology is effective in the smallest of protein domains.

To demonstrate the validity of this approach, we measure ¹H R_{DIPSI-2} relaxation rates in the human Pin1 WW domain. We also compare ¹H relaxation rates of β-CH₂ methylene groups with their corresponding 3-bond J couplings that are sensitive to the χ₁-dihedral angle, ³J_{Hα,Hβ}, ³J_{N,Hβ}, and ³J_{CO,Hβ}. The ¹H relaxation rates and J couplings attain maximum values when the protein side chain is rigid, whereas rapid rotation about the χ₁-dihedral angle markedly decreases both. Correlation between these two independent measures of side chain mobility indicates that ¹H relaxation rates, like J couplings, can be a reliable measure of side chain mobility.

We thus propose to measure T_{DIPSI-2} relaxation rates throughout a protein by appending a short DIPSI-2 mixing sequence to the beginning of any pulse sequence in which magnetization starts off on the ¹H nucleus of interest. The subsequent pulse sequence could be any 2D or 3D pulse sequence in which magnetization starts on ¹H (Figure 4.1b). Relaxation rates can be measured by comparing signal intensities of the conventional pulse sequence compared with decreased signal intensities when a short DIPSI-2 element is applied at the beginning.

Materials and Methods

A mutant form of human Pin1 WW domain with N-terminal His-tag was overexpressed in *E. coli* bacteria and purified as previously described, with uniform ^{15}N or $^{15}\text{N},^{13}\text{C}$ isotope enrichment¹⁰. For NMR samples, protein concentration was 1 mM, with 10 mM imidazole, pH 6.7, 100 mM KCl.

Standard Varian VnmrJ Biopack NMR experiments were run at 30°C using a Varian Inova 500 MHz spectrometer equipped with z-axis gradients and room temperature triple resonance probe.

For J coupling measurement experiments, the 3D out-and-back HNHB experiment was performed on an ^{15}N -labeled sample^{11,12}. The intensity of cross-peaks was proportional to $\sin^2(\pi Jt)$, where time $t = 28$ ms for evolution of the $^3J_{\text{N,HB}}$ coupling, and this was used as a numerator in a ratio. A 2D HNHB control experiment in which ^{15}N magnetization is not transferred to HB was used as the denominator. However, we decided not to use the data to calculate exact J couplings due to systematic error in comparing intensities in the 3D experiment with the 2D control. Instead, we elected to keep the ratio of the sine squared cross-peak to the diagonal peak control and present this normalized against the highest value found in the protein (which would correspond to the highest observed J coupling). Errors were estimated as the noise level divided by the cross-peak intensity.

We employed an approach like the HNHB experiment for the HN(CO)HB experiment. Again, we maintained the ratio of the $\sin^2(\pi Jt)$ cross-peak to the diagonal peak control. The time interval used for the evolution of the $^3J_{\text{CO,HB}}$ coupling was 25 ms.

For the $^3J_{\text{H}\alpha,\text{H}\beta}$ coupling, we used a different approach. We used the standard ^1H -TOCSY- ^{13}C -HSQC 3D pulse sequence but modified the TOCSY mixing element from DIPSI-2 to DIPSI-2rc, which is modified to contain delays in which the magnetization is aligned along the z-axis, to cancel out ROE and NOE effects. Unlike the HNHB and HN(CO)HB, which are out-and-back experiments, the ^1H -TOCSY- ^{13}C -HSQC experiment transfers magnetization to vicinal protons via strong coupling so that cross-peak intensity is dependent on $\sin(2\pi Jt)$ rather than $\sin^2(\pi Jt)$. It is not possible to run an equivalent 2D diagonal control spectrum as for HNHB and HN(CO)HB, so the data from this experiment is presented as the maximum observed ratio of $\text{H}\beta_2/\text{H}\beta_3$ or $\text{H}\beta_3/\text{H}\beta_2$ peak intensities.

For $T_{\text{DIPSI-2}}$ experiments, a 5 kHz ^1H DIPSI-2 pulse element was appended to the beginning of the 3D ^1H -TOCSY- ^{15}N -HSQC (22 ms mixing time, ^{15}N -labeled sample), 3D CBCACONH (22 ms mixing time), 2D ^{13}C -HSQC (11 ms mixing time, ^{15}N , ^{13}C -labeled sample in D_2O buffer), or 3D ^1H -TOCSY- ^{13}C -HSQC (11 ms mixing time, ^{15}N , ^{13}C -labeled sample in D_2O buffer). We kept the ^1H carrier frequency on water during the DIPSI-2 sequence, but upon further review, we think it would have been more appropriate to centre it at about 2.5 ppm to ensure better coverage of upfield ^1H signals. A recycle delay of 2 s was used. Peak intensities were compared to the same experiment without the DIPSI-2 pulse element, and the ratio was used to calculate $R_{\text{DIPSI-2}}$ rates. Like the J coupling data, these were also normalized to the fastest $R_{\text{DIPSI-2}}$ rate found in the protein. Errors for the normalized values were estimated based on the express

$$\sqrt{\left(\frac{\Delta_1}{I_1}\right)^2 + \left(\frac{\Delta_2}{I_2}\right)^2}$$

Where I_1 and I_2 are the intensities of the peaks used and Δ_1 and Δ_2 are the noise estimates of the associated spectra.

We recorded a 3D CBCA(CO)NH experiment to measure cross-correlation relaxation rates between ^1H - ^{13}C dipoles of $\text{C}\beta$ methylene groups using the indirect detection of ^{13}C triplet. The relaxation rates were calculated based on the deviation of the triplet intensities from 1:2:1 using the equation below^{6,7,13,14}.

$$\Gamma = -\frac{1}{4T} \ln \frac{\{4 \times I_{\text{out1}} \times I_{\text{out2}}\}}{\{I_c\}^2} \quad [7]$$

where I_{out1} , I_{out2} , and I_c are the intensities of the outer lines 1 and 2 and the central line, respectively, and T is the constant time evolution period for the $^{13}\text{C}\beta$ nucleus (22 ms), set to about $3/(4J)$ for the one-bond C-C J coupling.

Results and Discussion

Pin1 backbone dynamics

The mutant Pin1 WW domain protein used in this study contains 53 amino acid residues. The first 15 residues comprise a His-tag with a flexible Gly- and Ser-rich linker, which is not readily observable in the ^1H - ^{15}N HSQC spectrum due to solvent-amide exchange and signal

overlap. The native sequence begins at residue 16 in our construct. The structure comprises three antiparallel β -strands, with residues 22-53 visible in the X-ray crystal structure (Figure 4.2)¹⁵. In agreement with the crystal structure, ^{15}N T_1 and T_2 (from $T_{1\rho}$) times indicate that the backbone of these residues is rigid throughout the WW domain, with significant mobility seen only N- and C-terminal to the structured core (residues 16-21 and 53) (Figure 4.S1). The similarity of backbone ^{15}N T_1 and T_2 values indicate that this small protein (30-residue structured core) tumbles in the extreme narrowing regime at 30 °C, wherein the spectral densities $J(0)$ and $J(\omega_N)$ are of similar magnitude, as opposed to the slow tumbling regime in which $J(0)$ dominates.

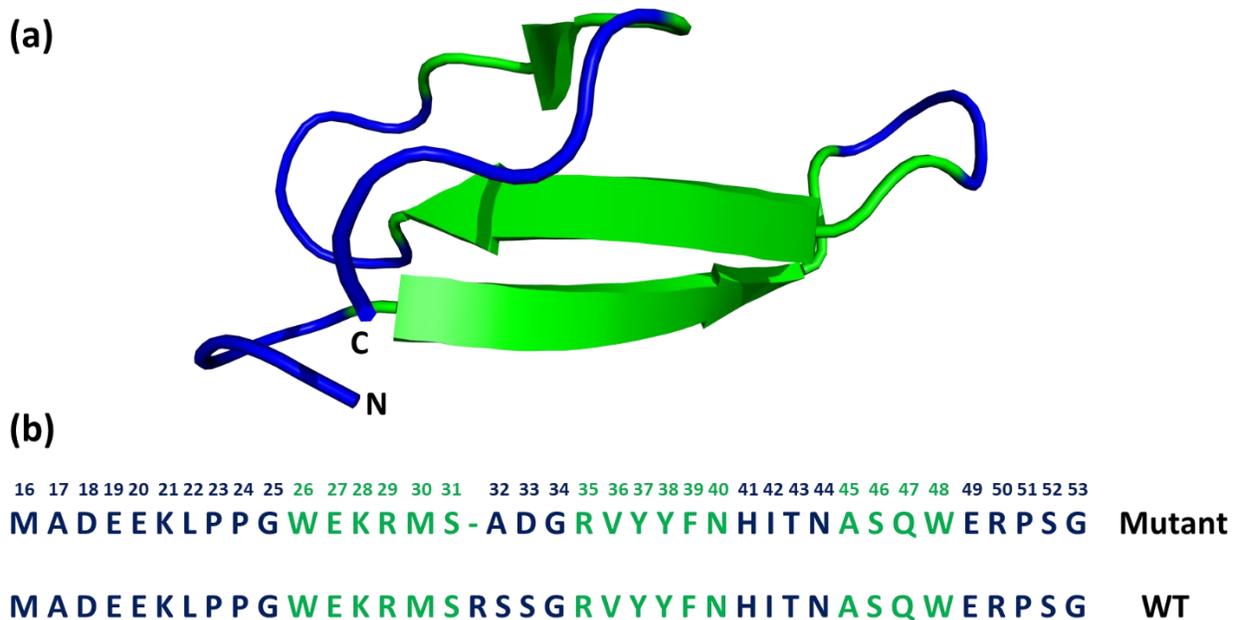


Figure 4.2 (a) Crystal structure of the human Pin1 protein rotated along its long axis showing the antiparallel β -strands in green (PDB id 2F21). (b) Sequence of the mutant Pin1 used in this study versus wild type. The antiparallel β -strands residues are highlighted in green

Table 4.1. Chemical shifts and normalized values for maximum observed ^3J -couplings and ^1H R_{DIPS12} relaxation rates for the mutant human Pin1 WW domain. $\text{C}\beta$ cross-correlated relaxation rates have not been normalized. In the last column, the predominant rotamer is determined by comparing $\text{H}\beta_2$ with $\text{H}\beta_3$ intensities in the NMR experiments used to derive J couplings, but this is not always possible if the NMR signals are overlapped or degenerate. We label the residue “mobile” or “restricted” in such cases based on the available relaxation data. Generally speaking, “mobile” residues have normalized values <0.55 , while “restricted” residues have normalized values >0.55 . The value is left blank if the residues has no β CH_2 group (A,G,I,V,T) or if there was insufficient data.

Residue	Atom	Chemical shifts (ppm)	$^3\text{J}_{\text{H}\alpha\text{H}\beta}$	$^3\text{J}_{\text{HNH}\beta}$	$^3\text{J}_{\text{HN}(\text{CO})\text{H}\beta}$	^1H R_{DIPS12} from ^{15}N TOCSY	^1H R_{DIPS12} from ^{13}C TOCSY	^1H R_{DIPS12} from CBCACONH	$\Gamma_{\text{C}\beta\text{C}\alpha(\text{CO})\text{NH}}$ (s^{-1})	Predominant Rotamer
M16	$\text{H}\beta_2/\text{H}\beta_3$	2.05/2.05	-	-	-	0.39	0.31	0.28	-	Mobile
A17	$\text{H}\beta$	1.42	^a	-	-	0.28	0.16	0.18	-	
D18	$\text{H}\beta_2/\text{H}\beta_3$	2.64/2.74	0.42	0.35	0.34	0.31	0.29	0.34	-1.46	Mobile
E19	$\text{H}\beta_2/\text{H}\beta_3$	2.08/1.99	-	0.40	-	0.31	0.30	0.35	-2.17	Mobile
E20	$\text{H}\beta_2/\text{H}\beta_3$	2.03/1.83	-	-	-	0.42	-	0.31	-1.13	Mobile
K21	$\text{H}\beta_2/\text{H}\beta_3$	1.87/1.79	-	0.33	-	0.31	0.13	0.42	-1.66	Mobile
L22	$\text{H}\beta_2/\text{H}\beta_3$	1.45/1.85	^a	^b	-	0.74	0.71	-	-	Restricted
P23	$\text{H}\beta_2/\text{H}\beta_3$	2.63/2.02	0.45	-	-	-	0.79	-	-	Restricted
P24	$\text{H}\beta_2/\text{H}\beta_3$	1.91/2.36	0.39	-	-	-	0.60	0.53	-4.42	Restricted
G25	$\text{H}\alpha_2/\text{H}\alpha_3$	4.05/3.33	-	-	-	0.62	0.53	0.68	-7.20	

W26	$H\beta_2/H\beta_3$	3.27/2.99	1.00	1.00	-	0.89	0.82	0.88	-10.64	Gauche+
E27	$H\beta_2/H\beta_3$	2.22/2.32	-	0.40	0.47	0.45	0.30	0.46	4.06	Mobile
K28	$H\beta_2/H\beta_3$	1.78/1.66	0.50	0.54	0.57	0.26	0.54	0.58	-2.53	Averaged between Trans and Gauche+
R29	$H\beta_2/H\beta_3$	0.09/1.32	0.70	0.79	0.27	0.69	0.70	0.78	-7.64	Gauche+
M30	$H\beta_2/H\beta_3$	1.98/1.89	0.38	0.41	0.50	0.16	0.28	0.37	0.30	Mobile
S31	$H\beta_2/H\beta_3$	4.53/4.23	-	-	-	-	0.22	-	-	Mobile
A32	$H\beta$	1.53	-	-	-	-	0.33	0.31	-	
D33	$H\beta_2/H\beta_3$	2.98/2.67	-	0.43	0.87	-	0.54	0.53	-5.85	Gauche-
G34	$H\alpha_2/H\alpha_3$	4.23/3.79	-	-	-	0.67	0.52	0.59	-5.34	
R35	$H\beta_2/H\beta_3$	1.96/2.05	-	-	-	0.32	-	0.63	-	Mobile
V36	$H\beta$	2.01	-	^b	-	0.23	0.24	0.38	-	
	$H\gamma_{11}/H\gamma_{21}$	0.81/1.07					0.31/0.15			
Y37	$H\beta_2/H\beta_3$	2.46/2.79	0.39	0.89	0.95	-	1.00	1.00	-8.76	Gauche-
Y38	$H\beta_2/H\beta_3$	2.69/2.94	0.88	0.90	-	0.75	0.73	0.93	-6.40	Gauche+

F39	H β_2 /H β_3	2.89/2.61	0.60	0.31	0.94	0.84	0.93	0.78	-7.98	Trans
N40	H β_2 H β_3	-0.06 2.00	1.00	0.31	1.00	1.00	0.65	0.84	-6.97	Trans
H41	H β_2 /H β_3	3.30/3.08	-	0.81	^c	-	-	0.62	-2.15	Gauche-
I42	H β H γ_{21} /H δ_{11}	2.02 0.79/0.76	^a	^b	-	-	0.46 0.30/0.12	0.42		
T43	H β H γ_{21}	4.25 0.96	^a	0.84	^b	-	0.40 0.26	0.28	-	
N44	H β_2 /H β_3	3.14/2.92	0.51	0.69	0.44	0.48	0.28	0.37	0.54	Averaged between Trans and Gauche+
A45	H β	1.26	^a	-	-	0.22	0.26	0.26	-	
S46	H β_2 /H β_3	3.84/3.79	-	0.46	0.40	0.43	0.67	0.47	0.54	Mobile
Q47	H β_2 /H β_3	2.24/2.56	-	0.62	0.79	-	-	0.68	-12.02	Gauche-
W48	H β_2 /H β_3	3.22/3.67	0.82	0.92	0.30	0.70	0.62	0.70	-9.03	Gauche+
E49	H β_2 /H β_3	1.91/1.91	-	0.68	-	0.49	0.36	-	-	Mobile
R50	H β_2 /H β_3	1.45/1.45	-	-	-	-	0.60	-	-	

P51	H β_2 /H β_3	0.84/0.63	0.61	-	-	-	0.47	0.49	-2.87	Mobile
S52	H β_2 /H β_3	3.81/3.75	0.36	0.30	0.45	0.49	0.35	0.26	-1.70	Mobile
G53	H α_2 /H α_3	3.78/3.78	-	-	-	0.11	0.15		-	

^a Single H β peak, but no second H β peak for quantitative 3J coupling estimation.

^b Peak present in HNHB spectrum, but the overlapped signal in 2D reference spectrum prevents quantitative 3J coupling measurement.

^c Peak present in HNCOHB spectrum but overlapped signal in 2D reference spectrum prevents quantitative 3J coupling measurement.

^1H $R_{\text{DIPSI-2}}$ relaxation rates

We measured initial ^1H relaxation rates during a DIPSI-2 time interval for the Pin1 WW domain, with normalized rates shown in Table 4.1.

The 3D ^1H -TOCSY- ^{15}N -HSQC experiment yielded ^1H R_{DIPSI2} relaxation rates for most $\text{H}\beta$ groups in the Pin1 WW domain. The experiment suffers from variable signal-to-noise depending on the size of the 3-bond ^1H J couplings used to relay magnetization from $\text{H}\beta$ to the backbone HN, smallest for the side chain χ_1 gauche⁻ rotamer and for helical backbone ϕ dihedral angles. Fortunately, in the small β sheet WW domain, signal-to-noise was sufficient for most residues. However, for larger helical proteins, pulse sequences that transfer magnetization via ^{13}C - ^{13}C couplings would be preferable. ^1H R_{DIPSI2} relaxation rates vary substantially as one goes through the sequence of the Pin1 WW domain (Table 4.1, Figure 4.S2a). The most rapid ^1H R_{DIPSI2} relaxation rate observed in the 3D ^1H -TOCSY- ^{15}N -HSQC belongs to Asn40 $\text{H}\beta_{2/3}$, with a relaxation rate of 28.8 s^{-1} , corresponding to a time constant of 34.7 ms. For comparison purposes, this rate was assigned a value of 1, and all other rates observed in the protein were normalized to this value. The $\text{H}\beta$ protons of Asn40 have unique and divergent chemical shifts, consistent with a rigid side chain. Additional residues with rapid ^1H R_{DIPSI2} relaxation rates include the following: Leu22 (0.74), Trp26 (0.89), Arg29 (0.69), Tyr38 (0.75), Phe39 (0.84), and Trp48 (0.70). These residues possess a rigid backbone structure, as indicated by the backbone ^{15}N relaxation (Figure 4.S1). Consistent with expectations, hydrophobic aromatic residues (F,W,Y) possess some of the fastest relaxation rates. Aromatic rings frequently make important contacts with both hydrophobic and polar residues. Changes in the side chain χ_1 dihedral angle drastically re-position the bulky inflexible aromatic ring, making transitions less likely. Interestingly, all the rigid non-aromatic residues possess $\text{H}\beta_2$ chemical shifts that are highly divergent from $\text{H}\beta_3$, whereas mobile residues have more similar chemical shifts. It is likely that divergent methylene chemical shifts are a specific indicator for side chain rigidity, whereas similar methylene chemical shifts are not necessarily specific for mobility.

Glycine residues are unique in that they have a methylene CH_2 group at the α position. According to backbone ^{15}N relaxation (Figure 4.S1), Gly25 and Gly34 residues tumble like the rest of the WW domain, whereas Gly53 at the C-terminus is flexible. Gly25 and Gly34 have normalized ^1H R_{DIPSI2} relaxation rates slower than the most rigid C^βH_2 groups, 0.62 and 0.67,

respectively. This is reflective of there being fewer vicinal ^1H - ^1H dipolar interactions as well as some increased mobility at the Gly positions, as gauged by ^{15}N relaxation. The ^1H R_{DIPS12} relaxation rates at non-glycine $\text{H}\alpha$ positions are generally slower than for $\text{H}\beta$ methylene groups due to the absence of short-range dipolar ^1H - ^1H interactions found in methylene groups (Table 4.S1). Relaxation rates at $\text{H}\alpha$ depend on vicinal and more distant ^1H - ^1H interactions that rely on backbone and side chain dihedral angles, making them a less reliable indicator of backbone dynamics than ^{15}N relaxation.

^1H R_{DIPS12} relaxation rates were also measured using the 3D ^1H -TOCSY- ^{13}C -HSQC sequence, as shown in Table 4.1 and Figure 4.S2b. Tyr37 had the fastest relaxation rate of 32.6 s^{-1} , and we normalized other relaxation rates in the protein against this value. Consistent with the 3D ^1H -TOCSY- ^{15}N -HSQC experiment, the same residues display rapid relaxation rates at the $\text{H}\beta$ position: Leu22 (0.71) Trp26 (0.82), Arg29 (0.70), Tyr38 (0.73), Phe39 (0.93), Asn40 (0.65), and Trp48 (0.62). Additional sites in the protein are also accessible to this NMR experiment, suggesting some rigidity in the following residues as well: Pro23 (0.79), Pro24 (0.60), Tyr37 (1.00), and Arg50 (0.60). There are many more NMR signals obtainable via the 3D ^1H -TOCSY- ^{13}C -HSQC experiment. For instance, proline residues are only accessible via the ^{13}C experiments, as are additional sites further down the larger side chains. For the most part, the ^1H relaxation rates along long side chains line up well with expectations, with slower relaxation rates observed as one moves further away from the backbone (Table 4.S1).

We initially hoped that the 3D ^1H -TOCSY- ^{13}C -HSQC experiment would yield better data than the 3D ^1H -TOCSY- ^{15}N -HSQC experiment because of superior signal-to-noise. However, spectral distortions arising from the water peak and the ^1H - ^1H diagonal make data obtained from the ^1H -TOCSY- ^{13}C -HSQC experiment less reliable. It was possible to obtain ^1H relaxation data for any ^1H atom via a diagonal peak or cross-peaks wherein magnetization was transferred to geminal or vicinal protons. The strongest signal-to-noise, often by an order of magnitude, is seen in the diagonal peaks, but we have found that data from these diagonal peaks are unreliable compared with that from cross-peaks or other data presented in this manuscript. Thus, for ^1H relaxation rates derived from the 3D ^1H -TOCSY- ^{13}C -HSQC experiment, we used the cross-peak with the strongest intensity if there was one above a minimum threshold. We averaged the value

with other cross-peaks within 50% of that intensity, with the standard deviation of those values used for error bar estimation (see Figure 4.S2b).

Figure 4.S2c compares the relaxation rates derived from 3D ^{15}N -TOCSY-HSQC against the ^{13}C TOCSY-HSQC experiment. It is important to note that there were some discrepancies between the ^1H R_{DIPS12} relaxation rates measured using the 3D ^1H -TOCSY- ^{13}C -HSQC versus the ^1H -TOCSY- ^{15}N -HSQC experiment, most notably for Lys28 H β (0.54 vs. 0.26), Asn40 H β (0.65 vs. 1.00), Asn44 (0.28 vs. 0.48), and Ser46 H β (0.67 vs 0.43). We suggest that for these discrepancies, the values obtained from the 3D ^1H -TOCSY- ^{13}C -HSQC experiment are less reliable due to spectral distortions, as discussed above.

