The Structure of Micelle-bound Proamylin and Advances in Nuclear Overhauser Effect Spectroscopy

By

Charles DeLisle
B.S., Grand Valley State University, 2012
M.S., University of Illinois at Chicago, 2014

Submitted as partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Chemistry
University of Illinois at Chicago
Chicago, IL, USA
Spring 2020

Defense Committee:

Justin L. Lorieau, Chair and Advisor, Chemistry
Stephanie M. Cologna, Chemistry
Leslie Wo-Mei Fung, Chemistry
Tim Keiderling, Chemistry Emeritus
Vadim Gaponenko, Biochemistry & Molecular Genetics

This work is wholly dedicated to my family and friends that have tirelessly supported me throughout the entirety of my graduate school career. I could not have done this without you.

Acknowledgements

First, I would like to thank my advisor Professor Justin L. Lorieau for all he has invested in me over the past five years. I joined his group with almost no prior knowledge of NMR, and now I leave a confident biomolecular NMR spectroscopist. He has been an exceptional leader, mentor, and motivator, and I was fortunate to have the opportunity to work with him. I would also like to thank the other members of my committee, Professor Stephanie Cologna, Professor Leslie Wo-Mei Fung, Professor Emeritus Tim Keiderling, and Professor Vadim Gaponenko for challenging me and offering me excellent insight and advice.

I would like to thank Dr. Dan McElheny, for the countless times he has gone beyond what is asked of him to help our lab group succeed. This institution is fortunate to have such a compassionate expert running the NMR research facilities.

A very special thanks to Dr. Pallavi Thiagarajan-Rosenkranz and Dr. Adrian Draney. Without their help and support early on, I wouldn't have completed the Ph.D. I would also like to thank all my current and former colleagues for the critical discussions that helped us to grow together: Sean, Bhagya, Indrani, Alec, Ed, Zoe, Medine, Hyunjin, Ewa, Liz, Svetlana, Ren, Yao, Yusi, and Shu-Lin.

To the original and honorary members of the Loomis Street boys; Adrian, Greyson, Tom, Drew, Noel, and Chris. Thank you for going through this with me and keeping me sane.

To my parents, Fred and JoAnn. Thank you for the years of guidance to shape me into the man I am today. Thank you for pushing me to achieve but providing a safety net for when I failed. I am forever grateful for the opportunities that you provided me.

To my local family, Mama and Dusty. Words cannot describe the role you played in this achievement. When I was alone for the holidays, you gave me a Christmas. When I was frustrated and shortsighted, you gave me new perspectives. Thank you for always being there. You will always be my family.

To my best friend and companion, Michael Farina. Thank you for the countless small things that you do to make my life easier. You have been my sounding board and my biggest supporter. Above all, thank you for your patience with me through these years.

Finally, to my daughter, Louise. Watching you grow up has been one of the most rewarding experiences in my life. You motivate me to be better man and remind me to appreciate the small things. Thank you for keeping my spirits high, even when things have been ruff.

Authors and Contributions

Chapter 1: Written exclusively for this dissertation.

Author contributions: Charles F. DeLisle wrote this chapter.

Chapter 2: Submitted to *The FEBS Journal* on October 12th, 2019, as "Micelle-bound pro-Islet Amyloid Polypeptide Contains a Carboxy-terminal Helical Segment", Charles F. DeLisle, Alexander L. Malooley, Indrani Banerjee, and Justin L. Lorieau, The FEBS Journal, 2019. *Under Revision*.

Author contributions: CFD and JLL planned the experiments. CFD and ALM prepared the protein samples. CFD, ALM, and IB collected and analyzed the data. CFD, ALM, IB, and JLL wrote the manuscript.

Chapter 3: Sections 3.1-3.6 were reproduced by permission from Springer Nature: "Super resolution NOESY spectra of proteins", Charles F. DeLisle, H. Bhagya Mendis, and Justin L. Lorieau, Journal of Biomolecular Nuclear Magnetic Resonance, 2019. 73(3): 105-116. Copyright © (2019) Springer Nature. Section 3.7 was written exclusively for this dissertation.