Table 4.1 also shows $^{13}\text{C}\beta$ R_{DIPS12} relaxation rates measured from the CBCACONH experiment. The most rapid C β to relax in Pin1 belongs to Tyr37 (32.6 s^{-1}), and we normalized other values found in the protein against it (Figure 4.S2c). Consistent with both ^1H relaxation experiments, same residues Trp26, Arg29, Tyr38, Phe39, Asn40, and Trp 48 display rapid R_{DIPS12} relaxation rates of 0.88, 0.78, 0.93, 0.78, 0.84, and 0.70. Importantly, the CBCACONH experiment was unaffected by spectral distortions observed in the ^{13}C TOCSY-HSQC experiment, making data obtained from this experiment reliable. We also measured C α R_{DIPS12} relaxation rates (Table 4.S1). We compare spectra from the CBCACONH experiment in Figure 4.3 to demonstrate the effect of the DIPS12 mixing scheme on a rigid residue, Y37, and a flexible residue, N44. After 22 ms, the C α and C β intensities of Y37 decreased by $\sim 33\%$ and $\sim 50\%$, whereas both in N44 reduced by $\sim 22\%$.

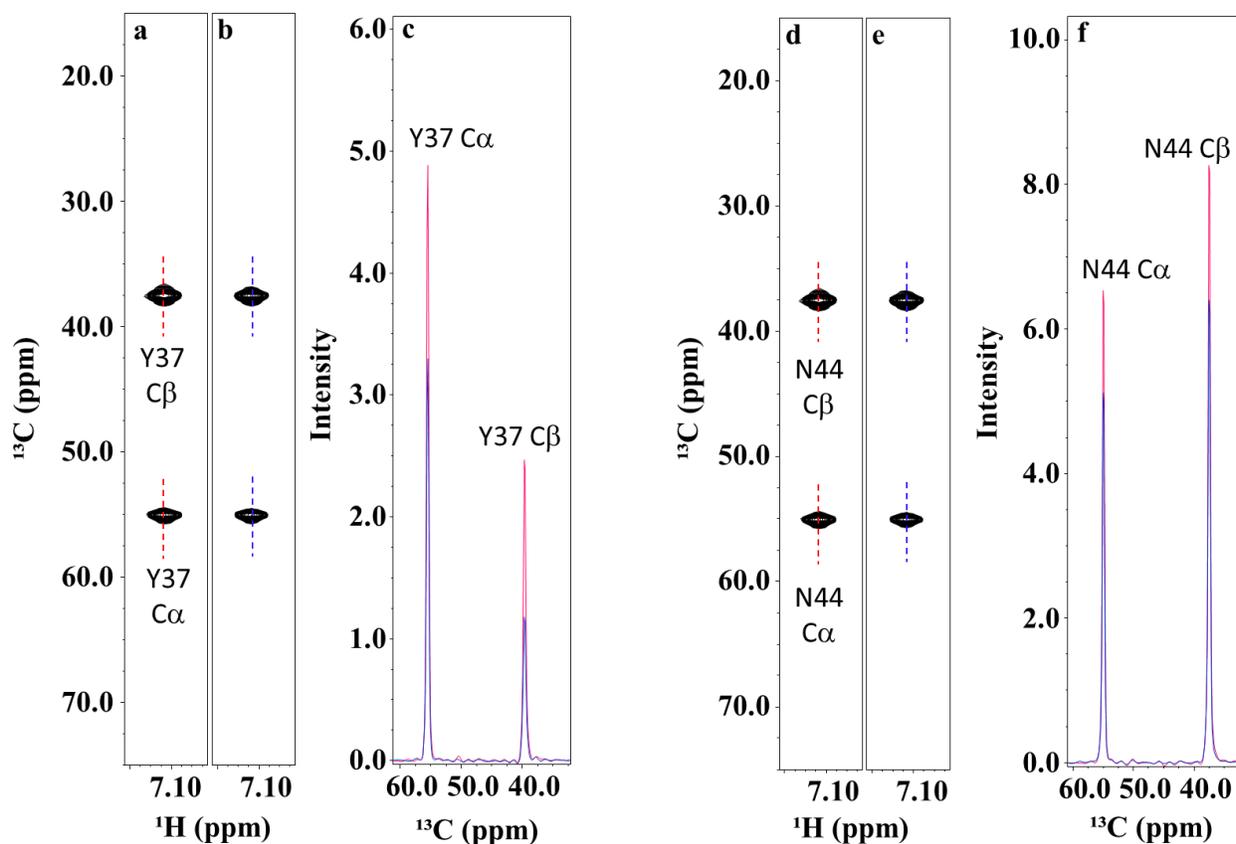


Figure 4.3 Effect of the DIPSI2 mixing scheme on rigid Y37 and a rotamer-averaging N44. CBCACONH strip plots taken at 0 ms (a), (d), and the corresponding region at 22 ms (b), (e). Cross-sections of the peaks highlighted in the strip plots are shown in (c) and (f).

Cross-correlated relaxation rates of Pin1 residues

Table 4.1 also displays cross-correlated relaxation rates Γ derived from a CBCA(CO)NH experiment without ^1H decoupling during ^{13}C evolution in the indirect ^{13}C dimension. Γ values range from $+4.06\text{ s}^{-1}$ for flexible residues like Glu27 down to -12.02 s^{-1} for Gln47 (-7.20 for Gly25 H α - note that only glycine residues have comparable cross-correlated relaxation at the α position in the CBCACONH experiment). The methylene groups of the rigid residues: Trp26, Arg29, Tyr37, Tyr38, Phe39, Asn40, Trp48 show cross-correlated relaxation rates of -10.64 , -7.64 , -8.76 , -6.40 , -7.98 , -6.97 , and -9.03 , while the flexible Glu27, Met30, Asn44, and Ser46 exhibit positive rates of $+4.06$, 0.30 , 0.54 , and 0.54 . It is interesting to note that in the original CBCACONH cross-correlated relaxation experiment described by Kay and co-workers, performed on drk SH3 domain, there were no positive cross-correlated relaxation rates,

indicating that the outer components of the ^{13}C triplet have slower relaxation rates than the central line. This would be expected for the case where there is no internal side chain mobility or limited side chain mobility. However, the presence of positive cross-correlated relaxation rates confirms the presence of very fast side chain rotamer transitions between the major rotamers, faster than the overall tumbling of the WW domain, so that the magnetic fields of anti-parallel ^1H spins in the central methylene ^{13}C transition destructively interfere, instead of the constructive interference one would expect based on the tetrahedral 109° angle between ^{13}C - ^1H dipoles. One would expect that these rapid transitions occur in all proteins and not just in the WW domain as we have observed, except that in larger proteins (which includes all folded domains), the $J(0)$ spectral density contributed by overall tumbling as described by the order parameter S^2 becomes more dominant, so that only negative cross-correlated relaxation rates are observed, even for a domain as small as an SH3 domain (though it is about double the size of the WW domain). Interestingly, Zheng and Yang observe positive cross-correlated relaxation rates in the flexible lysine side chains of intestinal fatty acid binding protein (Zheng and Yang 2004). For comparison to ^1H R_{DIPSI2} relaxation rates, we converted the values to a linear scale of 0 (for Glu27) to 1 (for Gln47) in Figure 4.S3.

χ_1 dihedral angle estimated using maximum observed ^3J couplings

Accurate measurements of 3-bond scalar coupling constants from $^3\text{J}_{\text{H}\alpha\text{-H}\beta}$, $^3\text{J}_{\text{N-H}\beta}$, and $^3\text{J}_{\text{CO-H}\beta}$ experiments are critical for stereospecific chemical shift assignment of β -methylene protons and determination of the side chain χ_1 rotameric state. J couplings also provide an alternative independent method to assess the validity of using relaxation measurements to gauge conformational dynamics.

For measuring J couplings, we used quantitative methods that transfer magnetization from ^{15}N or ^{13}CO to $\text{H}\beta_2$ and $\text{H}\beta_3$ and then back. For instance, the three-bond J coupling $^3\text{J}_{\text{N-H}\beta}$ is measured in the HNHB experiment by comparing the intensity ratio of the cross-peaks of the N-H β correlations in the 3D spectrum to the diagonal N correlation in the 2D reference spectrum (similar amount of time spent on the ^{15}N nucleus, but ^1H - ^{15}N couplings refocused):

$$\frac{I_{\text{N-H}\beta}}{I_{\text{N}}} = \sin^2(\pi J t)$$

where t is the time spent evolving the ^{15}N magnetization to $\text{H}\beta_{2/3}$, with a second identical interval, t , evolving the magnetization from $\text{H}\beta_{2/3}$ back to ^{15}N . Since the couplings reported in the literature for N-H β or CO-H β are $< 11 \text{ Hz}^{11,12,16}$, we simulated the intensity ratio for $0 < J < 11 \text{ Hz}$. Figure 4.4a shows a parabolic correlation comparing the $\sin^2(\pi Jt)$ term to J . By applying a square root, $\sqrt{\sin^2(\pi Jt)}$, an approximately linear correlation can be achieved for this range of Jt values, because $\sin \pi Jt \sim \pi Jt$ for small values of πJt . Thus, we use $\sqrt{\sin^2(\pi Jt)}$ as proportional to the J coupling value and normalize the result against the highest value found in the protein. We chose not to estimate absolute values for the J couplings because this would introduce systematic error in comparing the 3D HNHB experiment with its 2D reference spectrum, and we are mainly interested in the normalized values for this study.

According to the Karplus equation, 3-bond J couplings achieve their maximum value when the interacting nuclei are oriented trans to each other and take on smaller values when they are at 60° (corresponding to the other major rotameric states). Since the accuracy of our J couplings measurements depends on signal intensity, we chose to focus on the maximum observable J coupling. The maximum observable values in the protein correspond to rigid side chains in which interacting nuclei are fixed in a trans orientation. With increasing mobility, the nuclei “wiggle” away from the trans position or jump to other rotameric states, decreasing the maximum observed J coupling, down to 33% in theory, assuming a minimum J coupling of around 0 Hz and complete averaging about the 3 major rotameric states.

For $^3J_{\alpha\beta}$ couplings, we employed a ^1H DIPSI-2rc TOCSY- ^{13}C -HSQC sequence to relay magnetization from $\text{H}\beta$ to $\text{H}\alpha$ and then to HN , in which delays are introduced into the DIPSI-2 pulse train while the magnetization is oriented along the z-axis in such a way that NOE and ROE cancel out¹⁷. However, because this is not an out-and-back experiment, there is no 2D reference experiment. Thus, we used a different approach to estimating relative J -couplings by taking the larger of the cross-peak intensities (corresponding to $\text{H}\beta_2$ or $\text{H}\beta_3$) divided by the lesser intensity. The most rigid side chains give rise to the largest quotients, which tend towards 1 for rapid rotameric averaging. The $^3J_{\alpha\beta}$ couplings previously reported in the literature range between $< 4 \text{ Hz}$ for gauche substituents and $> 10 \text{ Hz}$ for trans substituents, with an intermediate 6-8 Hz for disordered side-chains in equilibrium between two or three of the major rotamer conformations¹⁸. Since the cross-peak intensity is dependent on $\sin(2\pi Jt)$, we simulated the

quotient $\sin(2\pi Jt)_{max} / \sin(2\pi Jt)_{min}$ versus J within the range $6 < J < 14$ Hz, corresponding to rotameric averaging between 3 states, with the trans state having a theoretical J coupling of 14 Hz (max) and the gauche states, 2 Hz (min) (Figure 4.4b). Taking the square root of the quotient $\sin(2\pi Jt)_{max} / \sin(2\pi Jt)_{min}$ converts an exponential-like function into a roughly linear correlation, suggesting the such a term would be roughly proportional to 3J -couplings.

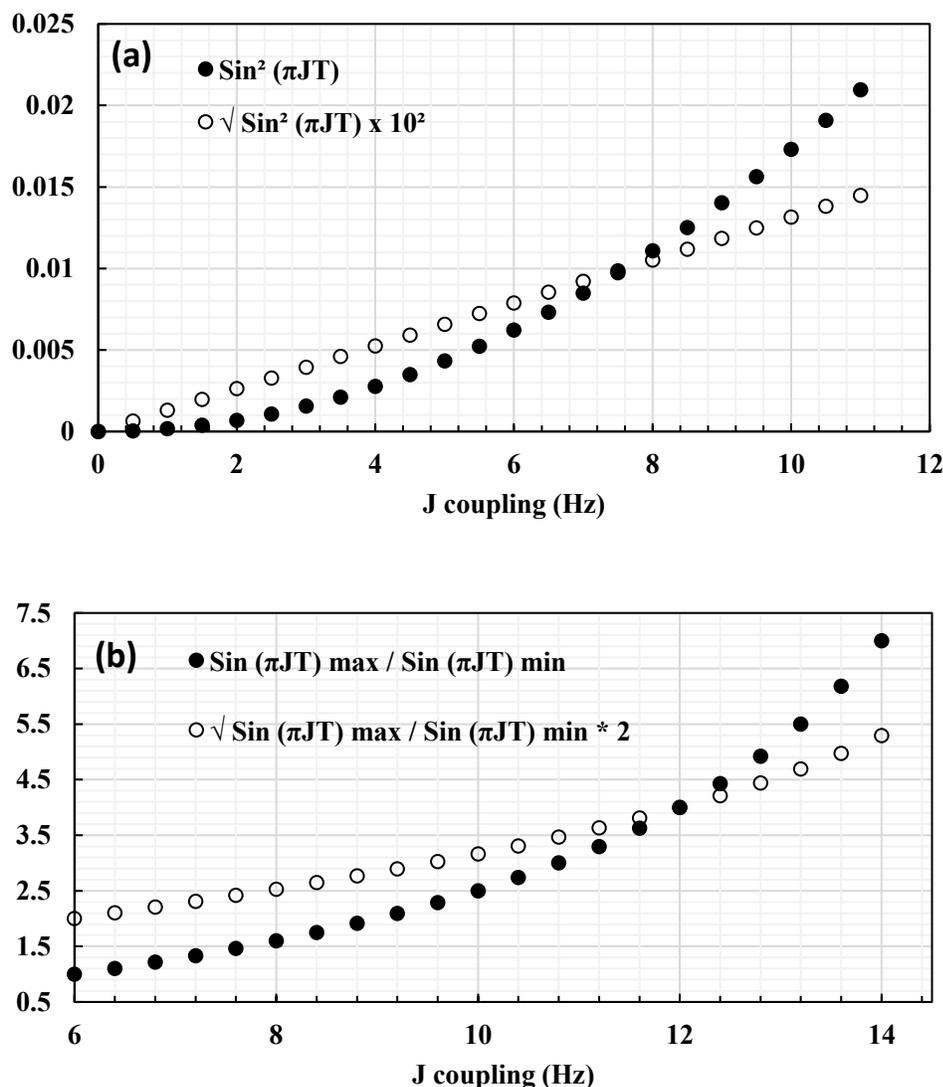


Figure 4.4 Correlation plot of simulated cross-peak intensity ratio against 3J couplings of HNHB/HN(CO)HB experiment (a) and H α H β experiment (b). In both plots, taking the square root of the simulation converts an exponential to a linear correlation, providing a reasonable estimate of J -couplings

The estimation of maximum measured 3J -coupling for Pin1 residues

We thus obtained measures proportional to $^3J_{N,H\beta}$, $^3J_{CO,H\beta}$, and $^3J_{\alpha,\beta}$ couplings and normalized them against the maximum values observed in the protein (Table 4.1 and Figure 4.S4). Residues with a single dominant χ_1 -rotamer are expected to display large couplings in two of the three J coupling measurement experiments, and the normalized couplings should be the same in theory (though for $^3J_{\alpha\beta}$ compared to the others, this is only approximate). We excluded residues for which we could not obtain precise measurements due to signal overlap or poor signal-to-noise. Figure 4.S5 shows the normalized maximum 3J -coupling plotted against the human Pin1 sequence for residues with a complete data set. Consistent with the 1H relaxation experiments, Trp26 and Asn40 residues possessed the maximum observed 3J values (normalized to 1.0), confirming that the side chains of these residues are the most immobile in the entire protein.

Based on the J-couplings, we can differentiate residues into three groups: 1) rigid residues with a single dominant χ_1 dihedral angle conformation (Trp26, Arg29, Asp33, Tyr37, Tyr38, Phe39, Asn40, His41, Gln47, and Trp48) as highlighted in Figure 4.5; 2) residues averaging between two χ_1 dihedral angle rotamers (Lys28 and Asn44); and 3) flexible residues averaging equally between all three χ_1 dihedral angle rotamers (Met16, Asp18, Glu19, Glu20, Lys21, Glu27, Arg35, Ser46, Glu49, Ser52) (Table 4.1). Excluded from these groups are residues with no χ_1 dihedral angle (Gly, Ala), residues without a β -methylene group (Ala, Thr, Ile, and Val), and any residues lacking sufficient data (Pro and others).

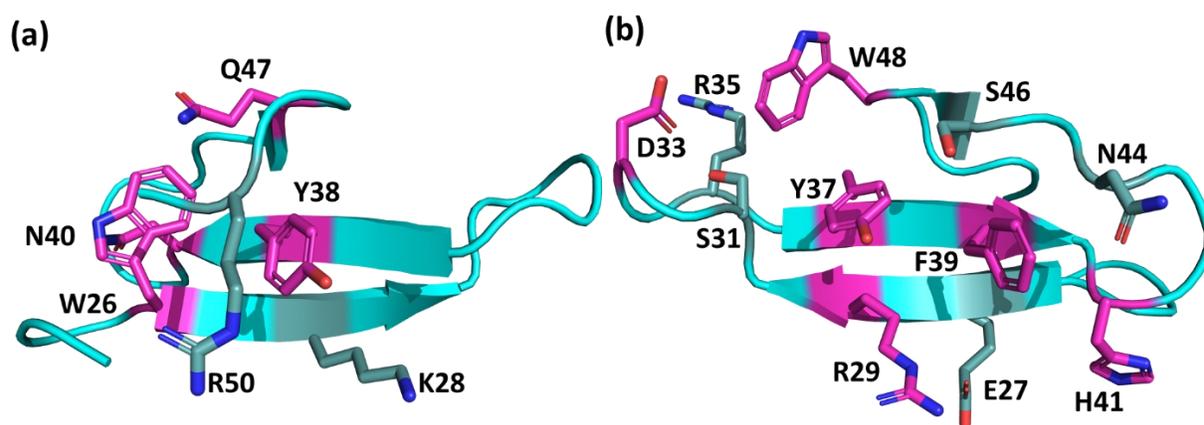


Figure 4.5 The crystal structure of the human Pin1 protein shows side chains for which NMR J-coupling experiments have defined a dominant χ_1 dihedral angle. The side chains of Trp26, Tyr38, Asn40, and Gln47 are clustered on one side of the sheet as shown in (a), while side chains

of Arg29, Asp33, Tyr37, Phe39, His41 and Trp48 constitute a surface on the other side as shown in (b). Flexible side chains interacting with the rigid residues are highlighted in light teal

Rigid Pin1 residues adopting a single dominant χ_1 dihedral angle rotamer

According to ^3J coupling experiments, the most restricted side chains in the Pin1 WW domain belong to Trp26 and Asn40. The side chain of Asn40 is packed against the plane of the Trp26 indole ring, giving Asn40 the most unique and upfield-shifted $\text{H}\beta$ chemical shifts in the entire protein (-0.60 and 2.00). The side chain amide HN of Asn40 hydrogen bonds with the π -cloud above the plane of the Trp26 indole ring nitrogen¹⁹, as shown in Figure 4.6. The other HN forms a hydrogen bond with the backbone carbonyl of Pro24. The side chain amide oxygen of Asn40 additionally forms two strong hydrogen bonds with the backbone amide HN of Ile42 and Thr43. Thus, the side chain of Asn40 forms no fewer than four hydrogen bonds, 3 of which are with the backbone. The backbone of Asn40 hydrogen bonds with the backbone HN of Asn44 and Ala45, forming an unusual 5-residue short loop structure that connects the second and third β -strands. The Trp26-Asn40 dyad, sitting on the first two β -strands, constitutes the central folding core of the WW domain (see Figure 4.6). The indole side chain of Trp26, which belongs to the first β -strand, forms packing interactions with Gln47 of the third β -strand, and Pro23 and Pro51 from the N- and C-terminal tails of the WW domain. Thus, all three β -strands and the N-terminal and C-terminal tails of the small WW domain appear to converge around the first tryptophan side chain for which the domain is named. Thus, the maximum χ_1 -sensitive ^3J couplings observed in the protein highlight the Trp26-Asn40 dyad that defines the folded core of the Pin1 WW domain (Figure 4.6)

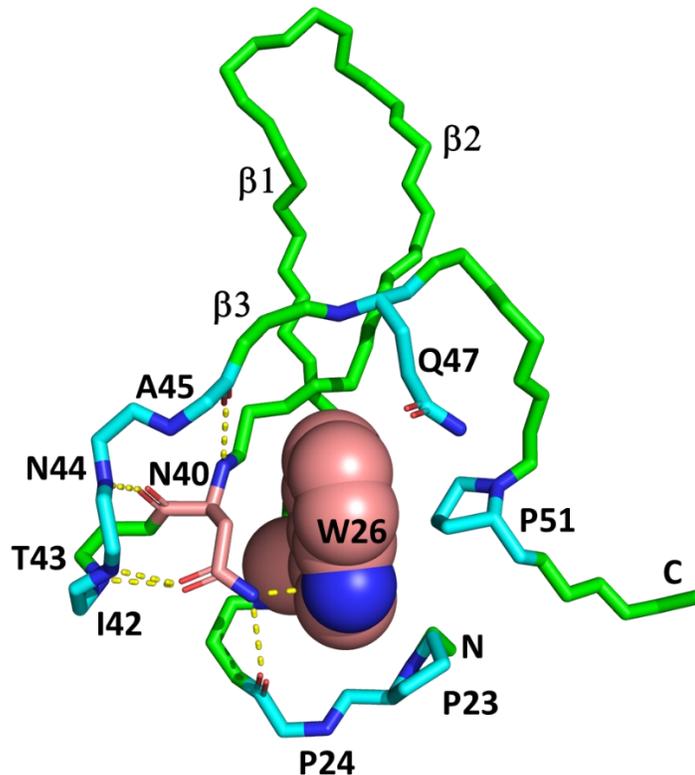


Figure 4.6 The folding core of the human Pin1 WW domain shows the buried residues: Trp26 (sphere) and Asn40 (salmon stick). Backbone atoms of Pro24, Ile42, Thr43, Asn44, and Ala45 forming yellow hydrogen bonds with Asn40 are shown in cyan, as are the side chains of residues closely interacting with Trp26 (Pro23, Gln47, and Pro51)

In contrast to Trp26, Trp48 is relatively solvent-exposed and displays more mobility than Trp26, as indicated by all relaxation and J coupling measurements (Table 4.1), though both have a dominant gauche⁺ rotamer. The side chains of the aromatic residues, Tyr37, Tyr38, and Phe39, appear to be rigid by both J couplings and relaxation, and their aromatic rings have significant packing interactions with polar side chains as well as remote regions of the backbone: the aromatic ring of Tyr37 contacts Arg29, Ser31, and Ser46; Tyr38 contacts Lys28 and Arg50; and Phe39 contacts Arg29, His41, Asn44, and Ser46 (Figure 4.5). The hydrophilic side chains are mobile, as assessed by ¹H relaxation and/or J couplings, except for Arg29. Besides contacting two aromatic residues, Arg29 also forms a salt bridge with Glu27. Arg29 also has a unique H β 2 chemical shift at 0.09 ppm due to its close approach to Tyr37.

It is noteworthy that the side chain of Asp33 is rigid, even though this residue is not part of the native Pin1 WW domain sequence. Asp33 is part of a deletion-substitution mutation found

to stabilize the WW domain¹⁵. The side chain of Asp33 is in the sterically disfavored gauche⁻ conformation (Figure 4.S6), which allows its carboxyl group to hydrogen bond with the side chain of Ser31, with both side chains forming electrostatic interactions with the ring protons of Trp48.

Figure 4.7a shows strip plots from 3D ¹H-TOCSY-¹³C-HSQC, HN(CO)Hβ, and HNHβ for Tyr37. The residue predominantly adopts the gauche⁻ χ₁-rotamer with downfield resonance assigned to Hβ₃. For comparison, strip plots for a side chain with rotameric averaging, Lys28, are shown in Figure 4.7b.

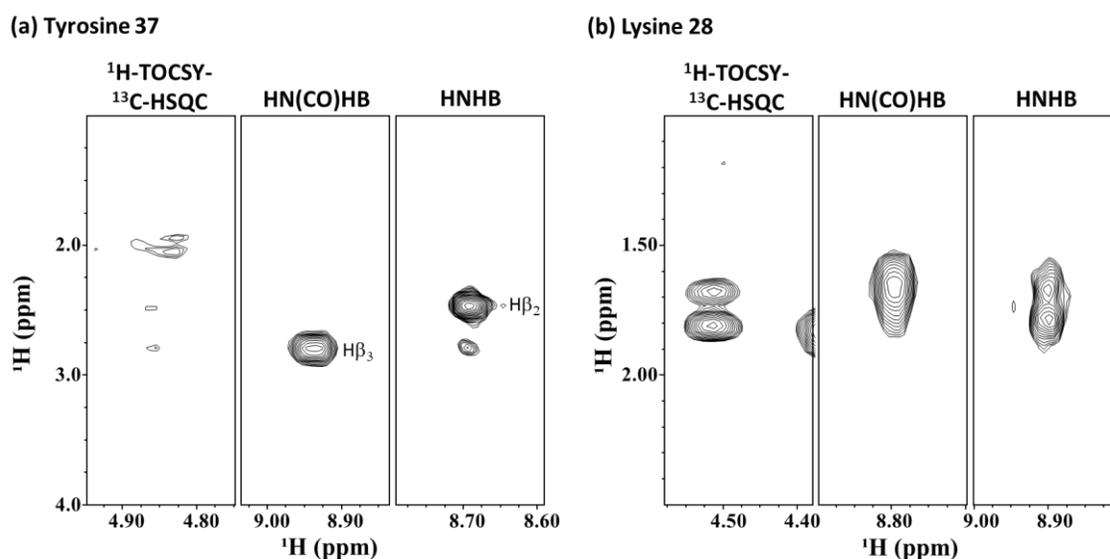


Figure 4.7 NMR strip plots from 3D ¹H-TOCSY-¹³C-HSQC, HN(CO)Hβ, and HNHβ for a rigid Pin1 residue, Tyr37, adopting the gauche⁻ χ₁ rotamer (a), and averaging Lys28 sampling two χ₁ rotamers (b). ¹H-TOCSY-¹³C-HSQC and HNHβ strips are taken at the Hα and NH frequencies of residue-*i*, respectively, while HN(CO)Hβ strips are taken at the NH frequency of residue *i*+1. The intensity of each correlation relates to the size of each ³J coupling.

Pin1 residues averaging between two χ₁ dihedral angle rotamers.

Based on J couplings, we could identify just two residues with rotameric mobility about two of the three major χ₁-dihedral angle positions: Lys28 and Asn44. Without stereospecific assignments for Hβ_{2/3}, both residues appear to have a slight preference for the sterically unfavorable gauche⁻ rotamer (see Figure 4.7 for Lys28). However, with stereospecific assignments, it becomes clear that the residues are instead averaging between the trans and gauche⁺-rotamers (Figure 4.8). We assigned stereospecific Hβ resonances using an isotope

labelling scheme we developed previously, which selectively protonates the H β_2 of Asp, Asn, Lys, and Met amino acid residues with deuteration at H β_3 using fumarate as a carbon source for *E. coli* in D₂O¹⁰. In the X-ray crystal structure, Lys28 is partially solvent-exposed on the first β -strand of Pin1 and interacts with Val36 and Tyr38 (Figure 4.8). All 3 rotameric positions are accessible to Lys28, but the trans conformation allows it to interact more closely with Val36, while the gauche⁺ conformation brings it into closer contact with Tyr38. The relaxation rates of Lys28 clearly indicate that its side chain is mobile, whereas the J couplings are intermediate between those of rigid residues and those that can freely access all three major rotamers.

Asn44 is found in the loop connecting the second and the third β -strands of Pin1 (Figure 4.2). The χ_1 dihedral angle for Asn44 is gauche⁺ from the crystal structure. Like Lys28, all 3 χ_1 dihedral angle rotamers are accessible to Asn44, but the gauche⁺ and trans rotamers allow its sidechain to form close contacts with the aromatic ring of Phe39 (Figure 4.8). As with Lys28, relaxation data indicate that the side chain of Asn44 is mobile, but the maximum observed J couplings are intermediate between those of rigid and mobile residues. It is noteworthy that for both Lys28 and Asn44, the two preferred rotamers allow the hydrophilic side chains to make contacts with other side chains, whereas the disfavoured gauche⁻ rotamer is more solvent-exposed.

The examples of Lys28 and Asn44 demonstrate that for mobile residues, one should not attempt to define the χ_1 -dihedral angle without independently obtained stereospecific assignments for H β residues. It is impossible to distinguish between there being one preferred rotamer with substantial access to the other two major rotamers, versus the other two rotamers being preferred with the stereospecific H β assignments reversed. In such cases, it is possible to obtain the wrong stereospecific assignments with the wrong rotamer determination.

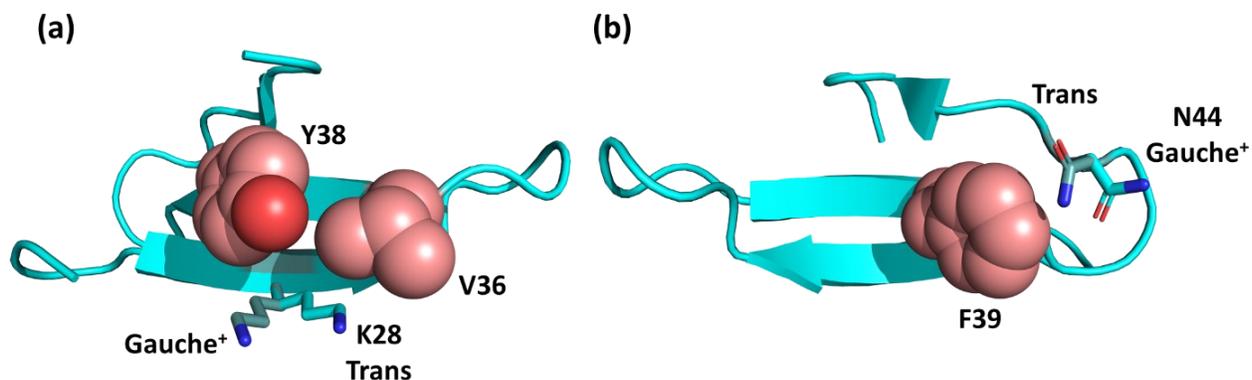


Figure 4.8 The alternative rotamers of Lys28 (a) and Asn44 (b). The trans/gauche⁺ rotamers are shown as sticks in cyan (crystal structure conformation) or light teal (sampled rotamer). The residues contacting Lys28 (Val36 and Tyr38) and Asn44 (Phe39) are shown as salmon spheres

Methyl-containing residues

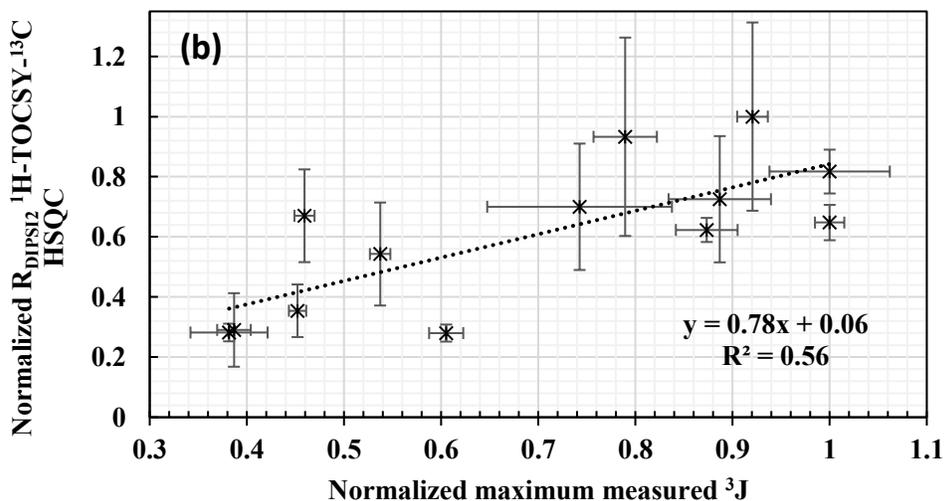
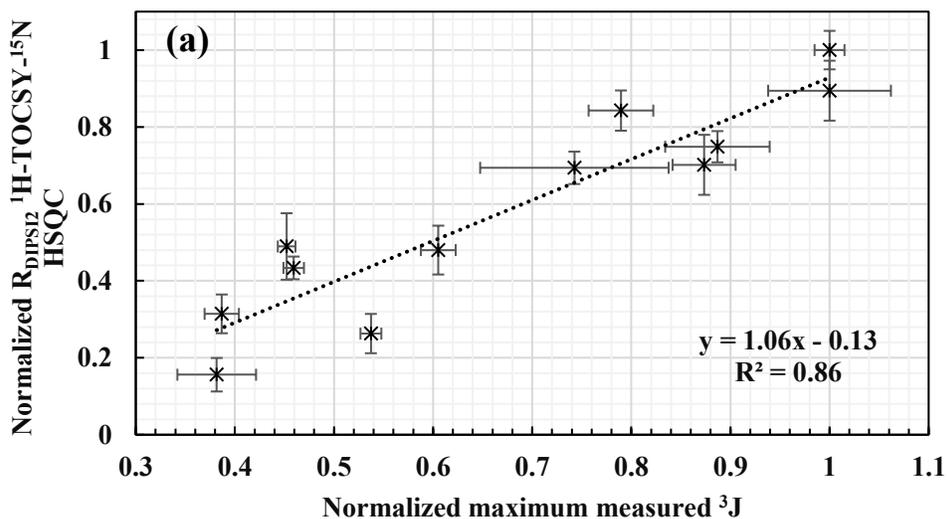
The alanine β -methyl group is rigidly attached to the backbone and should provide a gauge of backbone dynamics comparable to ^{15}N . There are three alanine residues in our mutant Pin1 WW domain, Ala17, Ala32, and Ala45. Of these, Ala32 and Ala45 are part of the folded WW domain and have normalized relaxation rates of 0.33 and 0.26, respectively (Table 4.1). These values are perhaps higher than what might be expected for a rapidly spinning methyl group, given that other residues with larger side chains have comparable or even slower rates (for example, D18, K28, M30). This may be because the initial relaxation rate of an alanine methyl group is dominated by its faster-relaxing components due to constructive interference of the ^1H magnetic fields, and this phenomenon is not present in the β -methylene side chains. Similar relaxation rates are observed for the γ -methyl groups of Val36, Ile42, and Thr43, suggesting that the side chains of these residues may be immobile. In theory, it should have been possible to determine the dominant rotamer for these residues, but this was hampered by signal overlap. It is important to note that we are able to obtain meaningful relaxation data for all 20 amino acid types via ^1H relaxation: β -methylene positions for most amino acid residues, α -methylene positions for Gly, β -methyl positions for Ala, and γ -methyl positions for Val, Ile, and Thr.

Correlations between maximum observed ^3J couplings and protein side-chain relaxation rates

We have compared the measured ^1H relaxation rates at $\text{C}\beta$ derived from 3D ^{15}N -TOCSY-HSQC with the maximum measured ^3J -couplings in Figure 4.9a. The figure shows a strong correlation, with a correlation coefficient of 0.86. Both independent mobility measures agree that the most rigid side chains in the entire Pin1 WW domain belong to the Trp26-Asn40 dyad at the folded center of the domain. The reason for the strong correlation is that rotation about the χ_1 -dihedral angle attenuates both maximum observed J couplings as well as the transverse magnetization ^1H relaxation rates. In theory, J coupling measurements are sensitive to a much larger range of motion timescales, everything from sub-nanosecond timescale motions up to tens of milliseconds (exchange broadening would become a confounding factor towards the slower end of this range). Thus, the fact that we observe a strong correlation between J couplings and ^1H relaxation indicates that almost all the side chain rotameric averaging in Pin1 WW domain is happening on a very fast time scale. Further studies in other systems would be needed to demonstrate if this is generally applicable to proteins or not. We note that Arseniev and coworkers¹⁴ also observed this correlation for a small 61-residue water-soluble protein neurotoxin II.

Figures 4.9b, 4.9c, and 4.9d show positive correlations between the other relaxation experiments and maximum observed ^3J values. The correlations for relaxation data derived from the ^1H -TOCSY- ^{13}C -HSQC (Figure 4.9b) and the CBCACONH-based cross-correlated relaxation (Figure 4.9d) experiments are not as strong as we observed for the ^1H -TOCSY- ^{15}N -HSQC and CBCACONH (Figure 4.9c) ^1H -DIPSI-2 relaxation experiments. As we noted earlier, we suspect that a major problem with the data obtained from the ^1H -TOCSY- ^{13}C -HSQC was spectral distortion arising from the water signal. Another issue with the ^{13}C -HSQC is that the water signal is not restored to +z-axis at the end of the pulse sequence (as for the 2D ^{15}N -HSQC), being instead dephased for solvent suppression. Thus, ^1H - ^1H NOE-aided T_1 spin-lattice relaxation would not be able to uniformly return all the protein ^1H spins close to their equilibrium positions prior to the pulse sequence, leading to inconsistent evolution and relaxation properties during the DIPSI-2 element. Whatever the reason, relaxation data from the 3D ^1H -TOCSY- ^{13}C -HSQC (Figure 4.9b) and the 2D ^{13}C -HSQC (not shown) was less reliable and yielded poorer correlations than data obtained from spectra utilizing the 2D ^{15}N -HSQC (Figures 4.9a and 4.9c).

This was very disappointing to us, as the ^{13}C -HSQC-based spectra possessed far superior signal-to-noise. It may be worthwhile to re-write the pulse sequences for experiments build on the ^{13}C -HSQC to utilize a water-selective flipback pulse for solvent suppression, even though this would mean loss of signal from αH protons. Finally, for the cross-correlated relaxation experiment, another issue could be that constructive and destructive interference of the magnetic fields from the two ^{13}C - ^1H dipoles results in a net magnetic field that is less sensitive to rotations about the χ_1 -dihedral angle and more sensitive to backbone dynamics.



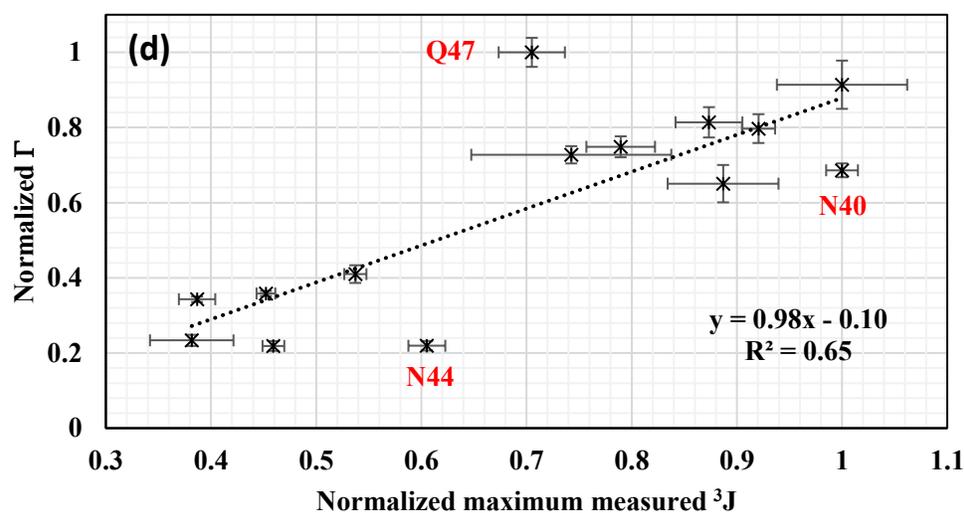
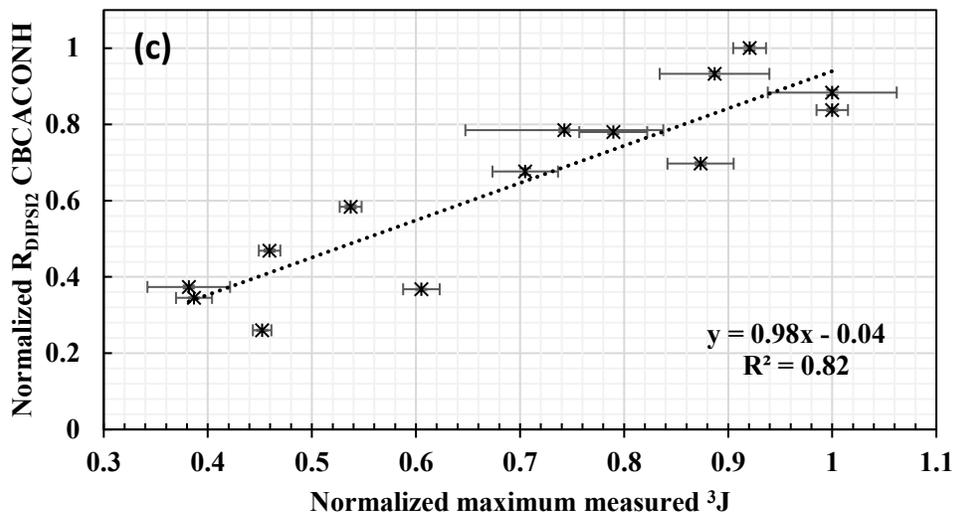


Figure 4.9 Correlation plot of the normalized maximum observed 3J coupling and normalized R_{DIPS12} derived from 1H -TOCSY- ^{15}N -HSQC (a), 1H -TOCSY- ^{13}C -HSQC (b), CBCACONH (c), and cross-correlated relaxation rates (d). Errors bars are indicated on the vertical and horizontal axis

Analyzing 3J -coupling measurements to determine χ_1 dihedral angle conformation has a rich history in NMR^{11,12,16,18}. However, most analyses have relied on the simplifying assumption of a dominant trans, gauche-, or gauche+ conformation, even though most protein side chains are mobile, averaging between two or three χ_1 dihedral angle rotamers^{20,21}. Few studies have used J

couplings to measure side chain dynamics, and even fewer have compared them to relaxation-based estimates of side chain dynamics^{22–24}. J coupling experiments suffer from reduced sensitivity due to their reliance on small ³J couplings, as well as an inability to obtain information on protein residues in which methylene protons have identical or very similar chemical shifts. ¹H relaxation measurements thus provide high quality data on protein residues that would otherwise be inaccessible to J coupling-based analysis.

Precise characterization of side-chain dynamics has proven to be a challenging process. We show that protein side-chain motions can be characterized using a simple method of measuring ¹H relaxation during a ¹H TOCSY element, complementing the dynamic information derived independently from ³J-coupling measurements. Historically, neither ¹H relaxation nor ³J couplings have been extensively used to probe side chain dynamics. We hope that with the implementation of ¹H relaxation, studies of side chain dynamics can become routine by solution NMR.

Acknowledgements

This work was supported by a Discovery Grant, RGPIN-2015- 06664, from NSERC (Natural Sciences and Engineering Research Council of Canada), startup funds from the Department of Medicine and Faculty of Medicine & Dentistry at the University of Alberta, and private funding from the Hwang Professional Corporation.

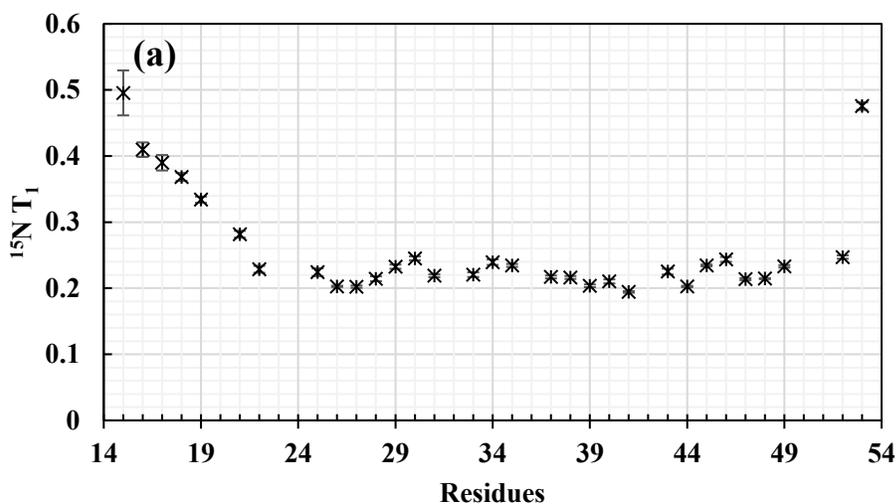
References

- (1) Mittermaier, A.; Kay, L. E. X1 Torsion Angle Dynamics in Proteins, From Dipolar Couplings. *J. Am. Chem. Soc.* **2001**, *123* (28), 6892–6903. <https://doi.org/10.1021/ja010595d>.
- (2) Li, F.; Grishaev, A.; Ying, J.; Bax, A. Side Chain Conformational Distributions of a Small Protein Derived from Model-Free Analysis of a Large Set of Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, *137* (46), 14798–14811. <https://doi.org/10.1021/jacs.5b10072>.
- (3) Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in ¹³C-Labeled and Fractionally ²H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **2002**, *124* (22), 6439–6448. <https://doi.org/10.1021/ja012497y>.
- (4) Skrynnikov, N. R.; Millet, O.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* **2002**, *124* (22), 6449–6460. <https://doi.org/10.1021/ja012498q>.
- (5) Tugarinov, V.; Kay, L. E. Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins. *ChemBioChem* **2005**, *6* (9), 1567–1577.