Author contributions: Portion reproduced from the Journal of Biomolecular Nuclear Magnetic Resonance, JLL designed the experimental research, CFD and HBM prepared the samples, CFD and HBM collected the data, CFD, HBM, and JLL analyzed the data, and CFD, HBM, and JLL wrote the manuscript. Section 3.7 was written by CFD.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

List of Tables

List of Figures

List of Figures (Continued)

List of Figures (Continued)

List of Abbreviations

aa, amino acid

ANN, artificial neural network

APBS, adaptive Poisson-Boltzmann solver

Arg, arginine

Asn, asparagine

AUC, analytical ultracentrifugation

BMRB, biological magnetic resonance bank

B_0, static magnetic field

BPP, bipolar pulse pair

CARA, computer aided resonance assignment

CD, circular dichroism

CHAPS, 3-cholamidopropyl dimethylammonio 1-propanesulfonate

CIS, complexation-induced changes in shift

CMC, critical micelle concentration

COSY, correlation spectroscopy

CPE, carboxypeptidase E

Cpro, C-terminal pro-islet amyloid polypeptide

cryo-EM, cryogenic electron microscopy

CSA, chemical shift anisotropy

CSI, chemical shift index

CSP, chemical shift perturbation

CYANA, combined assignment and dynamics algorithm for NMR applications

DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphocholine

DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine

DMPS, 1,2-dimyristoyl-sn-glycero-3-phosphoserine

DNA, deoxyribonucleic acid

DPC, dodecyl phosphocholine

E. coli, *Escherichia coli*

E. COSY, exclusive correlation spectroscopy

List of Abbreviations (Continued)

ESI-MS, electrospray ionization mass spectrometry

DOSY, diffusion ordered spectroscopy

FID, frequency induction decay

FPLC, fast-protein liquid chromatography

FT, Fourier transform

FWHH, full-width at half-height

FXa, Factor Xa

Gln, glutamine

GUI, graphical user interface

HAfp, hemagglutinin fusion peptide

HBDB, hydrogen bonding database

H-bond, hydrogen bond

hetNOE, heteronuclear nuclear Overhauser effect

His_6, hexa-histadine

HSQC, heteronuclear single quantum coherence

HX, hydrogen exchange

IAPP, islet amyloid polypeptide

IDP, intrinsically disordered protein

IDR, intrinsically disordered region

INEPT, insensitive nuclei enhanced by polarization transfer

IPAP-HSQC, in-phase anti-phase heteronuclear single quantum coherence

IPTG, isopropyl β-d-1-thiogalactopyranoside

IVM, internal variable module

K_4D, tetra-lysine aspartatic acid

LB, Luria-Bertani

LED, longitudinal eddy current delay

MALS, multi-angle light scattering

MC, monochromator

MD, molecular dynamics

List of Abbreviations (Continued)

MRE, mean residue elipticity

MW, molecular weight

MWCO, molecular weight cut-off

NOESY, nuclear Overhauser effect spectroscopy

NMR, nuclear magnetic resonance

Npro, N-terminal pro-islet amyloid polypeptide

NUS, non-uniform sampling

PAGE, polyacrylamide gel electrophoresis

PAM, peptidyl-glycine alpha-amidating monooxygenase

PAMP, pro-adrenomedullin N-terminal 20 peptide

PAS, principal axis system

PC, prohormone/proprotein convertase

PDB, protein data bank

PEM, photoelastic modulator

PFG, pulsed field gradient

PINE, probabilistic interaction network of evidence

ppm, parts per million

PRE, paramagnetic relaxation enhancement

proIAPP, pro-islet amyloid polypeptide

PSVS, protein structure validation software suite

QXI, quattro resonance inverse

RDC, residual dipolar coupling

rf, radio frequency

RMSD, root-mean-square deviation

RSP, regulated secretory pathway

SAG, stretched acrylamide gel

SDS, sodium dodecyl sulfate

SEC, size-exclusion chromatography

SMILE, sparse multidimensional iterative lineshape-enhanced

List of Abbreviations (Continued)