- <https://doi.org/10.1002/cbic.200500110>.
- (6) Yang, D.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. A Study of Protein Side-Chain Dynamics from New 2H Auto-Correlation and 13C Cross-Correlation NMR Experiments: Application to the N-Terminal SH3 Domain from Drk. *J. Mol. Biol.* **1998**, *276* (5), 939–954. <https://doi.org/10.1006/jmbi.1997.1588>.
 - (7) Zheng, Y.; Yang, D. Measurement of Dipolar Cross-Correlation in Methylene Groups in Uniformly 13C-, 15N-Labeled Proteins. *J. Biomol. NMR* **2004**, *28* (2), 103–116. <https://doi.org/10.1023/B:JNMR.0000013826.82936.0e>.
 - (8) Bothner-By, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **2002**, *106* (3), 811–813. <https://doi.org/10.1021/JA00315A069>.
 - (9) Bax, A. Homonuclear Hartmann-Hahn Experiments. *Methods Enzymol.* **1989**, *176* (C), 151–168. [https://doi.org/10.1016/0076-6879\(89\)76010-9](https://doi.org/10.1016/0076-6879(89)76010-9).
 - (10) Danmaliki, G. I.; Liu, P. B.; Hwang, P. M. Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for Escherichia Coli in D2O. *Biochemistry* **2017**, *56* (45). <https://doi.org/10.1021/acs.biochem.7b00991>.
 - (11) Archer, S. J.; Ikura, M.; Torchia, D. A.; Bax, A. An Alternative 3D NMR Technique for Correlating Backbone 15N with Side Chain H β Resonances in Larger Proteins. *J. Magn. Reson.* **1991**, *95* (3), 636–641. [https://doi.org/10.1016/0022-2364\(91\)90182-S](https://doi.org/10.1016/0022-2364(91)90182-S).
 - (12) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. Measurement of Homo- and Heteronuclear J Couplings from Quantitative J Correlation. *Methods Enzymol.* **1994**, *239* (C), 79–105. [https://doi.org/10.1016/S0076-6879\(94\)39004-5](https://doi.org/10.1016/S0076-6879(94)39004-5).
 - (13) Yang, D. Probing Protein Side Chain Dynamics Via 13C NMR Relaxation. *Protein Pept. Lett.* **2011**, *18* (4), 380–395. <https://doi.org/10.2174/092986611794653932>.
 - (14) Lesovoy, D. M.; Dubinnyi, M. A.; Nolde, S. B.; Bocharov, E. V.; Arseniev, A. S. Accurate Measurement of Dipole/Dipole Transverse Cross-Correlated Relaxation Γ_2 in Methylenes and Primary Amines of Uniformly 13C / 15N -Labeled Proteins. *J. Biomol. NMR* **2019**, *73* (5), 245–260. <https://doi.org/10.1007/s10858-019-00252-6>.
 - (15) Jäger, 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. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (28), 10648–10653. <https://doi.org/10.1073/pnas.0600511103>.
 - (16) Grzesiek, S.; Ikura, M.; Marius Clore, G.; Gronenborn, A. M.; Bax, A. A 3D Triple-Resonance NMR Technique for Qualitative Measurement of Carbonyl-H β J Couplings in Isotopically Enriched Proteins. *J. Magn. Reson.* **1992**, *96* (1), 215–221. [https://doi.org/10.1016/0022-2364\(92\)90307-S](https://doi.org/10.1016/0022-2364(92)90307-S).
 - (17) Cavanagh, J.; Rance, M. Suppression of Cross-Relaxation Effects in TOCSY Spectra via a Modified DIPSI-2 Mixing Sequence. *J. Magn. Reson.* **1992**, *96* (3), 670–678. [https://doi.org/10.1016/0022-2364\(92\)90357-D](https://doi.org/10.1016/0022-2364(92)90357-D).
 - (18) Clore, G. M.; Bax, A.; Gronenborn, A. M. Stereospecific Assignment of B-Methylene Protons in Larger Proteins Using 3D 15N-Separated Hartmann-Hahn and 13C-Separated Rotating Frame Overhauser Spectroscopy. *J. Biomol. NMR* **1991**, *1*, 13–22. <https://doi.org/10.1007/BF01874566>.
 - (19) Zhang, R. B.; Somers, K. R. F.; Kryachko, E. S.; Nguyen, M. T.; Zeegers-Huyskens, T.;

- Ceulemans, A. Hydrogen Bonding to π -Systems of Indole and 1-Methylindole: Is There Any OH \cdots Phenyl Bond? *J. Phys. Chem. A* **2005**, *109* (35), 8028–8034. <https://doi.org/10.1021/jp0525437>.
- (20) Tuttle, L. M.; Dyson, H. J.; Wright, P. E. Side-Chain Conformational Heterogeneity of Intermediates in the Escherichia Coli Dihydrofolate Reductase Catalytic Cycle. *Biochemistry* **2013**, *52* (20), 3464–3477. <https://doi.org/10.1021/bi400322e>.
- (21) Hu, J. S.; Bax, A. Determination of ϕ and χ_1 Angles in Proteins from ^{13}C - ^{13}C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc.* **1997**, *119* (27), 6360–6368. <https://doi.org/10.1021/ja970067v>.
- (22) Hu, H.; Hermans, J.; Lee, A. L. Relating Side-Chain Mobility in Proteins to Rotameric Transitions: Insights from Molecular Dynamics Simulations and NMR. *J. Biomol. NMR* **2005**, *32* (2), 151–162. <https://doi.org/10.1007/s10858-005-5366-0>.
- (23) Chou, J. J.; Case, D. A.; Bax, A. Insights into the Mobility of Methyl-Bearing Side Chains in Proteins from 3JCC and 3JCN Couplings. *J. Am. Chem. Soc.* **2003**, *125* (29), 8959–8966. <https://doi.org/10.1021/ja029972s>.
- (24) Smith, L. J.; van Gunsteren, W. F.; Hansen, N. On the Use of Side-Chain NMR Relaxation Data to Derive Structural and Dynamical Information on Proteins: A Case Study Using Hen Lysozyme. *ChemBioChem* **2021**, *22* (6), 1049–1064. <https://doi.org/10.1002/cbic.202000674>.

Supplementary Information



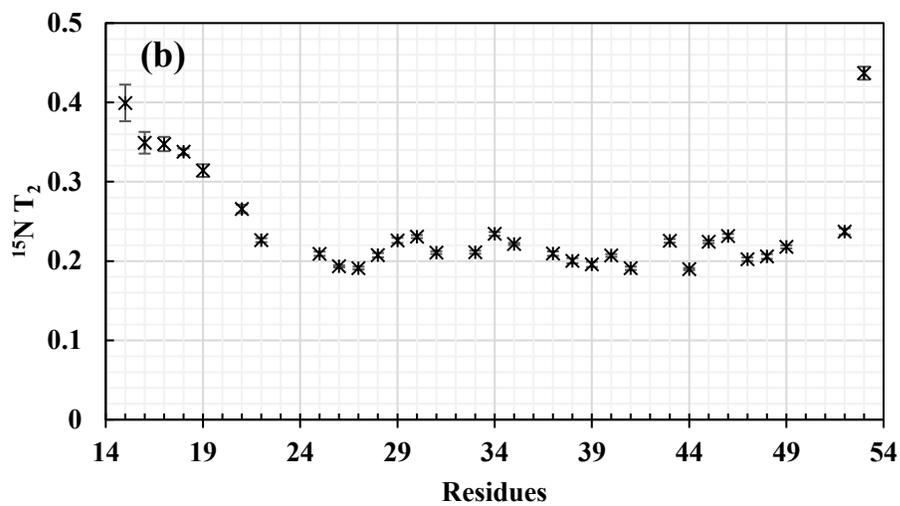


Figure 4.S1 Backbone dynamics of Pin1 measured by $^{15}\text{N } T_1$ (a) and $^{15}\text{N } T_2$ (b) relaxation. Errors are indicated as vertical bars. T_2 relaxation is calculated from $T_{1\rho}$

Table 4.S1. Chemical shifts and normalized values for maximum observed 3J -couplings and $^1H R_{DIPSI2}$ relaxation rates for the mutant human Pin1 WW domain. $C\beta$ cross-correlated relaxation rates have not been normalized. In the last column, the predominant rotamer is determined by comparing $H\beta_2$ with $H\beta_3$ intensities in the NMR experiments used to derive J couplings, but this is not always possible if the NMR signals are overlapped or degenerate. In such cases, we label the residue “mobile” or “restricted” based on the available relaxation data. Generally speaking, “mobile” residues have normalized values <0.55 , while “restricted” residues have normalized values >0.55 . The value is left blank if the residues has no β CH_2 group (A,G,I,V,T) or if there was insufficient data.

Residue	Atom	Chemical shifts (ppm)	$^3J_{H\alpha H\beta}$	$^3J_{HNH\beta}$	$^3J_{HN(CO)H\beta}$	$^1H R_{DIPSI2}$ from ^{15}N TOCSY	$^1H R_{DIPSI2}$ from ^{13}C TOCSY	$^1H R_{DIPSI2}$ from CBCACONH	$\Gamma_{C\beta C\alpha(CO)NH}$ (s^{-1})	Predominant Rotamer
M16	$H\alpha$	4.51				0.11	0.21*	0.27		
	$H\beta_2$	2.05	-	-	-	0.39	0.30	0.28	-	Mobile
	$H\beta_3$	2.05					0.30			
	$H\gamma_2$	2.48					0.13*			
	$H\gamma_3$	2.58								
	$H\epsilon_1$	2.10					-0.05*			
A17	$H\alpha$	4.35				0.02	0.08	0.22		
	$H\beta$	1.42	^a	-	-	0.28	0.16	0.18	-	
D18	$H\alpha$	4.6				0.10	0.16	0.29		
	$H\beta_2$	2.64	0.42	0.35	0.34	0.31	0.29	0.34	-1.46	Mobile
	$H\beta_3$	2.74								
E19	$H\alpha$	4.28				0.02	0.11	0.34		
	$H\beta_2$	2.10	-	0.4	-	0.31	0.30	0.35	-2.17	Mobile
	$H\beta_3$	1.99								
	$H\gamma_2$	2.49					0.15			
	$H\gamma_3$	2.49					0.15			

E20	H α	4.29				0.13	0.15*	-		
	H β_2	2.03	-	-	-	0.42	-	0.31	-1.13	Mobile
	H β_3	1.83								
	H γ_2	2.31					0.23*			
	H γ_3	2.31								
K21	H α	4.33				0.15	0.19	-		
	H β_2	1.87	-	0.33	-	0.31	0.13	0.42	-1.66	Mobile
	H β_3	1.79								
	H γ_2	1.48					0.29			
	H γ_3	1.48								
	H δ_2	1.73					0.15			
	H δ_3	1.73								
	H ϵ_2	3.04					-0.01			
	H ϵ_3	3.00								
L22	H α	4.33				0.23	0.32	-		
	H β_2	1.45	a	b	-	0.74	0.71	-	-	Restricted
	H β_3	1.85								
	H γ	1.87					0.42			
	H δ_{11}	1.11					0.35			
	H δ_{21}	0.83					0.37			
P23	H α	4.87					0.47	-		

	H β_2	2.63	0.45	-	-	-	0.79	-	-	Restricted
	H β_3	2.02								
	H γ_2	1.81					0.51			
	H γ_3	1.64								
	H δ_2	3.01					0.69			
	H δ_3	3.71								
P24	H α	4.40					0.12	0.38		
	H β_2	1.91	0.39	-	-	-	0.60	0.53	-4.42	
	H β_3	2.36								
	H γ_2	2.18					0.43			
	H γ_3	2.07								
	H δ_2	3.94					0.64			
	H δ_3	3.67								
G25	H α_2	4.05	-	-	-	0.62	0.53	0.68	-7.20	
	H α_3	3.33								
W26	H α	5.29				0.46	0.30	0.58		
	H β_2	3.27	1.00	1.00	-	0.89	0.82	0.88	-10.64	Gauche+
	H β_3	2.99								
E27	H α	4.88				0.46	0.36*	0.48		
	H β_2	2.22	-	0.40	0.47	0.45	0.30	0.46	4.06	Mobile
	H β_3	2.32								
	H γ_2	2.57					0.26*			
	H γ_3	2.24								

K28	H α	4.48				0.22	0.30	0.50			
	H β_2	1.78	0.50	0.54	0.57	0.26	0.54	0.58	-2.53	Averaged between Trans and Gauche+	
	H β_3	1.66									
	H γ_2	1.13					0.49				
	H γ_3	1.13									
	H δ_2	1.73					0.25				
	H δ_3	1.73									
	H ϵ_2	2.97					0.17				
	H ϵ_3	2.97									
R29	H α	4.45				0.53	0.31*	0.48			
	H β_2	0.09	0.70	0.79	0.27	0.69	0.70	0.78	-7.64	Gauche+	
	H β_3	1.32									
	H γ_2	1.25									
	H γ_3	1.41					0.60				
	H δ_2	2.89					0.33				
	H δ_3	2.61									
M30	H α	5.31				0.39	0.28*	0.41			
	H β_2	1.98	0.38	0.41	0.50	0.16	0.28	0.37	0.30	Mobile	
	H β_3	1.89									
	H γ_2	2.61					0.16				
	H γ_3	2.59									
	H ϵ_1	2.02					-0.02*				
S31	H α	4.75					0.21*				
	H β_2	4.53	-	-	-	-	0.22	-	-	Mobile	

	H β_3	4.23									
A32	H α	4.22					0.26	0.26			
	H β_1	1.53	-	-	-	-	0.27	0.31	-		
D33	H α	4.69				0.14	0.34*	0.46			
	H β_2	2.98	-	0.43	0.87	-	0.54	0.53	-5.85		Gauche-
	H β_3	2.67									
G34	H α_2	4.23	-	-	-	0.67	0.52	0.59	-5.34		
	H α_3	3.79									
R35	H α	4.34									
	H β_2	1.96	-	-	-	0.32	-	0.63	-		Mobile
	H β_3	2.05									
V36	H α	4.71					0.31	0.47			
	H β	2.01	-	b	-	0.23	0.24	0.38	-		
	H γ_{11}	0.81					0.22				
	H γ_{21}	1.07					0.23				
Y37	H α	4.84				0.52	-	0.52			
	H β_2	2.46	0.39	0.89	0.95	-	1.00	1.00	-8.76		Gauche-
	H β_3	2.79									
Y38	H α	5.27				0.37	0.53	0.54			
	H β_2	2.69	0.88	0.90	-	0.75	0.73	0.93	-6.40		Gauche+
	H β_3	2.94									

F39	H α	5.62				0.53	0.52	0.57		
	H β_2	2.89	0.60	0.31	0.94	0.84	0.93	0.78	-7.98	Trans
	H β_3	2.61								
N40	H α	4.44				0.44	0.43	0.59		
	H β_2	-0.60	1	0.31	1	1.00	0.65	0.84	-6.97	Trans
	H β_3	2.00								
H41	H α	4.11				-	-	-		
	H β_2	3.30	-	0.81	c	-	-	0.62	-2.15	Gauche-
	H β_3	3.08								
I42	H α	3.87				-	0.32	0.39		
	H β	2.02	a	b	-	-	0.46	0.42	-	
	Hg ₁₂	1.32								
	Hg ₁₃	1.02					0.51			
	H γ_{21}	0.79					0.30			
	H δ_{11}	0.76					0.15			
T43	H α	4.12				0.23	0.30*	0.35		
	H β	4.25	a	0.84	b	-	0.40	0.28	-	
	H γ_{21}	0.96					0.26			
N44	H α	4.12				0.15	0.22	0.35		
	H β_2	3.14	0.51	0.69	0.44	0.48	0.28	0.37	0.54	Average between Trans and Gauche+
	H β_3	2.92								
A45	H α	4.48				0.10	0.17	0.35		

	H β_1	1.26	^a	-	-	0.22	0.26	0.26	-	
S46	H α	6.01				0.40	0.43	0.50		
	H β_2	3.84	-	0.46	0.40	0.43	0.67	0.47	0.54	Mobile
	H β_3	3.79								
Q47	H α	4.60				0.19	0.48	0.44		
	H β_2	2.24	-	0.62	0.79	-	-	0.68	-12.02	Gauche-
	H β_3	2.56								
	H γ_2	2.05					0.59			
	H γ_3	2.00								
W48	H α	4.96				0.14	0.44	0.47		
	H β_2	3.22	0.82	0.92	0.30	0.70	0.62	0.70	-9.03	Gauche+
	H β_3	3.67								
E49	H α	4.48				0.27	0.33	-		
	H β_2	1.91	-	0.68	-	0.49	0.36	-	-	Mobile
	H β_3	1.91								
	H γ_2	2.34					0.13			
	H γ_3	2.27								
R50	H α	4.69				0.21	-	-		
	H β_2	1.45	-	-	-	-	0.60	-	-	
	H β_3	1.45								
	H γ_2	1.03					0.55			
	H γ_3	1.24								
	H δ_2	3.04					0.37			

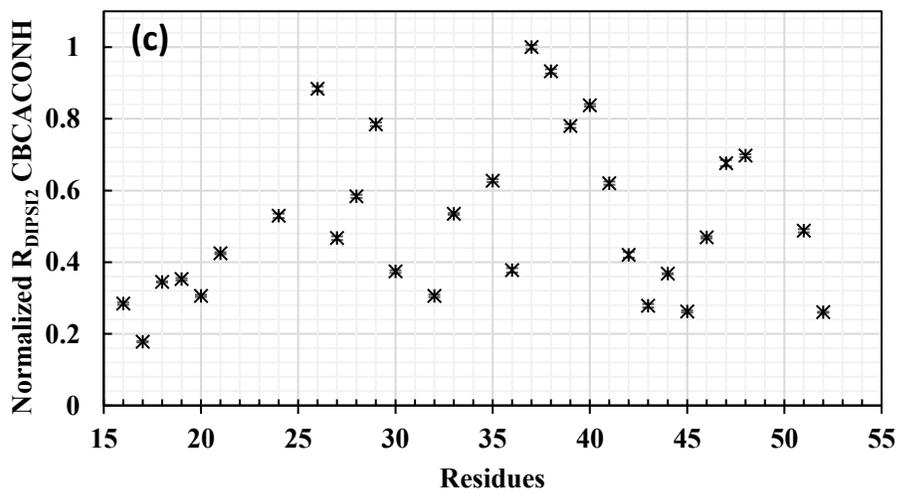
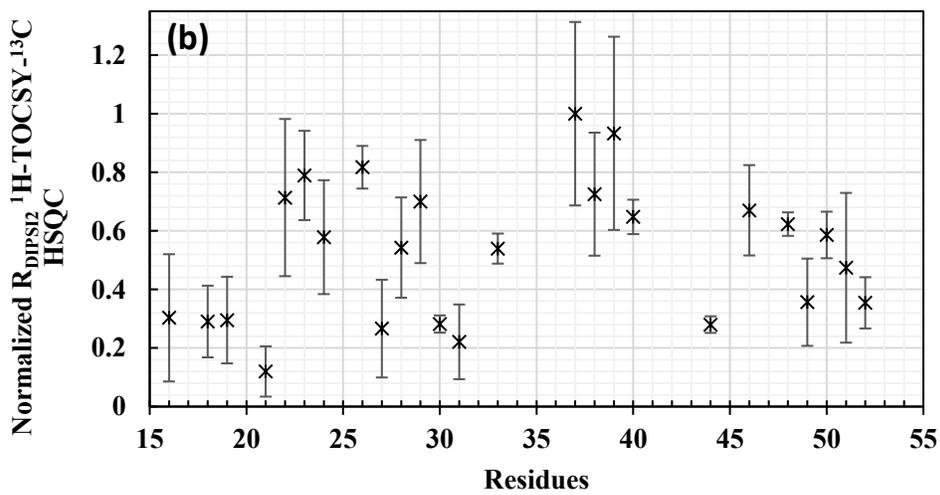
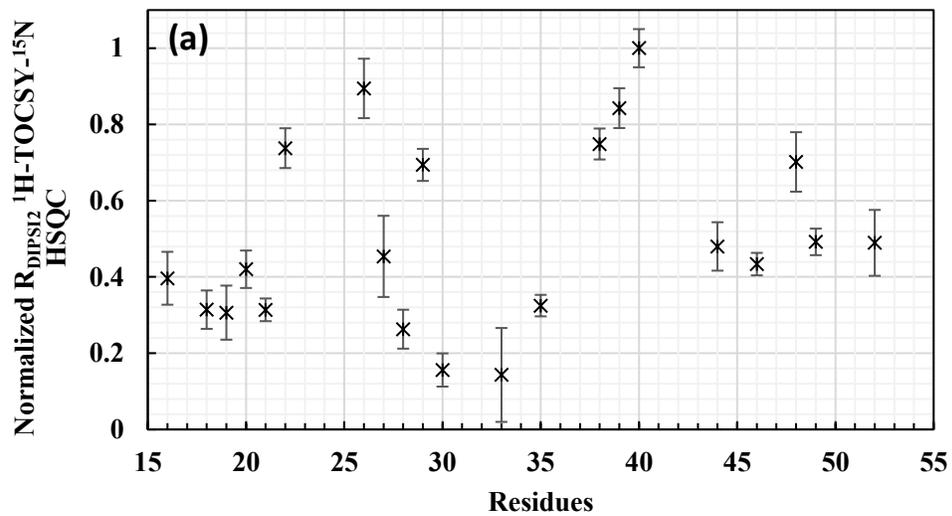
	H δ_3	3.04									
P51	H α	3.93					-0.02*	0.38			
	H β_2	0.84	0.61	-	-	-	0.47	0.49	-2.87		Mobile
	H β_3	0.63									
	H γ_2	-0.04					0.50				
	H γ_3	0.59									
	H δ_2	2.44					0.80				
	H δ_3	2.31									
S52	H α	4.33				0.25	0.19	0.29			
	H β_2	3.81	0.36	0.30	0.45	0.49	0.35	0.26	-1.70		Mobile
	H β_3	3.75									
G53	H α_2	3.78	-	-	-	0.11	0.15				
	H α_3	3.78									

^a Single H β peak, but no second H β peak for quantitative ³J coupling estimation.

^b Peak present in HNHB spectrum, but overlapped signal in 2D reference spectrum prevents quantitative ³J coupling measurement.

^c Peak present in HNCOHB spectrum, but overlapped signal in 2D reference spectrum prevents quantitative ³J coupling measurement.