srNOESY, super resolution nuclear Overhauser effect spectroscopy

SVD, singular value decomposition

TOCSY, total correlation spectroscopy

TorsionDB, torsion angle database

TPPI, time-proportional phase incrementation

Tris, 2-Amino-2-(hydroxymethyl)-1,3-propanediol

TROSY, transverse relaxation-optimized spectroscopy

TXI, triple resonance inverse

Ubq, ubiquitin

UV, ultraviolet

VDW, van der Waals

WATERGATE, water suppression by gradient tailored excitation

Summary

The chief focus of my dissertation research was to apply existing solution NMR techniques to study the molecular structure and dynamics of intrinsically disordered amyloid-forming proteins and develop new experiments to expand upon the NMR toolbox. Additional biophysical methods were employed to complement the NMR results in order to thoroughly characterize the protein systems of interest. Chapter 1 aims to introduce the theoretical framework behind the science, as well as to present practical examples of the workflow involved in structure determination by solution-state NMR spectroscopy.

The works presented in Chapters 2 and 3 are divided into two major areas: applications and methods development. Chapter 2 describes the use of NMR and other biophysical methods in the thorough structural characterization of pro-islet amyloid polypeptide in a membrane environment. We map the protein-lipid complex interface, and examine the effects of membrane curvature, charge, and pH on complex formation. Surprisingly, we found that the C-terminal pro-hormone segment of this protein can fold and associate to membranes independently of the rest of the molecule, and comment on potential implications.

Chapter 3 describes the development of the super-resolution NOESY experiment to improve data quality for protein structure determination. We use two model systems, ubiquitin and the hemagglunitin fusion peptide of the influenza A virus (HAfp) bound to large bicelles, in order to characterize the optimal experimental parameters with respect to molecular size for maximum resolution enhancement. Furthermore, we explain the theory, pitfalls, and contrast to apodization techniques.

Chapter 1

Experimental Techniques

1.1 Introduction to Structural Biology by NMR

The functional role of biological macromolecules is encompassed by their structure and conformational dynamics. Nuclear magnetic resonance (NMR) provides an unparalleled means to investigate both features simultaneously. Structures can be solved at atomic resolution, and dynamics can be probed on a timescale from picoseconds to years. NMR allows for study at multiple near-physiological temperatures and environments providing information with the potential to relate the findings directly to functional relevance.

One of the true strengths of NMR lies in the study of dynamic molecules with sparse populations of transiently structured states, i.e. intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions (IDRs). These are flexible and diverse polypeptide chains lacking a well-defined structure. Their conformational heterogeneity affords a large number of interacting partners, and yet precludes any detailed structural analysis by X-ray crystallography or cryo-electron microscopy (cryo-EM). An estimated 25% of the human proteome contains an intrinsically disordered region greater than 50 amino-acid (aa) residues in length, and 44% contains an IDR greater than 30 aa in length [1], [2]. Known IDPs include transcription factors, ribosomal proteins, hormones, and amyloid-forming proteins, but many IDP functions have yet to be determined. NMR provides a promising platform capable of

interrogating the fine structural and dynamic information from this important class of biomolecules.

NMR is not without disadvantages, specifically when it comes to molecular size. X-ray crystallography and cryo-EM have both been shown to be superior methods for the structure determination of very large molecules and molecular complexes. However, the sheer multitude of new strategies that have been developed to mitigate the size problem in the past three decades implies that NMR has not yet reached its limit.

The focus of this thesis has been on the study of proteins and protein-lipid complexes using solution-state NMR spectroscopy. The work contained within uses NMR to solve the dynamic structural ensemble of pro-Islet Amyloid Polypeptide (proIAPP), an IDP protein hormone, in a membrane environment and to probe the effects of pH, membrane charge, and membrane curvature using complementary biophysical techniques. Additionally, I present an improvement to the nuclear Overhauser effect spectroscopy (NOESY) class of NMR structural biology experiments. Though many exemplary textbooks exist on these topics, this chapter attempts to briefly introduce the relevant theory behind the research carried out in this thesis, with a particular focus on the modern practical aspects of the work.

1.1.1 Spin and Nuclear Magnetism

In addition to mass and charge, atomic nuclei also possess the property of spin. The behavior of nuclei to precess like a spinning charge gives rise to spin angular momentum ($\vec{L}$), the product of the radius of procession ($\vec{r}$) and the linear momentum, which is the product of velocity and mass ($\vec{v}m$) [3].

$$\vec{L} = \vec{r} \times \vec{v}m \tag{1}$$

In three dimensional space, the magnitude ($|\vec{L}|$) is given by the square-root of the sum of squares of the three directional components.

$$\left|\vec{L}\right| = \sqrt{L_x^2 + L_y^2 + L_z^2} \tag{2}$$

It is only possible to define $|\vec{L}|$ and one component (L_z) at any given time, due to the quantum nature of atoms.

$$\left|\vec{L}\right| = \hbar\sqrt{l(l+1)} \tag{3}$$

$$L_z = \hbar m_l \tag{4}$$

The angular momentum quantum number of the nucleus is given by l, the magnetic quantum number of the nucleus is m_l, and $\hbar$ is the reduced Planck's constant (h/2π). l adopts values of integers and half-integers (i.e. $\frac{1}{2}$, 1, $\frac{3}{2}$, …), while m_l adopts integer values ranging from -l to l with a total number of energy levels according to equation (5).

$$N_l = 2l + 1 \tag{5}$$

For example, a nucleus with $l = \frac{3}{2}$, has 4 energy levels $m_l = -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}$, and $\frac{3}{2}$.

These spin states are at equivalent energy levels in the absence of an external magnetic field (B_0). However once B_0 is applied, each spin state adopts a discrete, equally spaced energy level. This modulation of energy levels is known as the nuclear Zeeman splitting, and it is attributed to the interaction of the spin angular momentum of a magnetic dipole with an external magnetic field. Nuclear spins align in opposite directions within these energy levels in nearly equal populations. Most of the spins have net magnetic moments that cancel one another out, but a very small number will remain unpaired to create the observable bulk magnetic dipole moment. The population difference exhibits a temperature dependence through the Boltzmann distribution.

In classical electromagnetic theory of electric coils, the magnetic dipole moment ($\vec{\mu}$) is equivalent to the electrical current (I) over the size of a solenoid.

$$\vec{\mu} = I * A * N_{turns} \tag{6}$$

The size of the solenoid is determined by the area (A) and number of turns (N_{turns}). In the context of atomic magnetic moments, the magnetic susceptibility of a nucleus is proportional its gyromagnetic ratio (γ).

$$\gamma = {\vec{\mu}}/{\vec{L}} \tag{7}$$

At equilibrium, the dipole will precess about B_0 at the Larmor frequency (ω_0) and the direction of precession is encoded in γ.

$$\omega_0 = -\gamma B_0 \tag{8}$$

The energy associated with the dipole at equilibrium (E) is defined by the magnitude and direction of $\vec{\mu}$ and $\vec{B}$. Consider a ^{1}H nucleus (spin) in an 11.7 T at equilibrium, comprised of only z-axis magnetization.

$$E = -\vec{\mu} \cdot \vec{B} \tag{9}$$

Rearrangement and substitution of equation 9 with equation 7 yields equation 10.

$$E = -\gamma L_z B_0 \tag{10}$$

Further substitution with equation 4 gives equation 11.

$$E = -\gamma m \hbar B_0 \tag{11}$$

A ^{1}H atom has a $m_l = -\frac{1}{2}, \frac{1}{2}$, so the difference in Zeeman effect energy levels can be expressed as

$$\Delta E = -\gamma \left(-\frac{1}{2}\right) \hbar B_0 - -\gamma \left(\frac{1}{2}\right) \hbar B_0$$

$$\Delta E = -\gamma \hbar B_0 \tag{12}$$

Substitution of the relationship between frequency (υ) and energy, $\Delta E = h\upsilon$, gives the Larmor frequency of ^{1}H, expressed in units of MHz.

$$\omega_0 = -\frac{\gamma B_0}{2\pi} \tag{13}$$

γ_{1H} is $267.52 \cdot 10^6\ rad\ s^{-1}T^{-1}$, which yields a Larmor frequency of 500 MHz in a 11.7 T magnet.