* Diagonal peak used to calculate relaxation rate



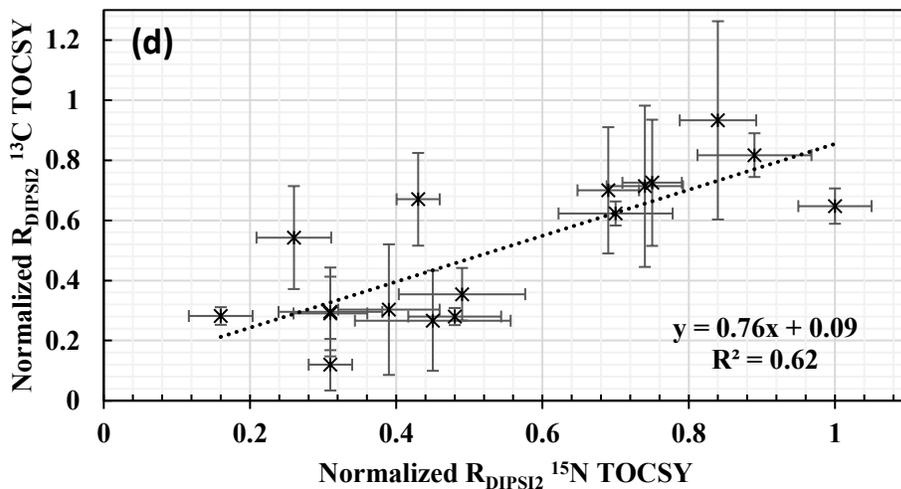


Figure 4.S2 Normalized β -methylene ^1H R_{DIPSI2} relaxation rates derived from 3D ^1H -TOCSY- ^{15}N -HSQC (a) 3D ^1H -TOCSY- ^{13}C -HSQC (b) CBCACONH (c) plotted against Pin1 sequence. In both figures, a value of one corresponds to a rigid side chain excluding methyl-containing residues. Error bars are displayed in the vertical axis. The plot of ^1H relaxation rates derived from 3D ^1H -TOCSY- ^{15}N -HSQC versus ^1H -TOCSY- ^{13}C -HSQC (d). Error bars are indicated in the vertical and horizontal axis

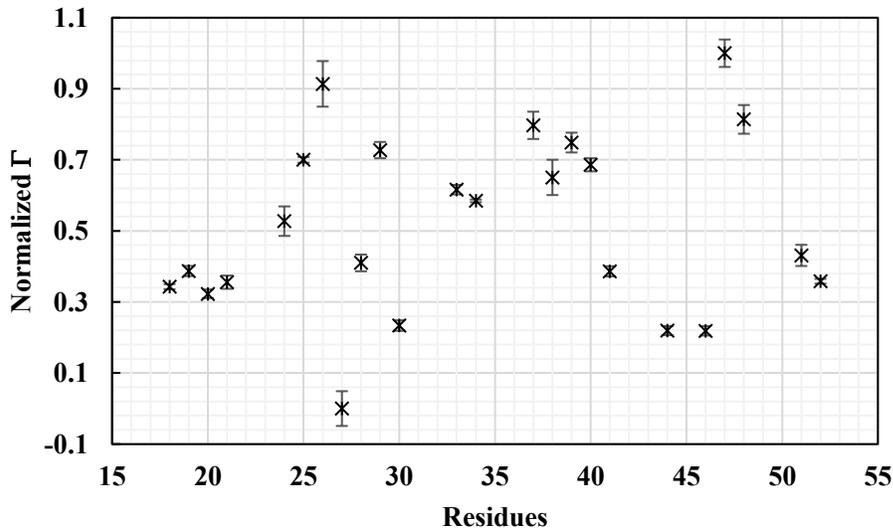
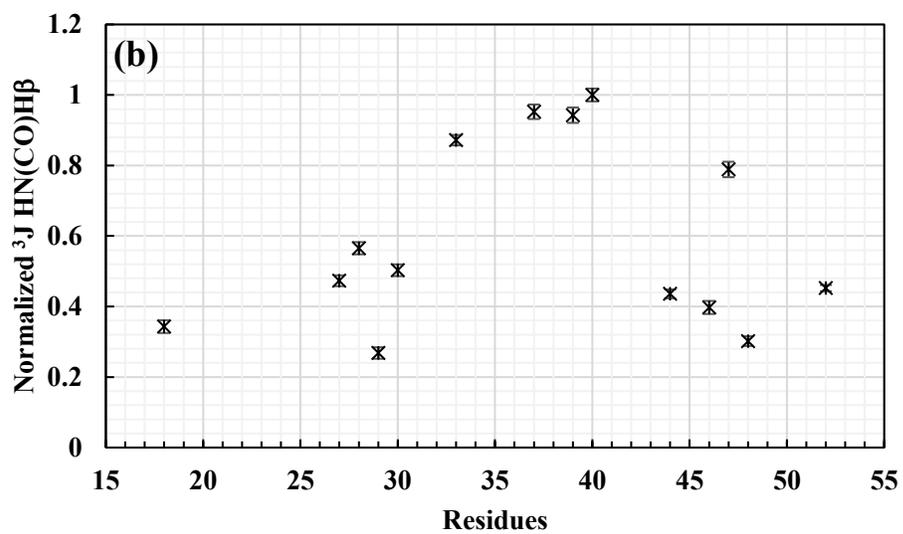
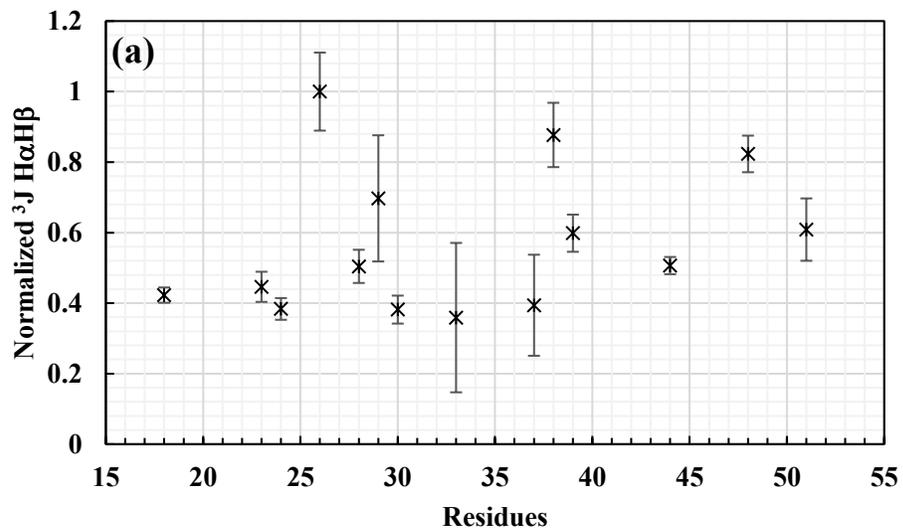


Figure 4.S3 Normalized cross-correlated relaxation rates plotted as a function of Pin1 residues. A value of zero indicates a less restricted motion of a flexible methylene group



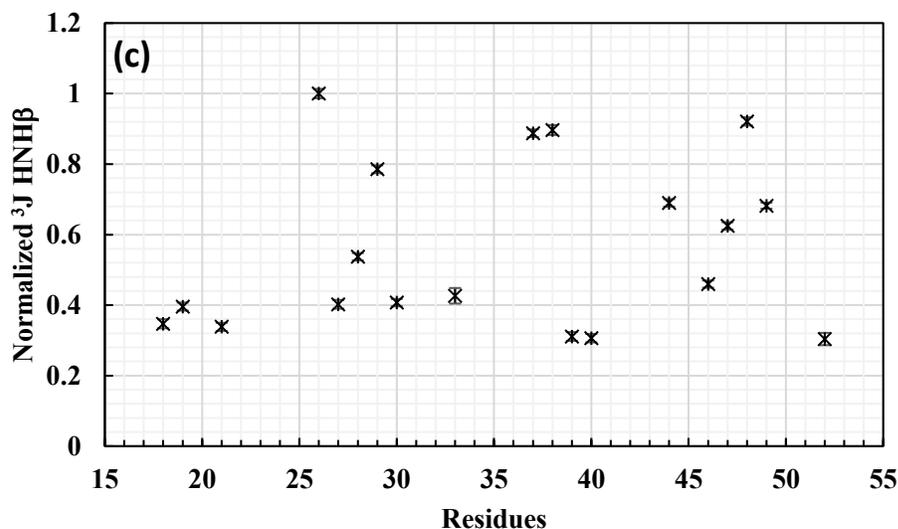


Figure 4.S4 Normalized 3J -coupling experiments: 3J H α H β (a), 3J HNCOH β (b), and 3J HNH β (c) plotted against the Pin1 sequence. Errors are indicated as vertical bars. These experiments provide a specific stereospecific assignment for the C β protons and define χ_1 dihedral angle for a given residue if the side chain is restricted

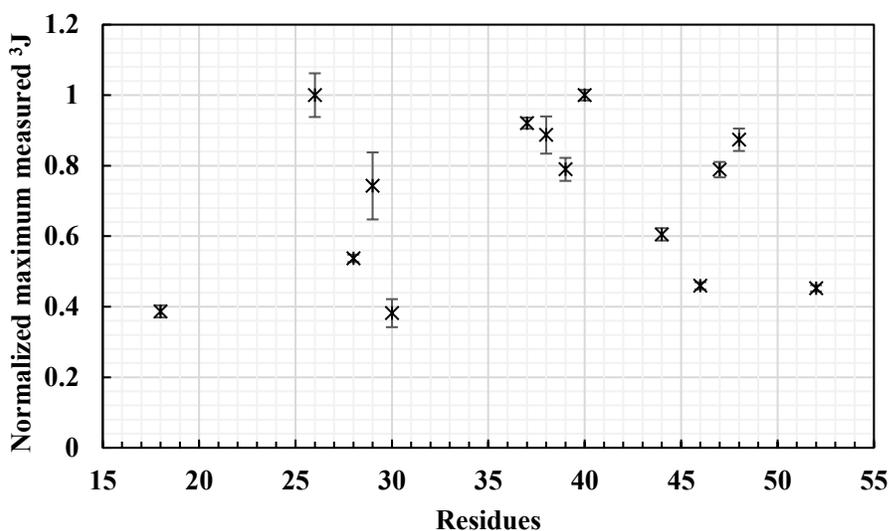


Figure 4.S5 The normalized maximum measured 3J -couplings (average of the two highest normalized values, which should in theory be the same) plotted against human Pin1 sequence. Errors are indicated with vertical bars—only data from amino acid residues with non-degenerate methylene H β chemical shifts were compared in this plot. A value of 1 corresponds to the most rigid residue in the protein

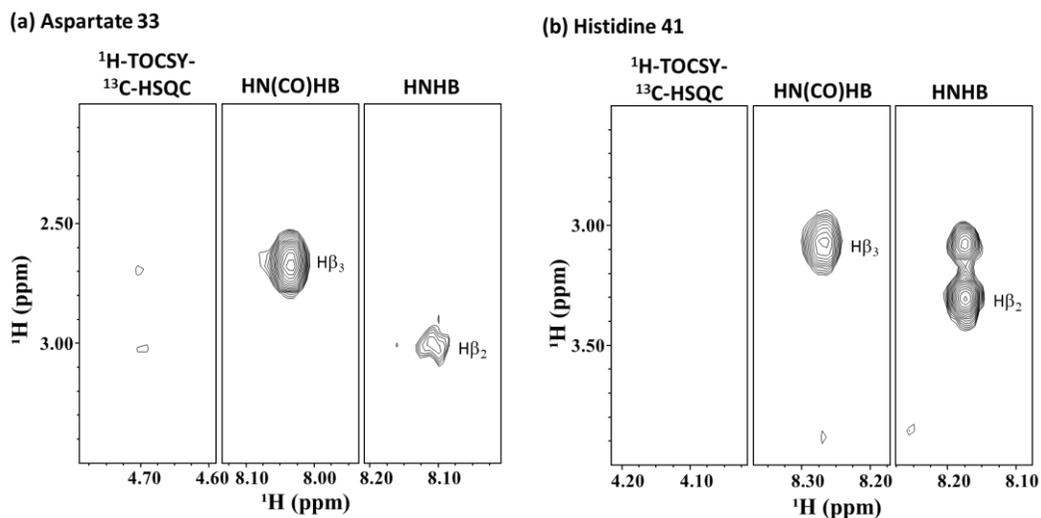


Figure 4.S6 NMR strip plots from 3D ¹H-TOCSY-¹³C-HSQC, HNCOHB, and HNHB experiments for Asp 33 (a) and His41 (b). Both residues adopt the gauche- χ_1 rotamer with the upfield resonances assigned to the H β_3 . The strips of ¹H-TOCSY-¹³C-HSQC and HNHB are taken at the Ha and NH frequency of residue-i, while HNCOHB strips are taken at the NH frequency of residue i+1

Chapter 5

Conclusion and future directions

Rapid signal decay in large proteins and inadequate side-chain information are the two significant limitations of solution NMR. We have developed biosynthetic methods to deuterate the side chains of amino acids in proteins stereospecifically. This approach facilitates the determination of structural restraints for all amino acids in high molecular weight systems.

In chapter two, we demonstrate a convenient and inexpensive technique for obtaining stereospecific β -methylene deuteration in Asp, Asn, Lys, and Met amino acid residues. We used rhamnose, pyruvate, and fumarate as carbon sources for *E.coli* in D_2O and added small molecule inhibitors: oxalate and malonate, to inhibit the scrambling of isotopes and the endogenous production of fumarate in the TCA cycle. This approach achieved stereoselective deuteration at the beta methylene proton (H_{β_3}) in the oxaloacetate family of amino acids. We also showed that the stereoselective 1H - ^{12}C labelling was helpful in NMR studies of large proteins and for determining side chain rotamers and dynamics of Asp, Asn, Lys, and Met amino acid residues in smaller proteins (developed in more detail in chapter 4).

Chapter three assessed Pin1 WW domain deuteration patterns in numerous growths using different carbon sources and metabolic inhibitors. We developed an isotopically enriched bacterial growth media for labelling proteins with isolated 1H - ^{12}C groups at non-methyl-containing amino acid residues (Asp, Asn, Met, Lys, Phe, Tyr, Trp, and His) in D_2O . We also introduced an efficient synthesis protocol for introducing isolated protons into the rings of aromatic amino acids using phenylpyruvate, hydroxyphenylpyruvate and anthranilate precursors. We showed that the labelling strategy is compatible with selective methyl group protonation. The isolated protons give rise to a long-lived 1H NMR signal, which is advantageous for obtaining solution NMR data on high molecular weight proteins.

We demonstrate the utility of our approach by labelling the outer membrane protein PagP in DPC detergent micelle (>50 kDa), specifically introducing targeted 1H - ^{12}C groups to compliment 1H - ^{13}C methyl and 1H - ^{15}N amide positions for chemical shift assignments and structure determination. We show that these groups can be connected via a multidimensional

NOESY experiment, yielding rich new structural restraints for solving a high-resolution membrane protein structure, which can be applicable to any protein of interest. We anticipate this will be the highest resolution structure of membrane protein obtained by solution NMR. The determination of the catalytic mechanism of this protein will allow it to be targeted by drugs in the future. Since PagP is a protein specifically used by disease-causing strains of Gram-negative bacteria, such therapies would target harmful bacteria while leaving the rest of the essential microbiome undisturbed.

While developing isotope labelling strategies to overcome the size limitation of solution NMR, we discovered a new NMR methodology for monitoring protein side-chain motion using ^1H relaxation rates, filling a significant gap in structural biology (chapter four). Solution NMR has been the most versatile technique used for monitoring protein dynamics. However, previous methods relied mainly on the measurement of backbone ^1H - ^{15}N amide groups because of their uniformity throughout the protein and the relative ease of magnetically isolating the spin pairs. Protein side chains are pretty flexible and involved in many interesting biological processes. However, the measurement of side-chain dynamics is complicated by ^{13}C - ^{13}C , ^1H - ^1H J couplings and relaxation of multiple ^{13}C - ^1H dipoles. Earlier attempts to overcome these limitations used $^{12}\text{C}/^{13}\text{C}$ and $^1\text{H}/^2\text{H}$ fractional labelling for protein systems <15 kDa or deuterium relaxation on methyl groups to probe side-chain motion.

Chapter four shows that protein side-chain dynamics can be measured using ^1H relaxation rates. We developed a new pulse sequence that uses a radiofrequency field as a TOCSY mixing element at the beginning of any multidimensional NMR experiment, compelling all proton atoms in a protein to behave similarly under the influence of strong coupling and maximum ROE contribution. So, differences in relaxation rates are due primarily to side-chain mobility. Our approach works well with uniformly labelled proteins, extending solution NMR studies of side-chain mobility to any protein system of interest, delineating how dynamics is critical to function and regulation.

The current artificial intelligence techniques, Rosetta and AlphaFold2, have limited accuracy in predicting side chain dynamics. While they can predict the overall protein structure based on high-resolution crystal structure templates, the flexibility and motions of individual side chains are more difficult to predict due to the complex interactions and movements involved.

Accurately predicting side chain dynamics typically requires additional experimental or computational methods.¹⁻⁶ Our approach for probing side-chain motion will permit the detailed characterization of structure and mobility. If our methods could be applied to many proteins in a database, this could greatly improve chemical shift-based protein structure and prediction methods. The problem with many studies correlating NMR chemical shifts with structure is that they do not adequately account for side chain motions.^{7,8} In the future, it will become possible for these methods to predict structure and dynamics solely based on NMR chemical shifts, incorporating this into a complete understanding of how protein dynamics is essential to function.

Limitations

A major limitation of our labelling technique is the replacement of the ^{13}C atom with ^{12}C . Most 3D NMR experiments are based on conventional uniform labelling patterns (^1H , ^{15}N , ^{13}C). Since ^{12}C has no magnetic moment, most multidimensional NMR will no longer be applicable, limiting the information that can be acquired. This limitation means that our experiments rely entirely on through-space NOEs to nearby ^1H - ^{15}N and ^1H - ^{13}C methyl groups for structure determination.

Our labelling protocol incorporates ^1H - ^{12}C groups into amino acid side chains in a largely deuterated background. Even though the sensitivity of ^1H - ^{12}C groups in a largely deuterated chemical environment is reasonably high, there is a potential problem of long T_1 relaxation time for that group. However, incorporating the methyl labelling protocol improves this limitation, serving as an efficient heat sink. The methyl groups relax efficiently and help nearby groups return quickly to equilibrium magnetization, thus minimizing the relaxation problems. Long T_1 relaxation times necessitate long recycle delay times, extending the time required for any experiment. Further experiments are needed to characterize ^1H T_1 relaxation times in proteins and how these change in pulse sequences that preserve the $^1\text{H}_2\text{O}$ magnetization along the +z-axis as a heat sink⁹⁻¹¹ to drive longitudinal relaxation versus pulses sequences in which $^1\text{H}_2\text{O}$ magnetization is dephased or that utilize samples dissolved in D_2O . The use of paramagnetic agents to enhance T_1 relaxation must also be further explored.^{12,13}

Future directions

1. We are developing biosynthetic methods for introducing stereoselective ^1H into the remaining amino acids: Glu, Gln, Pro, and Arg. We have tried several approaches to incorporate isolated ^1H - ^{12}C groups at the beta position of the α -ketoglutarate family of amino acids using various carbon sources and metabolic inhibitors. We tried growing *E. coli* cells on glutamate or α -ketoglutarate as carbon sources. However, the enzymatic activity of glutamate dehydrogenase exchanges glutamate α and β positions with solvent deuterons, leaving only the γ position with residual protonation from the pyruvate methyl group. The δ positions of arginine and proline and the γ of Lys are derived NADPH. This cofactor is required in the biosynthesis of several amino acids, impacting isotopic labelling at many different sites when protein expression is carried out in deuterated *E. coli* media, resulting in isotopic mixtures. Engineering a biosynthetic labelling scheme in *E. coli* for improving the labelling homogeneity at protein sites derived from NADPH would be an exciting future research direction.

There are five *E. coli* enzymes involved in the reduction of NADP^+ to NADPH: glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase of the oxidative pentose phosphate pathway, a malic enzyme, isocitrate dehydrogenase of the TCA cycle, and a membrane-bound proton-translocating transhydrogenase.^{14,15} The efficient synthesis of many amino acids and other commercial biotech products is limited by the rate of endogenous NADPH regeneration.^{16,17} One of the strategies for replenishing NADPH involves increasing metabolic flux through the pentose phosphate pathway and the TCA cycle and deleting competing pathways.^{14,15,18,19} Alternatively, genes encoding NADP-dependent enzymes can be overexpressed to channel significant flux toward cofactor regeneration.^{20,21} The addition of reduced compounds that can donate electrons to NADP^+ has also been shown to replenish NADPH in vitro.²² Finally, NADH-regenerating enzymes such as formate dehydrogenase and phosphite dehydrogenase have also been used to replenish NADPH.^{23,24} We could exploit these strategies to monitor improvements in isotope labeling at amino acid sites derived from NADPH.

2. Protein expression decreases in *E. coli* when TCA cycle intermediates are used as the sole energy source, severely affecting protein yield. This is a significant problem when expensive isotope labelling reagents are used. The bacteria lack the appropriate transporters to shuttle these

intermediates into the cell to be metabolized for energy. Developing an efficient *E. coli* strain that can utilize TCA cycle intermediates as an energy source will improve isotope labelling efficiency in the oxaloacetate and the α -ketoglutarate family of amino acids.

A recent study examined the growth kinetics of 3796 *E. coli* single-gene deletion mutants in 30 different carbon sources and found that deleting phosphoenol pyruvate (PEP) carboxylase (Δppc) hinders growth in most carbon sources except the TCA cycle intermediates: succinate, fumarate, malate, acetate, and α -ketoglutarate.²⁵ PEP carboxylase catalyzes an anaplerotic reaction which generates oxaloacetate from PEP, replenishing the substrates needed to drive the TCA cycle. This reaction is no longer needed when TCA cycle intermediates are used as carbon sources. In the future, this mutant can be examined for its isotope labelling efficiency on these carbon sources.

A different study identified conditions for making *E. coli* grow faster on secondary carbon sources using arginine or glutamate as a nitrogen source and small amounts of glucose. The bacteria utilize glucose first but grow faster after exhausting it, generating a reversed diauxic shift.²⁶ Glucose prevents *E. coli* from using other carbon sources, a phenomenon known as carbon catabolic repression.²⁷ Using glucose under nitrogen-limiting conditions leads to the build-up of TCA cycle intermediates, inhibiting cAMP synthesis by adenylate cyclase.^{26,28,29} This secondary messenger activates the transcription factor, Crp, which regulates over 200 operons with numerous functions, including alternative sugar utilization.^{30,31} In the future, we could grow *E. coli* on TCA cycle intermediates, add cAMP and monitor growth rate, protein yield, and labelling efficiency.

3. The issue of residual protonation at the Ile- γ_2 methyl group is known in the literature and supplementing chemically synthesized 2-hydroxy-2-ethyl-3-ketobutyrate precursor in *E. coli* growth media overcomes this problem. We have not solved the Ile- γ_2 methyl scrambling issue yet. Developing biosynthetic means of addressing this problem would be an exciting labelling scheme.