1.1.2 Magnetically Active Isotopes

Not all nuclei produce a bulk magnetic moment within a B_0 field. This occurs when *l* is 0, thus there is only 1 Zeeman energy level. Generally, the angular momentum quantum number can be determined by the atomic number and mass number of a given nucleus. If each value is an even number, typically *l* = 0 (i.e. ^{12}C). If the mass number is even and the atomic number is odd, typically *l* is an integer (i.e. ^{14}N, *l* = 1). If the mass number is odd, an even or an odd atomic number typically results in half-integer values of *l* (^{13}C, *l* = ½). Different nuclei precess at different frequencies within the same magnetic field, given their different charge-to-mass ratios. Their rate of precession is therefore encoded by γ. **Table 1.1** provides nuclei of biological significance and their isotopes, along with respective γ and relative natural abundancy [4].

Table 1.1. Magnetic Properties of Biologically Relevant Stable Nuclear Isotopes

Nucleus	I	γ ($10^6 \cdot rad \cdot s^{-1} \cdot T^{-1}$)	Natural Abundance (%)
^{1}H	1/2	267.53	99.980
^{2}H	1	41.06	0.016
^{12}C	0	N/A	98.892
^{13}C	1/2	67.28	1.108
^{14}N	1	19.34	99.630
^{15}N	1/2	-27.12	0.370
^{16}O	0	N/A	99.760
^{17}O	5/2	-36.28	0.040
^{19}F	1/2	251.79	100.000
^{31}P	1/2	108.41	100.000
^{32}S	0	N/A	95.020
^{33}S	3/2	20.55	0.760

1.1.3 The Chemical Shift

Application of an external B_0 field gives rise to secondary, induced magnetic fields at nuclei. An applied B_0 field induces a current in a molecule's electrons, generating an opposing local magnetic field at a nucleus. Local environment and chemical properties like bonding partners, bond angles, and bond lengths will modulate the local electron density around a nucleus. This phenomenon is known as chemical shielding, where the induced field from electrons shields the true field at a nucleus. Equation $E = -\vec{\mu} \cdot \vec{B}$ (9 can be modified to incorporate the effect of chemical shielding by

$$E = -\vec{\mu}(1 - \ddot{\sigma})\vec{B}$$

$$= \vec{\mu}\vec{B} + \vec{\mu}\ddot{\sigma}\vec{B} \quad (14)$$

where $\ddot{\sigma}$ represents the nuclear shielding tensor, which is orientationally dependent (anisotropic). This orientational dependence is described by a tensor, written in the principal axis system (PAS) as

$$\ddot{\sigma}_{PAS} = \begin{pmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{yy} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix} \quad (15)$$

where the x, y, and z directions are with respect to the local geometry of the nucleus. In the solution-state, isotropic tumbling of the nucleus averages this orientation dependence to its isotropic average.

$$\sigma_{avg} = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3} \quad (16)$$

Differences in chemical environment give rise to differences in σ, which in turn produce variations in the chemical shift (δ) for the same type of nucleus, providing the basis for all NMR experimentation.

$$\delta = \frac{\nu_{sample} - \nu_{ref}}{\omega_0} \cdot 10^6 = \left(\sigma_{ref} - \sigma_{sample}\right) \cdot 10^6 \quad (17)$$

The Larmor frequency is indicated by ω_0. This is typically expressed in parts per million (ppm) to standardize resonances at different B_0 fields. The chemical shift is an extremely sensitive measurement which affords unique insight into the local electronic environment of the nucleus.

1.1.4 Chemical Exchange

In the solution-state, the isotropic chemical shift represents a population weighted average of all conformations. A two state example is provided for populations *A* and *B* of the same nucleus.

$$\delta_{iso} = p_A \delta_A + p_B \delta_B \quad (18)$$

Chemical exchange kinetics are determined by the rate of exchange between populations and the frequencies of the spin populations. If chemical exchange rates between population *A* and *B* of the same spin are faster or slower than the difference between the spin frequencies, these are referred to fast or slow exchange kinetics, respectively [5]. Exchange kinetics on the order of the difference between the two frequencies are called intermediate exchange, and these often result in a loss of signal due to line broadening. In the slow exchange regime, a distinct chemical shift is observed for both *A* and *B*, while a single peak with the population weighted average chemical shift is observed for fast exchange [6] (Figure 1.1).