4. Obtaining sequence-specific resonance assignments of methyl groups remains a significant challenge in large protein structural and dynamic studies by solution NMR. The current methodologies rely on extensive mutagenesis or the divide-and-conquer approach for the assignment process. Introducing ^1H - ^{12}C groups strategically into proteins offers a more efficient and user-friendly approach to the assignment process, linking methyl groups to the amide resonances while extending the range of molecules amenable to study by NMR.

5. In our labelling scheme, ^1H - ^{12}C groups possess the longest-lived ^1H magnetization, followed by ^1H - ^{13}C methyl groups and then backbone ^1H - ^{15}N groups. We have already shown that these groups can be connected via NOE-based experiments, and NOE-based magnetization transfer improves efficiency with increasing molecular weight. The superiority of our labelling scheme will thus be substantially displayed when applied to tackle large protein systems (0.5-1.0 MDa). The obvious next step is to use the labelling scheme on the 20S or 11S proteasome and to develop improved methods for resonance assignment of our newly introduced ^1H magnetizations.

6. To fully exploit the benefits afforded by our labelling scheme, the sensitivity and resolution of our 3D NMR experiments can be significantly improved by designing pulse sequences that start on ^1H - ^{15}N amide groups or ^1H - ^{13}C methyl groups and end on the longest-lived ^1H magnetizations introduced by our isolated ^1H - ^{12}C groups. These developments in isotope labelling and improved assignment strategies will undoubtedly push solution NMR's size limitation to previously unachievable levels. Structural biologists can now use high-quality NMR data to complement X-ray crystallography and cryo-EM for studying high molecular weight proteins and their assemblies.

7. The discovery of novel metabolic precursors for protein labelling represents a promising opportunity for commercial development. One option could be to produce and sell the precursors for academic and industrial NMR spectroscopists. Another possibility could be to incorporate the precursors into existing NMR kits, making them more accessible and user-friendly for researchers. Another avenue for commercialization could be to partner with pharmaceutical companies to use the precursors in drug discovery efforts, as the improved NMR data provided

by the labelling could lead to the identification of new drug targets and the development of more effective treatments. Overall, the potential applications of the novel metabolic precursors are vast, and their commercial development will likely contribute to the advancement of NMR-based structural biology and drug discovery.

8. In conclusion, developing a method for labelling proteins with isolated ^1H - ^{12}C groups at non-methyl-containing amino acid sites using simple, efficient, and cheap metabolic precursors represent a significant advancement in NMR-based structural biology and drug discovery. This approach can enhance the accuracy and resolution of NMR spectroscopy, allowing for a more comprehensive understanding of protein structure and function. Furthermore, the simplicity and affordability of this labelling method make it accessible to a broader range of researchers, allowing for wider dissemination of NMR-based findings and more collaborative efforts in the field. The impact of this work is likely to be far-reaching, potentially leading to the discovery of new drug targets and the development of more effective treatments for a range of diseases.

References

- (1) Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* (80-.). **2001**, *294* (5540), 93–96. <https://doi.org/10.1126/science.1065659>.
- (2) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **2004**, *383* (2003), 66–93. [https://doi.org/10.1016/S0076-6879\(04\)83004-0](https://doi.org/10.1016/S0076-6879(04)83004-0).
- (3) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (4) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.
- (5) Ali, A. A. A. I.; Hoffmann, F.; Sch, L. V.; Mulder, F. A. A. Probing Methyl Group Dynamics in Proteins by NMR Cross- Correlated Dipolar Relaxation and Molecular Dynamics Simulations. *J Chem Theory Comput.* **2022**, *18* (12), 7722–7732. <https://doi.org/10.1021/acs.jctc.2c00568>.
- (6) Hoffmann, F.; Mulder, F. A. A.; Schäfer, L. V. Predicting NMR Relaxation of Proteins

- from Molecular Dynamics Simulations with Accurate Methyl Rotation Barriers. *J. Chem. Phys.* **2020**, *152* (8), 084102. <https://doi.org/10.1063/1.5135379>.
- (7) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. Consistent Blind Protein Structure Generation from NMR Chemical Shift Data. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (12), 4685–4690. <https://doi.org/10.1073/pnas.0800256105>.
 - (8) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (23), 9615–9620. <https://doi.org/10.1073/pnas.0610313104>.
 - (9) Kainosho, M.; Miyanoiri, Y.; Terauchi, T.; Takeda, M. Perspective: Next Generation Isotope-Aided Methods for Protein NMR Spectroscopy. *J. Biomol. NMR* **2018**, *71* (3), 119–127. <https://doi.org/10.1007/s10858-018-0198-x>.
 - (10) Miyanoiri, Y.; Takeda, M.; Terauchi, T.; Kainosho, M. Recent Developments in Isotope-Aided NMR Methods for Supramolecular Protein Complexes –SAIL Aromatic TROSY. *Biochim. Biophys. Acta - Gen. Subj.* **2020**, *1864* (2), 129439. <https://doi.org/10.1016/j.bbagen.2019.129439>.
 - (11) Feigensohn, G. W.; Chan, S. I. Nuclear Magnetic Relaxation Behavior of Lecithin Multilayers. *J. Am. Chem. Soc.* **1974**, *96* (5), 1312–1319. <https://doi.org/10.1021/ja00812a009>.
 - (12) Kocman, V.; Di Mauro, G. M.; Veglia, G.; Ramamoorthy, A. Use of Paramagnetic Systems to Speed-up NMR Data Acquisition and for Structural and Dynamic Studies. *Solid State Nucl. Magn. Reson.* **2019**, *102* (June), 36–46. <https://doi.org/10.1016/j.ssnmr.2019.07.002>.
 - (13) Cai, S.; Seu, C.; Kovacs, Z.; Sherry, A. D.; Chen, Y. Sensitivity Enhancement of Multidimensional NMR Experiments by Paramagnetic Relaxation Effects. *J. Am. Chem. Soc.* **2006**, *128* (5), 13474–13478. <https://doi.org/10.1021/ja0634526>.
 - (14) Aslan, S.; Noor, E.; Bar-Even, A. Holistic Bioengineering: Rewiring Central Metabolism for Enhanced Bioproduction. *Biochem. J.* **2017**, *474* (23), 3935–3950. <https://doi.org/10.1042/BCJ20170377>.
 - (15) Lindner, S. N.; Ramirez, L. C.; Krüsemann, J. L.; Yishai, O.; Belkhef, S.; He, H.; Bouzon, M.; Döring, V.; Bar-Even, A. NADPH-Auxotrophic *E. Coli*: A Sensor Strain for Testing in Vivo Regeneration of NADPH. *ACS Synth. Biol.* **2018**, *7* (12), 2742–2749. <https://doi.org/10.1021/acssynbio.8b00313>.
 - (16) Kabus, A.; Georgi, T.; Wendisch, V. F.; Bott, M. Expression of the *Escherichia Coli* PntAB Genes Encoding a Membrane-Bound Transhydrogenase in *Corynebacterium Glutamicum* Improves L-Lysine Formation. *Appl. Microbiol. Biotechnol.* **2007**, *75* (1), 47–53. <https://doi.org/10.1007/s00253-006-0804-9>.
 - (17) Park, S. H.; Kim, H. U.; Kim, T. Y.; Park, J. S.; Kim, S. S.; Lee, S. Y. Metabolic Engineering of *Corynebacterium Glutamicum* for L-Arginine Production. *Nat. Commun.* **2014**, *5*. <https://doi.org/10.1038/ncomms5618>.
 - (18) Spaans, S. K.; Weusthuis, R. A.; van der Oost, J.; Kengen, S. W. M. NADPH-Generating Systems in Bacteria and Archaea. *Front. Microbiol.* **2015**, *6* (JUL), 1–27. <https://doi.org/10.3389/fmicb.2015.00742>.
 - (19) Lee, W. H.; Kim, M. D.; Jin, Y. S.; Seo, J. H. Engineering of NADPH Regenerators in *Escherichia Coli* for Enhanced Biotransformation. *Appl. Microbiol. Biotechnol.* **2013**, *97*

- (7), 2761–2772. <https://doi.org/10.1007/s00253-013-4750-z>.
- (20) Fang, H.; Xie, X.; Xu, Q.; Zhang, C.; Chen, N. Enhancement of Cytidine Production by Coexpression of Gnd, Zwf, and Prs Genes in Recombinant Escherichia Coli CYT15. *Biotechnol. Lett.* **2013**, *35* (2), 245–251. <https://doi.org/10.1007/s10529-012-1068-3>.
- (21) Lim, S. J.; Jung, Y. M.; Shin, H. D.; Lee, Y. H. Amplification of the NADPH-Related Genes Zwf and Gnd for the Oddball Biosynthesis of PHB in an E. Coli Transformant Harboring a Cloned PhbCAB Operon. *J. Biosci. Bioeng.* **2002**, *93* (6), 543–549. [https://doi.org/10.1016/S1389-1723\(02\)80235-3](https://doi.org/10.1016/S1389-1723(02)80235-3).
- (22) Van Der Donk, W. A.; Zhao, H. Recent Developments in Pyridine Nucleotide Regeneration. *Curr. Opin. Biotechnol.* **2003**, *14* (4), 421–426. [https://doi.org/10.1016/S0958-1669\(03\)00094-6](https://doi.org/10.1016/S0958-1669(03)00094-6).
- (23) Tishkov, V. I.; Popov, V. O. Protein Engineering of Formate Dehydrogenase. *Biomol. Eng.* **2006**, *23* (2–3), 89–110. <https://doi.org/10.1016/j.bioeng.2006.02.003>.
- (24) Johannes, Tyler W. Woodyer, R. D.; Zhao, H. Efficient Regeneration of NADPH Using an Engineered Phosphite Dehydrogenase. *Biotechnol. Bioeng.* **2007**, *96* (1), 18–26. <https://doi.org/10.1002/bit.21168>.
- (25) Tong, M.; French, S.; El Zahed, S. S.; Ong, W. K.; Karp, P. D.; Brown, E. D. Gene Dispensability in Escherichia Coli Grown in Thirty Different Carbon Environments. *MBio* **2020**, *11* (5), 1–20. <https://doi.org/10.1128/mBio.02259-20>.
- (26) Bren, A.; Park, J. O.; Towbin, B. D.; Dekel, E.; Rabinowitz, J. D.; Alon, U. Glucose Becomes One of the Worst Carbon Sources for E.Coli on Poor Nitrogen Sources Due to Suboptimal Levels of CAMP. *Sci. Rep.* **2016**, *6* (April), 2–11. <https://doi.org/10.1038/srep24834>.
- (27) Stülke, J.; Hillen, W. Carbon Catabolite Repression in Bacteria. *Curr. Opin. Microbiol.* **1999**, *2* (2), 195–201. [https://doi.org/10.1016/S1369-5274\(99\)80034-4](https://doi.org/10.1016/S1369-5274(99)80034-4).
- (28) You, C.; Okano, H.; Hui, S.; Zhang, Z.; Kim, M.; Gunderson, C. W.; Wang, Y. P.; Lenz, P.; Yan, D.; Hwa, T. Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling. *Nature* **2013**, *500* (7462), 301–306. <https://doi.org/10.1038/nature12446>.
- (29) Daniel, J.; Danchin, A. 2-Ketoglutarate as a Possible Regulatory Metabolite Involved in Cyclic AMP-Dependent Catabolite Repression in Escherichia Coli K12. *Biochimie* **1986**, *68* (2), 303–310. [https://doi.org/10.1016/S0300-9084\(86\)80027-X](https://doi.org/10.1016/S0300-9084(86)80027-X).
- (30) Görke, B.; Stülke, J. Carbon Catabolite Repression in Bacteria: Many Ways to Make the Most out of Nutrients. *Nat. Rev. Microbiol.* **2008**, *6* (8), 613–624. <https://doi.org/10.1038/nrmicro1932>.
- (31) Kremling, A.; Geiselmann, J.; Ropers, D.; de Jong, H. Understanding Carbon Catabolite Repression in Escherichia Coli Using Quantitative Models. *Trends Microbiol.* **2015**, *23* (2), 99–109. <https://doi.org/10.1016/j.tim.2014.11.002>.

Bibliography

- Abraham, A. *The Principles of Nuclear Magnetism*; Clarendon Press, Oxford, **1961**.
- Aghazadeh, B.; Zhu, K.; Kubiseski, T. J.; Liu, G. A.; Pawson, T.; Zheng, Y.; Rosen, M. K. Structure and Mutagenesis of the Dbl Homology Domain. *Nat. Struct. Biol.* **1998**, *5* (12), 1098–1107. <https://doi.org/10.1038/4209>.
- Ahn, V. E.; Lo, E. I.; Engel, C. K.; Chen, L.; Hwang, P. M.; Kay, L. E.; Bishop, R. E.; Prive, G. G. A Hydrocarbon Ruler Measures Palmitate in the Enzymatic Acylation of Endotoxin. *EMBO J.* **2004**, *23* (15), 2931–2941. <https://doi.org/10.1038/sj.emboj.7600320>.
- Ali, A. A. A. I.; Hoffmann, F.; Sch, L. V.; Mulder, F. A. A. Probing Methyl Group Dynamics in Proteins by NMR Cross- Correlated Dipolar Relaxation and Molecular Dynamics Simulations. *J Chem Theory Comput.* **2022**, *18* (12), 7722–7732. <https://doi.org/10.1021/acs.jctc.2c00568>.
- Amrhein, N.; Deus, B.; Gehrke, P.; Steinrücken, H. C. The Site of the Inhibition of the Shikimate Pathway by Glyphosate. *Plant Physiol.* **1980**, *66* (5), 830–834. <https://doi.org/10.1104/pp.66.5.830>.
- Archer, S. J.; Ikura, M.; Torchia, D. A.; Bax, A. An Alternative 3D NMR Technique for Correlating Backbone ¹⁵N with Side Chain H β Resonances in Larger Proteins. *J. Magn. Reson.* **1991**, *95* (3), 636–641. [https://doi.org/10.1016/0022-2364\(91\)90182-S](https://doi.org/10.1016/0022-2364(91)90182-S).
- Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. Structure of Outer Membrane Protein A Transmembrane Domain by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8* (4), 334–338. <https://doi.org/10.1038/86214>.
- Aslan, S.; Noor, E.; Bar-Even, A. Holistic Bioengineering: Rewiring Central Metabolism for Enhanced Bioproduction. *Biochem. J.* **2017**, *474* (23), 3935–3950. <https://doi.org/10.1042/BCJ20170377>.
- Ayala, I.; Chiari, L.; Kerfah, R.; Boisbouvier, J.; Gans, P.; Hamelin, O. Asymmetric Synthesis of Methyl Specifically Labelled L-Threonine and Application to the NMR Studies of High Molecular Weight Proteins. *ChemistrySelect* **2020**, *5* (17), 5092–5098. <https://doi.org/10.1002/slct.202000827>.
- 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.* **2012**, *48* (10), 1434–1436. <https://doi.org/10.1039/c1cc12932e>.
- Ayala, I.; Sounier, R.; Usé, N.; Gans, P.; Boisbouvier, J. An Efficient Protocol for the Complete Incorporation of Methyl-Protonated Alanine in Perdeuterated Protein. *J. Biomol. NMR* **2009**, *43* (2), 111–119. <https://doi.org/10.1007/s10858-008-9294-7>.
- Babu, U. M.; Johnston, R. B. Nuclear Magnetic Resonance Studies of D₂O-Substrate Exchange Reactions Catalyzed by Glutamic Pyruvic and Glutamic Oxaloacetic

Transaminases. *Biochemistry* **1976**, 15 (25), 5671–5678. <https://doi.org/10.1021/bi00670a037>.

Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* (80). **2001**, 294 (5540), 93–96. <https://doi.org/10.1126/science.1065659>.

Bax, A. Homonuclear Hartmann-Hahn Experiments. *Methods Enzymol.* **1989**, 176 (C), 151–168. [https://doi.org/10.1016/0076-6879\(89\)76010-9](https://doi.org/10.1016/0076-6879(89)76010-9).

Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. Measurement of Homo- and Heteronuclear J Couplings from Quantitative J Correlation. *Methods Enzymol.* **1994**, 239 (C), 79–105. [https://doi.org/10.1016/S0076-6879\(94\)39004-5](https://doi.org/10.1016/S0076-6879(94)39004-5).

Bayrhuber, M.; Meins, T.; Habeck, M.; Becker, S.; Giller, K.; Villinger, S.; Vonrhein, C.; Griesinger, C.; Zweckstetter, M.; Zeth, K. Structure of the Human Voltage-Dependent Anion Channel. *Proc. Natl. Acad. Sci.* **2008**, 105 (40), 15370–15375. <https://doi.org/10.1073/pnas.0808115105>.

Bibow, S.; Hiller, S. A Guide to Quantifying Membrane Protein Dynamics in Lipids and Other Native-like Environments by Solution-State NMR Spectroscopy. *FEBS J.* **2019**, 286 (9), 1610–1623. <https://doi.org/10.1111/febs.14639>.

Bishop, R. E. The Lipid A Palmitoyltransferase PagP: Molecular Mechanisms and Role in Bacterial Pathogenesis. *Mol. Microbiol.* **2005**, 57 (4), 900–912. <https://doi.org/10.1111/j.1365-2958.2005.04711.x>.

Bishop, R. E.; Gibbons, H. S.; Guina, T.; Trent, M. S.; Miller, S. I.; Raetz, C. R. H. Transfer of Palmitate from Phospholipids to Lipid A in Outer Membranes of Gram-Negative Bacteria. *EMBO J.* **2000**, 19 (19), 5071–5080. <https://doi.org/10.1093/emboj/19.19.5071>.

Bothner-By, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **2002**, 106 (3), 811–813. <https://doi.org/10.1021/JA00315A069>.

Bren, A.; Park, J. O.; Towbin, B. D.; Dekel, E.; Rabinowitz, J. D.; Alon, U. Glucose Becomes One of the Worst Carbon Sources for E.Coli on Poor Nitrogen Sources Due to Suboptimal Levels of CAMP. *Sci. Rep.* **2016**, 6, 2–11. <https://doi.org/10.1038/srep24834>.

Brüschweiler, R.; Ernst, R. R. Molecular Dynamics Monitored by Cross-Correlated Cross Relaxation of Spins Quantized along Orthogonal Axes. *J. Chem. Phys* **1992**, 96, 1758. <https://doi.org/10.1063/1.462131>.

Cai, F.; Li, M. X.; Pineda-Sanabria, S. E.; Gelozia, S.; Lindert, S.; West, F.; Sykes, B. D.; Hwang, P. M. Structures Reveal Details of Small Molecule Binding to Cardiac Troponin. *J. Mol. Cell. Cardiol.* **2016**, 101, 134–144. <https://doi.org/10.1016/j.yjmcc.2016.10.016>.

Cai, M.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R. An Efficient and Cost-Effective Isotope Labeling Protocol for Proteins Expressed in Escherichia Coli. *J. Biomol. NMR* **1998**, 11, 97–102. <https://doi.org/10.1023/a:1008222131470>.

Cai, S.; Seu, C.; Kovacs, Z.; Sherry, A. D.; Chen, Y. Sensitivity Enhancement of Multidimensional NMR Experiments by Paramagnetic Relaxation Effects. *J. Am. Chem. Soc.* **2006**, 128 (5), 13474–13478. <https://doi.org/10.1021/ja0634526>.

Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104 (23), 9615–9620. <https://doi.org/10.1073/pnas.0610313104>.

Cavanagh, J.; Rance, M. Suppression of Cross-Relaxation Effects in TOCSY Spectra via a Modified DIPSI-2 Mixing Sequence. *J. Magn. Reson.* **1992**, 96 (3), 670–678. [https://doi.org/10.1016/0022-2364\(92\)90357-D](https://doi.org/10.1016/0022-2364(92)90357-D).

Chou, J. J.; Case, D. A.; Bax, A. Insights into the Mobility of Methyl-Bearing Side Chains in Proteins from 3JCC and 3JCN Couplings. *J. Am. Chem. Soc.* **2003**, 125 (29), 8959–8966. <https://doi.org/10.1021/ja029972s>.

Clore, G. M.; Bax, A.; Gronenborn, A. M. Stereospecific Assignment of B-Methylene Protons in Larger Proteins Using 3D ¹⁵N-Separated Hartmann-Hahn and ¹³C-Separated Rotating Frame Overhauser Spectroscopy. *J. Biomol. NMR* **1991**, 1, 13–22. <https://doi.org/10.1007/BF01874566>.

Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Wingfield, P. T. Analysis of the Backbone Dynamics of Interleukin-1 β Using Two-Dimensional Inverse Detected Heteronuclear ¹⁵N-¹H NMR Spectroscopy. *Biochemistry* **1990**, 29 (32), 7387–7401. <https://doi.org/10.1021/bi00484a006>.

Clore, G. M.; Gronenborn, A. M. NMR Structure Determination of Proteins and Protein Complexes Larger than 20 KDa. *Curr. Opin. Chem. Biol.* **1998**, 2 (5), 564–570. [https://doi.org/10.1016/S1367-5931\(98\)80084-7](https://doi.org/10.1016/S1367-5931(98)80084-7).

Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J. Am. Chem. Soc.* **1990**, 112 (12), 4989–4991. https://doi.org/10.1021/JA00168A070/ASSET/JA00168A070.FP.PNG_V03.

Conte, L. Lo; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol.* **1999**, 285 (5), 2177–2198. <https://doi.org/10.1006/jmbi.1998.2439>.

Cooper, A. J. L. Proton Magnetic Resonance Studies of Glutamate Alanine Transaminase Catalyzed Deuterium Exchange. Evidence for Proton Conservation during Prototropic Transfer from the α Carbon of L Alanine to the C4 Position of Pyridoxal 5' Phosphate. *J. Biol. Chem.* **1976**, 251 (4), 1088–1096. [https://doi.org/10.1016/s0021-9258\(17\)33805-x](https://doi.org/10.1016/s0021-9258(17)33805-x).

Cornell, N. W.; Zuurendonk, P. F.; Kerich, M. J.; Straight, C. B. Selective Inhibition of Alanine Aminotransferase and Aspartate Aminotransferase in Rat Hepatocytes. *Biochem. J.* **1984**, 220 (3), 707–716. <https://doi.org/10.1042/bj2200707>.

Dalvit, C. ^1H to ^{15}N Polarization Transfer via ^1H Chemical-Shift Anisotropy- ^1H - ^{15}N Dipole-Dipole Cross Correlation. *J. Magn. Reson.* **1992**, 97, 645–650. [https://doi.org/10.1016/0022-2364\(92\)90046-A](https://doi.org/10.1016/0022-2364(92)90046-A).

Daniel, J.; Danchin, A. 2-Ketoglutarate as a Possible Regulatory Metabolite Involved in Cyclic AMP-Dependent Catabolite Repression in Escherichia Coli K12. *Biochimie* **1986**, 68 (2), 303–310. [https://doi.org/10.1016/S0300-9084\(86\)80027-X](https://doi.org/10.1016/S0300-9084(86)80027-X).

Danmaliki, G. I.; Hwang, P. M. Proton TOCSY NMR Relaxation Rates Quantitate Protein Side Chain Mobility in the Pin1 WW Domain. *J. Biomol. NMR* **2022**, 76 (4), 121–135. <https://doi.org/10.1007/s10858-022-00400-5>.