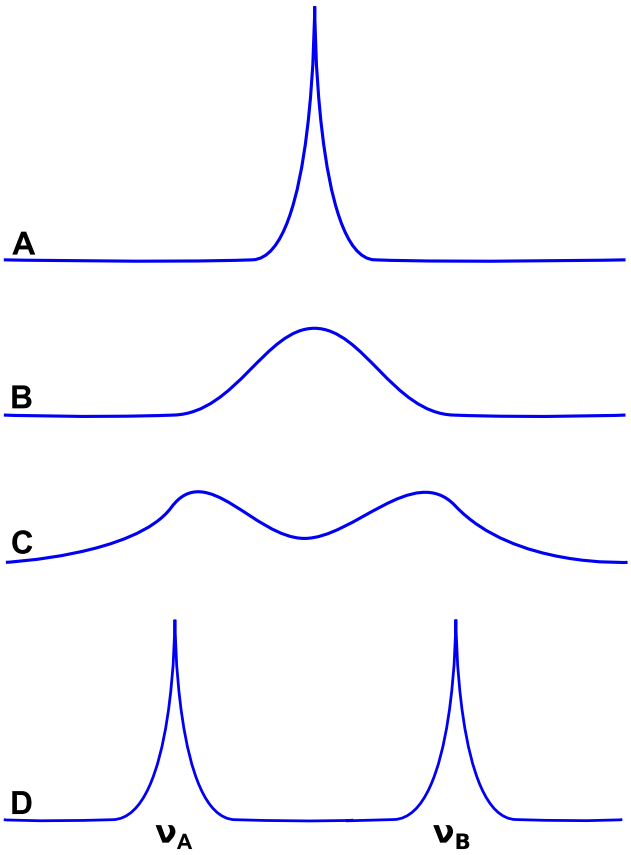


Figure 1.1. Manifestations of chemical exchange. In fast exchange **(A)**, one peak is observed at the weighted population-averaged frequency of the species. This peak broadens and eventually splits into two peaks during intermediate exchange **(B-C)**. In slow exchange **(D)**, two discrete peaks are observed for each species.

1.1.5 Hydrogen Exchange

All biological activity occurs in an aqueous environment. For this reason, organic solvents that would denature protein systems are not used in solution NMR. Most proteins are measured in solvent systems comprised of $^{1}H_2O$ supplemented with at least 5% $^{2}H_2O$ for spectrometer locking. An immediate challenge with using mostly $^{1}H_2O$ is the abundance of ^{1}H signal from the solvent, which can dominate the NMR signal. A number of water suppression pulse program sequences (e.g. Spin-echo, WATERGATE, etc.) have been created to nullify the strong water resonance [7], [8]. Alternatively, a solvent of purely $^{2}H_2O$ can be used to eliminate the water signal. However this strategy comes at the sacrifice of labile ^{1}H resonances, such as those bound to nitrogen, oxygen, or sulfur atoms. ^{1}H frequencies bound to carbons are non-labile and can still be detected.

Amide hydrogen exchange (HX) is a chemical exchange event that occurs when backbone amide ^{1}H atoms exchange with bulk water in the solvent. This results in multiple populations of a given amide resonance that average into a single broad peak at noise level. The rate of amide hydrogen exchange has been shown to have a direct correlation with temperature, pH, and local secondary structure [9], [10]. Of those factors, pH has the most dramatic effect on the exchange rate (**Figure 1.2**). The optimal NMR pH for hydrogen exchange is ~ 3, however the physiological significance of a protein structure solved in acid should be carefully considered. Additionally, there is a concern for hydrolysis of the protein amide backbone in strongly acidic or alkaline solutions. For a stable, well-folded protein, amide resonances can be measured up to a pH of ~ 8 depending on the system [11]. Unfolded proteins, or proteins with regions of many

flexible loops, are going to have a much greater number of missing amide resonances, making the assignment process much more tedious.