Danmaliki, G. I.; Liu, P. B.; Hwang, P. M. Stereoselective Deuteration in Aspartate, Asparagine, Lysine, and Methionine Amino Acid Residues Using Fumarate as a Carbon Source for Escherichia Coli in D_2O . *Biochemistry* **2017**, 56 (45). <https://doi.org/10.1021/acs.biochem.7b00991>.

Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, 6 (3), 277–293. <https://doi.org/10.1007/BF00197809>.

Fang, H.; Xie, X.; Xu, Q.; Zhang, C.; Chen, N. Enhancement of Cytidine Production by Coexpression of Gnd, Zwf, and Prs Genes in Recombinant Escherichia Coli CYT15. *Biotechnol. Lett.* **2013**, 35 (2), 245–251. <https://doi.org/10.1007/s10529-012-1068-3>.

Farmer, B. T.; Venters, R. A. NMR of Perdeuterated Large Proteins. In: Krishna, N.R., Berliner, L.J. (eds) *Modern Techniques in Protein NMR*. Biological Magnetic Resonance, Springer, Boston, MA. **2002**, 16, 75–120. https://doi.org/10.1007/0-306-47083-7_3.

Farrow, N. A.; Muhandiram, R.; Pascal, S. M.; Kay, L. E.; Singer, A. U.; Forman-Kay, J. D.; Kay, C. M.; Gish, G.; Pawson, T.; Shoelson, S. E. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by ^{15}N NMR Relaxation. *Biochemistry* **1994**, 33 (19), 5984–6003. <https://doi.org/10.1021/bi00185a040>.

Farrow, N. A.; Muhandiram, R.; Pascal, S. M.; Kay, L. E.; Singer, A. U.; Forman-Kay, J. D.; Kay, C. M.; Gish, G.; Pawson, T.; Shoelson, S. E. Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by ^{15}N NMR Relaxation. *Biochemistry* **1994**, 33 (19), 5984–6003. https://doi.org/10.1021/BI00185A040/SUPPL_FILE/BI00185A040_SI_001.PDF.

Feigenson, G. W.; Chan, S. I. Nuclear Magnetic Relaxation Behavior of Lecithin Multilayers. *J. Am. Chem. Soc.* **1974**, 96 (5), 1312–1319. <https://doi.org/10.1021/ja00812a009>.

Fernández, C.; Adeishvili, K.; Wüthrich, K. Transverse Relaxation-Optimized NMR Spectroscopy with the Outer Membrane Protein OmpX in Dihexanoyl Phosphatidylcholine Micelles. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98 (5), 2358–2363. <https://doi.org/10.1073/pnas.051629298>.

Fernández, C.; Hilty, C.; Wider, G.; Güntert, P.; Wüthrich, K. NMR Structure of the Integral Membrane Protein OmpX. *J. Mol. Biol.* **2004**, 336 (5), 1211–1221. <https://doi.org/10.1016/j.jmb.2003.09.014>.

Fernández, C.; Wider, G. TROSY in NMR Studies of the Structure and Function of Large Biological Macromolecules. *Curr. Opin. Struct. Biol.* **2003**, 13 (5), 570–580. <https://doi.org/10.1016/j.sbi.2003.09.009>.

Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. NMR Analysis of a 900K GroEL-GroES Complex. *Nature* **2002**, 418 (6894), 207–211. <https://doi.org/10.1038/nature00860>.

Fischer, M.; Kloiber, K.; Häusler, J.; Ledolter, K.; Konrat, R.; Schmid, W. Synthesis of a ¹³C-Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. *ChemBioChem* **2007**, 8 (6), 610–612. <https://doi.org/10.1002/cbic.200600551>.

Foster, M. P.; McElroy, C. A.; Amero, C. D. Solution NMR of Large Molecules and Assemblies. *Biochemistry* **2007**, 46 (2), 331–340. <https://doi.org/10.1021/bi0621314>.

Gans, P.; Hamelin, O.; Sounier, R.; Ayala, I.; Durá, M. A.; Amero, C. D.; Noirclerc-Savoie, M.; Franzetti, B.; Plevin, M. J.; Boisbouvier, J. Stereospecific Isotopic Labeling of Methyl Groups for NMR Spectroscopic Studies of High-Molecular-Weight Proteins. *Angew. Chemie - Int. Ed.* **2010**, 49 (11), 1958–1962. <https://doi.org/10.1002/anie.200905660>.

Gardner, K. H.; Kay, L. E. Production and Incorporation of ¹⁵N, ¹³C, ²H (¹H- δ ¹ Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **1997**, 119 (32), 7599–7600.

Gardner, K. H.; Kay, L. E. The Use of ²H, ¹³C, ¹⁵N Multidimensional NMR to Study the Structure and Dynamics of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, 27, 357–406. <https://doi.org/10.1146/annurev.biophys.27.1.357>.

Gardner, K. H.; Rosen, M. K.; Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR. *Biochemistry* **1997**, 36 (6), 1389–1401. <https://doi.org/10.1021/bi9624806>.

Gauto, D. F.; Macek, P.; Barducci, A.; Fraga, H.; Hessel, A.; Terauchi, T.; Gajan, D.; Miyanoiri, Y.; Boisbouvier, J.; Lichtenecker, R.; Kainosho, M.; Schanda, P. Aromatic Ring Dynamics, Thermal Activation, and Transient Conformations of a 468 KDa Enzyme by Specific ¹H-¹³C Labeling and Fast Magic-Angle Spinning NMR. *J. Am. Chem. Soc.* **2019**, 141 (28), 11183–11195. <https://doi.org/10.1021/jacs.9b04219>.

Gelis, I.; Bonvin, A. M. J. J.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **2007**, 131 (4), 756–769. <https://doi.org/10.1016/j.cell.2007.09.039>.

Gifford, J. L.; Ishida, H.; Vogel, H. J. Fast Methionine-Based Solution Structure Determination of Calcium-Calmodulin Complexes. *J. Biomol. NMR* **2011**, 50 (1), 71–81. <https://doi.org/10.1007/s10858-011-9495-3>.

Golichowski, A.; Harruff, R. C.; Jenkins, W. T. The Effects of PH on the Rates of Isotope Exchange Catalyzed by Alanine Aminotransferase. *Arch. Biochem. Biophys.* **1977**, 178 (2), 459–467. [https://doi.org/10.1016/0003-9861\(77\)90216-8](https://doi.org/10.1016/0003-9861(77)90216-8).

Görke, B.; Stülke, J. Carbon Catabolite Repression in Bacteria: Many Ways to Make the Most out of Nutrients. *Nat. Rev. Microbiol.* **2008**, 6 (8), 613–624. <https://doi.org/10.1038/nrmicro1932>.

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 (δ 1). *J. Biomol. NMR* **1999**, No. 13, 369–374. <https://doi.org/10.1023/a:1008393201236>.

Goto, N. K.; Kay, L. E. New Developments in Isotope Labeling Strategies for Protein Solution NMR Spectroscopy. *Curr. Opin. Struct. Biol.* **2000**, 10 (5), 585–592. [https://doi.org/10.1016/S0959-440X\(00\)00135-4](https://doi.org/10.1016/S0959-440X(00)00135-4).

Gross, J. D.; Gelev, V. M.; Wagner, G. A Sensitive and Robust Method for Obtaining Intermolecular NOEs between Side Chains in Large Protein Complexes. *J. Biomol. NMR* **2003**, 25 (3), 235–242. <https://doi.org/10.1023/A:1022890112109>.

Grzesiek, S.; Ikura, M.; Marius Clore, G.; Gronenborn, A. M.; Bax, A. A 3D Triple-Resonance NMR Technique for Qualitative Measurement of Carbonyl- $H\beta$ J Couplings in Isotopically Enriched Proteins. *J. Magn. Reson.* **1992**, 96 (1), 215–221. [https://doi.org/10.1016/0022-2364\(92\)90307-S](https://doi.org/10.1016/0022-2364(92)90307-S).

Guo, L.; Lim, K. B.; Poduje, C. M.; Daniel, M.; Gunn, J. S.; Hackett, M.; Miller, S. I. Lipid A Acylation and Bacterial Resistance against Vertebrate Antimicrobial Peptides. *Cell* **1998**, 95 (2), 189–198. [https://doi.org/10.1016/S0092-8674\(00\)81750-X](https://doi.org/10.1016/S0092-8674(00)81750-X).

Hajduk, P. J.; Augeri, D. J.; Mack, J.; Mendoza, R.; Yang, J.; Betz, S. F.; Fesik, S. W. NMR-Based Screening of Proteins Containing ^{13}C -Labeled Methyl Groups. *J. Am. Chem. Soc.* **2000**, 122 (33), 7898–7904. <https://doi.org/10.1021/JA000350L>.

Hattab, G.; Warschawski, D. E.; Moncoq, K.; Miroux, B. Escherichia Coli as Host for Membrane Protein Structure Determination: A Global Analysis. *Sci. Rep.* **2015**, 5, 12097. <https://doi.org/10.1038/srep12097>.

Hayashi, H.; Inoue, K.; Nagata, T.; Kuramitsu, S.; Kagamiyama, H. Escherichia Coli Aromatic Amino Acid Aminotransferase: Characterization and Comparison with Aspartate Aminotransferase. *Biochemistry* **1993**, 32 (45), 12229–12239. <https://doi.org/10.1021/bi00096a036>.

Hiyama, Y.; Niu, C. H.; Silverton, J. V.; Bavoso, A.; Torchia, D. A. Determination of ^{15}N Chemical Shift Tensor via ^{15}N - 2H Dipolar Coupling in Boc-Glycylglycyl[^{15}N]Glycine

Benzyl Ester. *J. Am. Chem. Soc.* **1988**, 110 (8), 2378–2383. <https://doi.org/10.1021/ja00216a006>.

Hoffmann, F.; Mulder, F. A. A.; Schäfer, L. V. Predicting NMR Relaxation of Proteins from Molecular Dynamics Simulations with Accurate Methyl Rotation Barriers. *J. Chem. Phys.* **2020**, 152 (8), 084102. <https://doi.org/10.1063/1.5135379>

Hu, H.; Hermans, J.; Lee, A. L. Relating Side-Chain Mobility in Proteins to Rotameric Transitions: Insights from Molecular Dynamics Simulations and NMR. *J. Biomol. NMR* **2005**, 32 (2), 151–162. <https://doi.org/10.1007/s10858-005-5366-0>.

Hu, J. S.; Bax, A. Determination of ϕ and χ_1 Angles in Proteins from ^{13}C - ^{13}C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc.* **1997**, 119 (27), 6360–6368. <https://doi.org/10.1021/ja970067v>.

Hwang, P. M.; Bishop, R. E.; Kay, L. E. The Integral Membrane Enzyme PagP Alternates between Two Dynamically Distinct States. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101 (26), 9618–9623. <https://doi.org/10.1073/pnas.0402324101>.

Hwang, P. M.; Choy, W.; Lo, E. I.; Chen, L.; Forman-kay, J. D.; Raetz, C. R. H.; Bishop, R. E.; Kay, L. E.; Prive, G. G. Solution Structure and Dynamics of the Outer Membrane Enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 13460–13565. <https://doi.org/10.1073/pnas.212344499>.

Jäger, 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. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103 (28), 10648–10653. <https://doi.org/10.1073/pnas.0600511103>.

Johannes, Tyler W. Woodyer, R. D.; Zhao, H. Efficient Regeneration of NADPH Using an Engineered Phosphite Dehydrogenase. *Biotechnol. Bioeng.* **2007**, 96 (1), 18–26. <https://doi.org/10.1002/bit.21168>.

Johnson, B. A.; Blevins, R. A. NMR View: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR* **1994**, 4 (5), 603–614. <https://doi.org/10.1007/BF00404272>.

Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

Jursic, B. S.; Sagiraju, S.; Ancalade, D. K.; Clark, t.; Stevens, E. D. Practical Preparation of Z- α -(N-Acetylamino)- and Z- α -(N-Benzoylamino)- α,β -unsaturated Acids. *Synthetic Communications* **2007**, 37 (10), 1709-1714. <https://doi.org/10.1080/00397910701265895>

Kabus, A.; Georgi, T.; Wendisch, V. F.; Bott, M. Expression of the Escherichia Coli PntAB Genes Encoding a Membrane-Bound Transhydrogenase in Corynebacterium Glutamicum Improves L-Lysine Formation. *Appl. Microbiol. Biotechnol.* **2007**, 75 (1), 47–53. <https://doi.org/10.1007/s00253-006-0804-9>.

Kainosho, M.; Miyanoiri, Y.; Terauchi, T.; Takeda, M. Perspective: Next Generation Isotope-Aided Methods for Protein NMR Spectroscopy. *J. Biomol. NMR* **2018**, 71 (3), 119–127. <https://doi.org/10.1007/s10858-018-0198-x>.

Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P. Optimal Isotope Labelling for NMR Protein Structure Determinations. *Nature* **2006**, 440 (7080), 52–57. <https://doi.org/10.1038/nature04525>.

Kasperek, G. J.; Pratt, R. F. The Fumarase Reaction. A Nuclear Magnetic Resonance Experiment for Biological Chemistry Students. *J. Chem. Educ.* **1977**, 54 (8), 515. <https://doi.org/10.1021/ed054p515>.

Kawale, A. A.; Burmann, B. M. Characterization of Backbone Dynamics Using Solution NMR Spectroscopy to Discern the Functional Plasticity of Structurally Analogous Proteins. *STAR Protoc.* **2021**, 2 (4). <https://doi.org/10.1016/j.xpro.2021.100919>.

Kay, L. E. Solution NMR Spectroscopy of Supra-Molecular Systems, Why Bother? A Methyl-TROSY View. *J. Magn. Reson.* **2011**, 210 (2), 159–170. <https://doi.org/10.1016/j.jmr.2011.03.008>.

Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.* **1990**, 89 (3), 496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-5](https://doi.org/10.1016/0022-2364(90)90333-5).

Kay, L. E.; Torchia, D. A.; Bax, A. Backbone Dynamics of Proteins As Studied by ¹⁵N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **1989**, 28 (23), 8972–8979. <https://doi.org/10.1021/bi00449a003>.

Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. Methyl-Specific Isotopic Labeling: A Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* **2015**, 32, 113–122. <https://doi.org/10.1016/j.sbi.2015.03.009>.

Kleckner, I. R.; Foster, M. P. An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, 1814 (8), 942–968. <https://doi.org/10.1016/j.bbapap.2010.10.012>.

Kocman, V.; Di Mauro, G. M.; Veglia, G.; Ramamoorthy, A. Use of Paramagnetic Systems to Speed-up NMR Data Acquisition and for Structural and Dynamic Studies. *Solid State Nucl. Magn. Reson.* **2019**, 102 (June), 36–46. <https://doi.org/10.1016/j.ssnmr.2019.07.002>.

Kotsuki, H.; Datta, P. K.; Hayakawa, H.; Suenaga, H. An Efficient Procedure for Palladium-Catalyzed Reduction of Aryl/Enol Triflates. *Synthesis* **1995**, 11, 1348-1350. <https://doi.org/10.1055/s-1995-4120>

Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein Dynamics and Function from Solution State NMR Spectroscopy. *Q. Rev. Biophys.* **2016**, 49, e6. <https://doi.org/10.1017/S0033583516000019>.

Kowalski, J. A.; Liu, K.; Kelly, J. W. NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the x-Ray Crystal Structures of Pin1. *Biopolymers* **2002**, 63 (2), 111–121. <https://doi.org/10.1002/bip.10020>.

Kremling, A.; Geiselmann, J.; Ropers, D.; de Jong, H. Understanding Carbon Catabolite Repression in Escherichia Coli Using Quantitative Models. *Trends Microbiol.* **2015**, 23 (2), 99–109. <https://doi.org/10.1016/j.tim.2014.11.002>.

Lee, W. H.; Kim, M. D.; Jin, Y. S.; Seo, J. H. Engineering of NADPH Regenerators in Escherichia Coli for Enhanced Biotransformation. *Appl. Microbiol. Biotechnol.* **2013**, 97 (7), 2761–2772. <https://doi.org/10.1007/s00253-013-4750-z>.

LeMaster, D. M. Chiral β and Random Fractional Deuteration for the Determination of Protein Sidechain Conformation by NMR. *FEBS Lett.* **1987**, 223 (1), 191–196. [https://doi.org/10.1016/0014-5793\(87\)80534-3](https://doi.org/10.1016/0014-5793(87)80534-3).

LeMaster, D. M.; Richards, F. M. NMR Sequential Assignment of Escherichia Coli Thioredoxin Utilizing Random Fractional Deuteriation. *Biochemistry* **1988**, 27 (1), 142–150. <https://doi.org/10.1021/bi00401a022>.

Lesovoy, D. M.; Dubinnyi, M. A.; Nolde, S. B.; Bocharov, E. V.; Arseniev, A. S. Accurate Measurement of Dipole/Dipole Transverse Cross-Correlated Relaxation Γ_2 in Methylenes and Primary Amines of Uniformly ^{13}C / ^{15}N -Labeled Proteins. *J. Biomol. NMR* **2019**, 73 (5), 245–260. <https://doi.org/10.1007/s10858-019-00252-6>.

Li, F.; Grishaev, A.; Ying, J.; Bax, A. Side Chain Conformational Distributions of a Small Protein Derived from Model-Free Analysis of a Large Set of Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, 137 (46), 14798–14811. <https://doi.org/10.1021/jacs.5b10072>.

Lichtenecker, R. J. Synthesis of Aromatic $^{13}\text{C}/^2\text{H}$ - α -Ketoacid Precursors to Be Used in Selective Phenylalanine and Tyrosine Protein Labelling. *Org. Biomol. Chem.* **2014**, 12 (38), 7551–7560. <https://doi.org/10.1039/c4ob01129e>.

Lichtenecker, R. J.; Coudevylle, N.; Konrat, R.; Schmid, W. Selective Isotope Labelling of Leucine Residues by Using α -Ketoacid Precursor Compounds. *ChemBioChem* **2013**, 14 (7), 818–821. <https://doi.org/10.1002/cbic.201200737>.

Lichtenecker, R. J.; Weinhäupl, K.; Schmid, W.; Konrat, R. α -Ketoacids as Precursors for Phenylalanine and Tyrosine Labelling in Cell-Based Protein Overexpression. *J. Biomol. NMR* **2013**, 57 (4), 327–331. <https://doi.org/10.1007/s10858-013-9796-9>.

Lichtenecker, R.; Ludwiczek, M. L.; Schmid, W.; Konrat, R. Simplification of Protein NOESY Spectra Using Bioorganic Precursor Synthesis and NMR Spectral Editing. *J. Am. Chem. Soc.* **2004**, 126 (17), 5348–5349. <https://doi.org/10.1021/ja049679n>.

- Lim, S. J.; Jung, Y. M.; Shin, H. D.; Lee, Y. H. Amplification of the NADPH-Related Genes *Zwf* and *Gnd* for the Oddball Biosynthesis of PHB in an E. Coli Transformant Harboring a Cloned *PhbCAB* Operon. *J. Biosci. Bioeng.* **2002**, 93 (6), 543–549. [https://doi.org/10.1016/S1389-1723\(02\)80235-3](https://doi.org/10.1016/S1389-1723(02)80235-3).
- Lindner, S. N.; Ramirez, L. C.; Krüsemann, J. L.; Yishai, O.; Belkhef, S.; He, H.; Bouzon, M.; Döring, V.; Bar-Even, A. NADPH-Auxotrophic E. Coli: A Sensor Strain for Testing in Vivo Regeneration of NADPH. *ACS Synth. Biol.* **2018**, 7 (12), 2742–2749. <https://doi.org/10.1021/acssynbio.8b00313>.
- Linsler, 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.* **2014**, 136 (32), 11308–11310. <https://doi.org/10.1021/ja504791j>.
- Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, 104 (17), 4559–4570. <https://doi.org/10.1021/ja00381a010>.
- Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, 104 (17), 4546–4559. <https://doi.org/10.1021/ja00381a009>.
- Liu, W.; Zheng, Y.; Cistola, D. P.; Yang, D. Measurement of Methyl ¹³C-¹H Cross-Correlation in Uniformly ¹³C-, ¹⁵N-, Labeled Proteins Weidong. *J. Biomol. NMR* **2003**, 27, 351–364. <https://doi.org/10.1023/a:1025884922203>.
- Malaisse, W. J.; Malaisse-Lagae, F.; Liemans, V.; Ottinger, R.; Willem, R. Phosphoglucosyltransferase-Catalyzed Interconversion of Hexose Phosphates: Isotopic Discrimination between Hydrogen and Deuterium. *Mol. Cell. Biochem.* **1990**, 93 (2), 153–165. <https://doi.org/10.1007/BF00226187>.
- Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J. Specific Labeling and Assignment Strategies of Valine Methyl Groups for NMR Studies of High Molecular Weight Proteins. *J. Biomol. NMR* **2013**, 57 (3), 251–262. <https://doi.org/10.1007/s10858-013-9785-z>.
- Massi, F.; Johnson, E.; Wang, C.; Rance, M.; Palmer, A. G. NMR R_{1ρ} Rotating-Frame Relaxation with Weak Radio Frequency Fields. *J. Am. Chem. Soc.* **2004**, 126 (7), 2247–2256. <https://doi.org/10.1021/ja038721w>.
- Melville, D. B.; Rachele, J. R.; Keller, E. B. A Synthesis of Methionine Containing Radiocarbon in the Methyl Group. *J. Biol. Chem.* **1947**, 169 (2), 419–426. [https://doi.org/10.1016/s0021-9258\(17\)35042-1](https://doi.org/10.1016/s0021-9258(17)35042-1).
- 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.* **2004**, 126 (34), 10560–10570. <https://doi.org/10.1021/ja047904v>.
- Mildvan, A. S.; Scrutton, M. C.; Utter, M. F. Pyruvate Carboxylase. *J. Biol. Chem.* **1966**, 241 (15), 3488–3498. [https://doi.org/10.1016/s0021-9258\(18\)99859-5](https://doi.org/10.1016/s0021-9258(18)99859-5).

Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. Interior and Surface of Monomeric Proteins. *J. Mol. Biol.* **1987**, 196 (3), 641–656. [https://doi.org/10.1016/0022-2836\(87\)90038-6](https://doi.org/10.1016/0022-2836(87)90038-6).

Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in ¹³C-Labeled and Fractionally ²H-Enriched Proteins in Solution. *J. Am. Chem. Soc.* **2002**, 124 (22), 6439–6448. <https://doi.org/10.1021/ja012497y>.

Mittermaier, A.; Kay, L. E. X1 Torsion Angle Dynamics in Proteins, From Dipolar Couplings. *J. Am. Chem. Soc.* **2001**, 123 (28), 6892–6903. <https://doi.org/10.1021/ja010595d>.

Miyanoiri, Y.; Takeda, M.; Jee, J.; Ono, A. M.; Okuma, K.; Terauchi, T.; Kainosho, M. Alternative SAIL-Trp for Robust Aromatic Signal Assignment and Determination of the X2 Conformation by Intra-Residue NOEs. *J. Biomol. NMR* **2011**, 51 (4), 425–435. <https://doi.org/10.1007/s10858-011-9568-3>.

Miyanoiri, Y.; Takeda, M.; Okuma, K.; Ono, A. M.; Terauchi, T.; Kainosho, M. Differential Isotope-Labeling for Leu and Val Residues in a Protein by E. Coli Cellular Expression Using Stereo-Specifically Methyl Labeled Amino Acids. *J. Biomol. NMR* **2013**, 57 (3), 237–249. <https://doi.org/10.1007/s10858-013-9784-0>.

Miyanoiri, Y.; Takeda, M.; Terauchi, T.; Kainosho, M. Recent Developments in Isotope-Aided NMR Methods for Supramolecular Protein Complexes –SAIL Aromatic TROSY. *Biochim. Biophys. Acta - Gen. Subj.* **2020**, 1864 (2), 129439. <https://doi.org/10.1016/j.bbagen.2019.129439>.

Montelione, G. T.; Winkler, M. E.; Rauenbuehler, P.; Wagner, G. Accurate Measurements of Long-Range Heteronuclear Coupling Constants from Homonuclear 2D NMR Spectra of Isotope-Enriched Proteins. *J. Magn. Reson.* **1989**, 82 (1), 198–204. [https://doi.org/10.1016/0022-2364\(89\)90183-2](https://doi.org/10.1016/0022-2364(89)90183-2).

Morris, G. A.; Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **1979**, 101 (3), 760–762. <https://doi.org/10.1021/ja00497a058>.

Muhandiram, D. R.; Yamazaki, T.; Kay, L. E.; Sykes, B. D. Measurement of ²H T1 and T1ρ Relaxation Times in Uniformly ¹³C-Labeled and Fractionally ²H-Labeled Proteins in Solution. *J. Am. Chem. Soc.* **1995**, 117 (46), 11536–11544. <https://doi.org/10.1021/ja00151a018>.

Narindrasorasak, S.; Bridger, W. A. Probes of the Structure of Phosphoenolpyruvate Synthetase: Effects of a Transition State Analogue on Enzyme Conformation. *Can. J. Biochem.* **1978**, 56 (8), 816–819. <https://doi.org/10.1139/o78-124>.

Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wüthrich, K. Stereospecific Nuclear Magnetic Resonance Assignments of the Methyl Groups of Valine and Leucine in the DNA-Binding Domain of the 434 Repressor by Biosynthetically Directed Fractional ¹³C Labeling. *Biochemistry* **1989**, 28 (19), 7510–7516. <https://doi.org/10.1021/bi00445a003>.

Ogata, K.; Yajima, Y.; Nakamura, S.; Kaneko, R.; Goto, M.; Ohshima, T.; Yoshimune, K. Inhibition of Homoserine Dehydrogenase by Formation of a Cysteine-NAD Covalent Complex. *Sci. Rep.* **2018**, 8 (1), 1–2. <https://doi.org/10.1038/s41598-018-24063-1>.

Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: Explanation and Experimental Verification. *Magn. Reson. Chem.* **2003**, 41 (10), 843–852. <https://doi.org/10.1002/mrc.1256>.

Oschkinat, H.; Griesinger, C.; Kraulis, P. J.; Sørensen, O. W.; Ernst, R. R.; Gronenborn, A. M.; Clore, G. M. Three-Dimensional NMR Spectroscopy of a Protein in Solution. *Nature* **1988**, 332 (6162), 374–376. <https://doi.org/10.1038/332374a0>.

Otten, R.; Chu, B.; Krewulak, K. D.; Vogel, H. J.; Mulder, F. A. A. Comprehensive and Cost-Effective NMR Spectroscopy of Methyl Groups in Large Proteins. *J. Am. Chem. Soc.* **2010**, 132 (9), 2952–2960. <https://doi.org/10.1021/ja907706a>.

Palmer Arthur G, I. I. I. Probing Molecular Motion by NMR. *Curr. Opin. Struct. Biol.* **1997**, 7 (5), 732–737.

Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **2004**, 104 (8), 3623–3640. <https://doi.org/10.1021/cr030413t>.

Palmer, A. G.; Massi, F. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* **2006**, 106 (5), 1700–1719. <https://doi.org/10.1021/cr0404287>.

Pandey, A.; Shin, K.; Patterson, R. E.; Liu, X. Q.; Rainey, J. K. Current Strategies for Protein Production and Purification Enabling Membrane Protein Structural Biology1. *Biochem. Cell Biol.* **2016**, 94 (6), 507–527. <https://doi.org/10.1139/bcb-2015-0143>.

Pardee, A. B.; Potter, V. R. Malonate Inhibition of Oxidations in the Krebs Tricarboxylic Acid Cycle. *J. Biol. Chem.* **1949**, 178 (1), 241–250. [https://doi.org/10.1016/s0021-9258\(18\)56954-4](https://doi.org/10.1016/s0021-9258(18)56954-4).

Park, S. H.; Kim, H. U.; Kim, T. Y.; Park, J. S.; Kim, S. S.; Lee, S. Y. Metabolic Engineering of *Corynebacterium Glutamicum* for L-Arginine Production. *Nat. Commun.* **2014**, 5. <https://doi.org/10.1038/ncomms5618>.

PDB Statistics <https://www.rcsb.org/stats>.

Pervushin, K. Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a Technique in Structural Biology. *Q. Rev. Biophys.* **2000**, 33 (2), 161–197. <https://doi.org/10.1017/S0033583500003619>.

Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, 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. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (23), 12366–12371. <https://doi.org/10.1073/pnas.94.23.12366>.

- 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.* **1998**, 120 (25), 6394–6400. <https://doi.org/10.1021/ja980742g>.
- Renault, M.; Saurel, O.; Czaplicki, J.; Demange, P.; Gervais, V.; Löhr, F.; Réat, V.; Piotto, M.; Milon, A. Solution State NMR Structure and Dynamics of KpOmpA, a 210 Residue Transmembrane Domain Possessing a High Potential for Immunological Applications. *J. Mol. Biol.* **2009**, 385 (1), 117–130. <https://doi.org/10.1016/j.jmb.2008.10.021>.
- Renner, C.; Schleicher, M.; Moroder, L.; Holak, T. A. Practical Aspects of the 2D ^{15}N - $\{^1\text{H}\}$ NOE Experiment. *J. Biomol. NMR* **2002**, 23 (1), 23–33.
- Riek, R.; Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. Solution NMR Techniques for Large Molecular and Supramolecular Structures. *J. Am. Chem. Soc.* **2002**, 124 (41), 12144–12153. <https://doi.org/10.1021/ja026763z>.
- Riek, R.; Wider, G.; Pervushin, K.; Wüthrich, K. W. Polarization Transfer by Cross-Correlated Relaxation in Solution NMR with Very Large Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 4918–4923. <https://doi.org/10.1073/pnas.96.9.4918>.
- Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **2004**, 383 (2003), 66–93. [https://doi.org/10.1016/S0076-6879\(04\)83004-0](https://doi.org/10.1016/S0076-6879(04)83004-0).
- Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* **1996**, 263 (5), 627–636. <https://doi.org/10.1006/jmbi.1996.0603>.
- Ruschak, A. M.; Velyvis, A.; Kay, L. E. A Simple Strategy for ^{13}C , ^1H Labeling at the Ile- γ_2 Methyl Position in Highly Deuterated Proteins. *J. Biomol. NMR* **2010**, 48 (3), 129–135. <https://doi.org/10.1007/s10858-010-9449-1>.
- Sattler, M.; Fesik, S. W. Use of Deuterium Labeling in NMR: Overcoming a Sizeable Problem. *Structure* **1996**, 4 (11), 1245–1249. [https://doi.org/10.1016/s0969-2126\(96\)00133-5](https://doi.org/10.1016/s0969-2126(96)00133-5).
- Sauer, U.; Eikmanns, B. J. The PEP-Pyruvate-Oxaloacetate Node as the Switch Point for Carbon Flux Distribution in Bacteria. *FEMS Microbiol. Rev.* **2005**, 29 (4), 765–794. <https://doi.org/10.1016/j.femsre.2004.11.002>.
- Schlegel, S.; Hjelm, A.; Baumgarten, T.; Vikström, D.; De Gier, J. W. Bacterial-Based Membrane Protein Production. *Biochim. Biophys. Acta - Mol. Cell Res.* **2014**, 1843 (8) 1739–1749. <https://doi.org/10.1016/j.bbamcr.2013.10.023>.
- Schönbrunn, E.; Eschenburg, S.; Shuttleworth, W. A.; Schloss, J. V.; Amrhein, N.; Evans, J. N. S.; Kabsch, W. Interaction of the Herbicide Glyphosate with Its Target Enzyme 5-Enolpyruvylshikimate 3-Phosphate Synthase in Atomic Detail. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98 (4), 1376–1380. <https://doi.org/10.1073/pnas.98.4.1376>.

Schörghuber, J.; Geist, L.; Bisaccia, M.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Anthranilic Acid, the New Player in the Ensemble of Aromatic Residue Labeling Precursor Compounds. *J. Biomol. NMR* **2017**, 69 (1), 13–22. <https://doi.org/10.1007/s10858-017-0129-2>.

Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late Metabolic Precursors for Selective Aromatic Residue Labeling. *J. Biomol. NMR* **2018**, 71 (3), 129–140. <https://doi.org/10.1007/s10858-018-0188-z>.

Schörghuber, J.; Geist, L.; Platzer, G.; Konrat, R.; Lichtenecker, R. J. Highly Selective Stable Isotope Labeling of Histidine Residues by Using a Novel Precursor in E. Coli-Based Overexpression Systems. *ChemBioChem* **2017**, 18 (15), 1487–1491. <https://doi.org/10.1002/cbic.201700192>.

Schörghuber, J.; Sára, T.; Bisaccia, M.; Schmid, W.; Konrat, R.; Lichtenecker, R. J. Novel Approaches in Selective Tryptophan Isotope Labeling by Using Escherichia Coli Overexpression Media. *ChemBioChem* **2015**, 16 (5), 746–751. <https://doi.org/10.1002/cbic.201402677>.

Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, 577 (7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>.

Shekhtman, A.; Ghose, R.; Goger, M.; Cowburn, D. NMR Structure Determination and Investigation Using a Reduced Proton (REDPRO) Labeling Strategy for Proteins. *FEBS Lett.* **2002**, 524 (1–3), 177–182. [https://doi.org/10.1016/S0014-5793\(02\)03051-X](https://doi.org/10.1016/S0014-5793(02)03051-X).

Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. Consistent Blind Protein Structure Generation from NMR Chemical Shift Data. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105 (12), 4685–4690. <https://doi.org/10.1073/pnas.0800256105>.

Sinha, K.; Jen-Jacobson, L.; Rule, G. S. Specific Labeling of Threonine Methyl Groups for NMR Studies of Protein-Nucleic Acid Complexes. *Biochemistry* **2011**, 50 (47), 10189–10191. <https://doi.org/10.1021/bi201496d>.

Skrynnikov, N. R.; Millet, O.; Kay, L. E. Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions. *J. Am. Chem. Soc.* **2002**, 124 (22), 6449–6460. <https://doi.org/10.1021/ja012498q>.

Smith, L. J.; van Gunsteren, W. F.; Hansen, N. On the Use of Side-Chain NMR Relaxation Data to Derive Structural and Dynamical Information on Proteins: A Case Study Using Hen Lysozyme. *ChemBioChem* **2021**, 22 (6), 1049–1064. <https://doi.org/10.1002/cbic.202000674>.

Spaans, S. K.; Weusthuis, R. A.; van der Oost, J.; Kengen, S. W. M. NADPH-Generating Systems in Bacteria and Archaea. *Front. Microbiol.* **2015**, 6:742. <https://doi.org/10.3389/fmicb.2015.00742>.

Sprangers, R.; Velyvis, A.; Kay, L. E. Solution NMR of Supramolecular Complexes: Providing New Insights into Function. *Nat. Methods* **2007**, 4 (9), 697–703. <https://doi.org/10.1038/nmeth1080>.

Stiffin, R. M.; Sullivan, S. M.; Carlson, G. M.; Holyoak, T. Differential Inhibition of Cytosolic PEPCCK by Substrate Analogues. Kinetic and Structural Characterization of Inhibitor Recognition. *Biochemistry* **2008**, 47 (7), 2099–2109. <https://doi.org/10.1021/bi7020662>.

Stülke, J.; Hillen, W. Carbon Catabolite Repression in Bacteria. *Curr. Opin. Microbiol.* **1999**, 2 (2), 195–201. [https://doi.org/10.1016/S1369-5274\(99\)80034-4](https://doi.org/10.1016/S1369-5274(99)80034-4).

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-Labeling and Production of Perdeuterated Proteins in H₂O. *J. Biomol. NMR* **2011**, 50 (1), 35–42. <https://doi.org/10.1007/s10858-011-9477-5>.

Takeda, M.; Ono, A. M.; Terauchi, T.; Kainosho, M. Application of SAIL Phenylalanine and Tyrosine with Alternative Isotope-Labeling Patterns for Protein Structure Determination. *J. Biomol. NMR* **2010**, 46 (1), 45–49. <https://doi.org/10.1007/s10858-009-9360-9>.

Tanamoto, K.; Azumi, S. Salmonella -Type Heptaacylated Lipid A Is Inactive and Acts as an Antagonist of Lipopolysaccharide Action on Human Line Cells . *J. Immunol.* **2000**, 164 (6), 3149–3156. <https://doi.org/10.4049/jimmunol.164.6.3149>.

Thorn, M. B. Inhibition by Malonate of Succinic Dehydrogenase in Heart-Muscle Preparations. *Biochem. J.* **1953**, 54 (4), 540–547. <https://doi.org/10.1042/bj0540540>.

Tishkov, V. I.; Popov, V. O. Protein Engineering of Formate Dehydrogenase. *Biomol. Eng.* **2006**, 23 (2–3), 89–110. <https://doi.org/10.1016/j.bioeng.2006.02.003>.

Tong, M.; French, S.; El Zahed, S. S.; Ong, W. K.; Karp, P. D.; Brown, E. D. Gene Dispensability in Escherichia Coli Grown in Thirty Different Carbon Environments. *MBio* **2020**, 11 (5), 1–20. <https://doi.org/10.1128/mBio.02259-20>.

Torizawa, T.; Ono, A. M.; Terauchi, T.; Kainosho, M. NMR Assignment Methods for the Aromatic Ring Resonances of Phenylalanine and Tyrosine Residues in Proteins. *J. Am. Chem. Soc.* **2005**, 127 (36), 12620–12626. <https://doi.org/10.1021/ja051386m>.

Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, 1 (2), 749–754. <https://doi.org/10.1038/nprot.2006.101>.

Tugarinov, V.; Kay, L. E. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *J. Biomol. NMR* **2004**, 28 (2), 165–172. <https://doi.org/10.1023/B:JNMR.0000013824.93994.1f>.

Tugarinov, V.; Kay, L. E. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **2003**, 125 (45), 13868–13878. <https://doi.org/10.1021/ja030345s>.

Tugarinov, V.; Kay, L. E. Methyl Groups as Probes of Structure and Dynamics in NMR Studies of High-Molecular-Weight Proteins. *ChemBioChem* **2005**, 6 (9), 1567–1577. <https://doi.org/10.1002/cbic.200500110>.

Tuttle, L. M.; Dyson, H. J.; Wright, P. E. Side-Chain Conformational Heterogeneity of Intermediates in the Escherichia Coli Dihydrofolate Reductase Catalytic Cycle. *Biochemistry* **2013**, 52 (20), 3464–3477. <https://doi.org/10.1021/bi400322e>.

Valls-Lacalle, L.; Barba, I.; Miró-Casas, E.; Ruiz-Meana, M.; Rodríguez-Sinovas, A.; García-Dorado, D. Selective Inhibition of Succinate Dehydrogenase in Reperfused Myocardium with Intracoronary Malonate Reduces Infarct Size. *Sci. Rep.* **2018**, 8 (1), 1–10. <https://doi.org/10.1038/s41598-018-20866-4>.

Van Der Donk, W. A.; Zhao, H. Recent Developments in Pyridine Nucleotide Regeneration. *Curr. Opin. Biotechnol.* **2003**, 14 (4), 421–426. [https://doi.org/10.1016/S0958-1669\(03\)00094-6](https://doi.org/10.1016/S0958-1669(03)00094-6).

Velyvis, A.; Ruschak, A. M.; Kay, L. E. An Economical Method for Production of ^2H , $^{13}\text{CH}_3$ -Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 KDa Proteasome. *PLoS One* **2012**, 7 (9), 1–8. <https://doi.org/10.1371/journal.pone.0043725>.

Verdecia, M. A.; Bowman, M. E.; Ping, K.; Hunter, T.; Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **2000**, 7 (8), 639–643. <https://doi.org/10.1038/77929>.

Vuister, G. W.; Kim, S. J.; Wu, C.; Bax, A. 2D and 3D NMR Study of Phenylalanine Residues in Proteins by Reverse Isotopic Labeling. *J. Am. Chem. Soc.* **1994**, 116 (20), 9206–9210. <https://doi.org/10.1021/ja00099a041>.

Wagner, G. Prospects for NMR of Large Proteins. *J. Biomol. NMR* **1993**, 3 (4), 375–385. <https://doi.org/10.1007/BF00176005>.

Wagner, S.; Klepsch, M. M.; Schlegel, S.; Appel, A.; Draheim, R.; Tarry, M.; Wijk, K. J. Van; Slotboom, D. J.; Persson, J. O.; Gier, J. De. Tuning Escherichia Coli for Membrane. *Proc. Natl. Acad. Sci. U. S. A* **2008**, 105 (38), 14371–14376. <https://doi.org/10.1073/pnas.0804090105>.

Wand, A. J.; Urbauer, J. L.; McEvoy, R. P.; Bieber, R. J. Internal Dynamics of Human Ubiquitin Revealed by ^{13}C -Relaxation Studies of Randomly Fractionally Labeled Protein. *Tech. Protein Chem.* **1997**, 8 (C), 715–725. [https://doi.org/10.1016/S1080-8914\(97\)80070-1](https://doi.org/10.1016/S1080-8914(97)80070-1).

Weininger, U. Site-Selective ^{13}C Labeling of Histidine and Tryptophan Using Ribose. *J. Biomol. NMR* **2017**, 69 (1), 23–30. <https://doi.org/10.1007/s10858-017-0130-9>.

Weininger, U.; Liu, Z.; McIntyre, D. D.; Vogel, H. J.; Akke, M. Specific $^{12}\text{C}\beta\text{D}_2$ $^{12}\text{C}\gamma\text{D}_2\text{S}^{13}\text{C}\epsilon\text{HD}_2$ Isotopomer Labeling of Methionine To Characterize Protein Dynamics by

¹H and ¹³C NMR Relaxation Dispersion. *J. Am. Chem. Soc.* **2012**, 134, 18562–18565. <https://doi.org/10.1021/ja309294u>.

Wider, G. NMR Techniques Used with Very Large Biological Macromolecules in Solution. *Methods Enzymol.* **2005**, 394, 382–398. [https://doi.org/10.1016/S0076-6879\(05\)94015-9](https://doi.org/10.1016/S0076-6879(05)94015-9).

Wiesner, S.; Sprangers, R. Methyl Groups as NMR Probes for Biomolecular Interactions. *Curr. Opin. Struct. Biol.* **2015**, 35, 60–67. <https://doi.org/10.1016/j.sbi.2015.08.010>.

Wintjens, R.; Wieruszkeski, J. M.; Drobecq, H.; Rousselot-Pailley, P.; Buée, L.; Lippens, G.; Landrieu, I. ¹H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* **2001**, 276 (27), 25150–25156. <https://doi.org/10.1074/jbc.M010327200>.

Yang, D. Probing Protein Side Chain Dynamics Via ¹³C NMR Relaxation. *Protein Pept. Lett.* **2011**, 18 (4), 380–395. <https://doi.org/10.2174/092986611794653932>.

Yang, D.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. A Study of Protein Side-Chain Dynamics from New 2H Auto-Correlation and 13C Cross-Correlation NMR Experiments: Application to the N-Terminal SH3 Domain from Drk. *J. Mol. Biol.* **1998**, 276 (5), 939–954. <https://doi.org/10.1006/jmbi.1997.1588>.

Yang, Z.; Floyd, D. L.; Loeber, G.; Tong, L. Structure of a Closed Form of Human Malic Enzyme and Implications for Catalytic Mechanism. *Nat. Struct. Biol.* **2000**, 7 (3), 251–257. <https://doi.org/10.1038/73378>.

You, C.; Okano, H.; Hui, S.; Zhang, Z.; Kim, M.; Gunderson, C. W.; Wang, Y. P.; Lenz, P.; Yan, D.; Hwa, T. Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling. *Nature* **2013**, 500 (7462), 301–306. <https://doi.org/10.1038/nature12446>.

Yu, H. Extending the Size Limit of Protein Nuclear Magnetic Resonance. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96 (2), 332–334. <https://doi.org/10.1073/pnas.96.2.332>.

Zahran, S.; Pan, J. S.; Liu, P. B.; Hwang, P. M. Combining a PagP Fusion Protein System with Nickel Ion-Catalyzed Cleavage to Produce Intrinsically Disordered Proteins in E. Coli. *Protein Expr. Purif.* **2015**, 116, 133–138. <https://doi.org/10.1016/j.pep.2015.08.018>.

Zeczycki, T. N.; St. Maurice, M.; Attwood, P. V. Inhibitors of Pyruvate Carboxylase. *Open Enzym. Inhib. J.* **2010**, 3 (1), 8–26. <https://doi.org/10.2174/1874940201003010008>.

Zhang, R. B.; Somers, K. R. F.; Kryachko, E. S.; Nguyen, M. T.; Zeegers-Huyskens, T.; Ceulemans, A. Hydrogen Bonding to π -Systems of Indole and 1-Methylindole: Is There Any OH \cdots Phenyl Bond? *J. Phys. Chem. A* **2005**, 109 (35), 8028–8034. <https://doi.org/10.1021/jp0525437>.

Zhang, X.; Sui, X.; Yang, D. Probing Methyl Dynamics from ¹³C Autocorrelated and Cross-Correlated Relaxation. *J. Am. Chem. Soc.* **2006**, 128 (15), 5073–5081. <https://doi.org/10.1021/ja057579r>.

Zhang, Z.; Kuipers, G.; Niemiec, Ł.; Baumgarten, T.; Slotboom, D. J.; De Gier, J.-W.; Hjelm, A. High-Level Production of Membrane Proteins in E. Coli BL21(DE3) by Omitting the Inducer IPTG. *Microb Cell Fact* **2015**, 14, 142. <https://doi.org/10.1186/s12934-015-0328-z>.

Zheng, Y.; Yang, D. Measurement of Dipolar Cross-Correlation in Methylene Groups in Uniformly ¹³C-, ¹⁵N-Labeled Proteins. *J. Biomol. NMR* **2004**, 28 (2), 103–116. <https://doi.org/10.1023/B:JNMR.0000013826.82936.0e>.