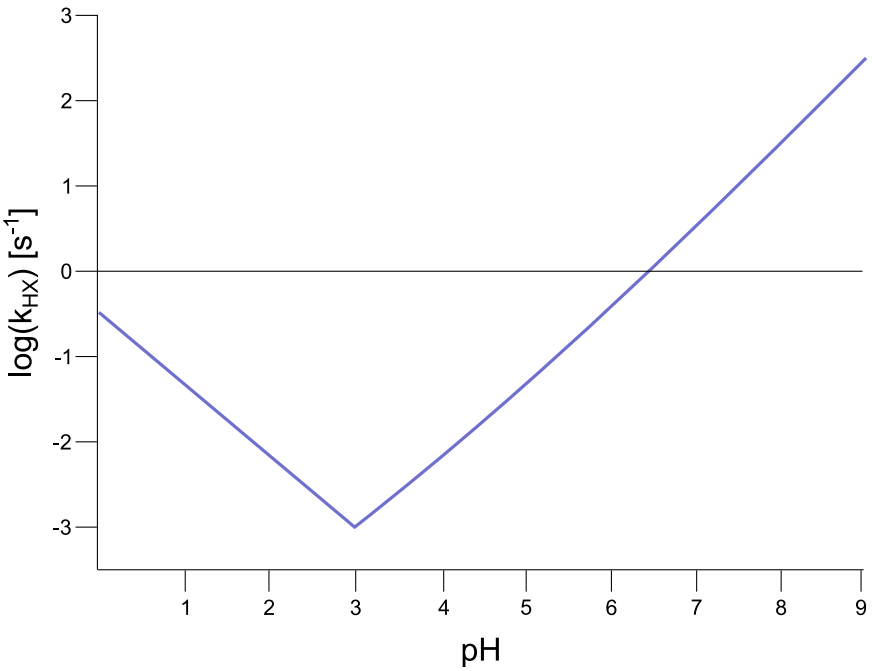


Figure 1.2. Amide hydrogen exchange rate as a function of pH. As pH increases, HX decreases until minimizing at pH 3. As pH increases beyond 3, HX increases. Above pH 7, the HX rate becomes faster than the timescale of the NMR measurement [12].

1.2 Protein Assignment

1.2.1 Multidimensional NMR

When studying molecules like peptides and proteins, often very little information can be obtained from a 1D experiment. This is mainly due to signal overlap, which prevents any sort of detailed spectral analysis. Multidimensional NMR can be utilized to increase resolution by spreading the NMR signal over a 2D surface or within a higher order dimensional space (3D, 4D, etc.). The resulting spectra offer considerably easier resonance assignment and analysis.

Two distinct types of dipole interactions dominate the NMR signal. Direct spin-spin coupling between non-zero spin nuclei, known as the dipolar (*D*) coupling, and indirect (*J*) coupling manifested from electrons shared in the bonds between nuclei. *D*-coupling is isotropically averaged to zero in the solution state, but this interaction remains a source of spin-spin relaxation. Similarly the orientation dependence of the *J*-coupling is averaged in the solution state, which reduces the tensor to a scalar quantity. The *J*-coupling can span multiple bonds, typically observable up to four bonds with magnitude inversely proportional to the bond number. Multidimensional NMR takes advantage of the constant value *J*-couplings in order to transfer magnetization between nuclei.

Consider a simple 2D homonuclear NMR experiment, the correlation spectroscopy (COSY) pulse sequence [13]. The pulse program consists of a relaxation delay period (d_1) for spin equilibration, a 90° hard pulse (p_1) and a incremented indirect dimension evolution period (t_1), another 90° hard pulse (p_2) to transfer magnetization to *J*-coupled spins, and a period (t_2) of frequency induction decay (FID) detection (**Figure 1.3**). Spins with connectivity up to $^3J_{HH}$ can usually be detected in this way. The 2D projection reads out on the magnitude of *J* in Hz as well as the number of spin active (in this case ^{1}H) neighbors encoded in the multiplicity of the signal.

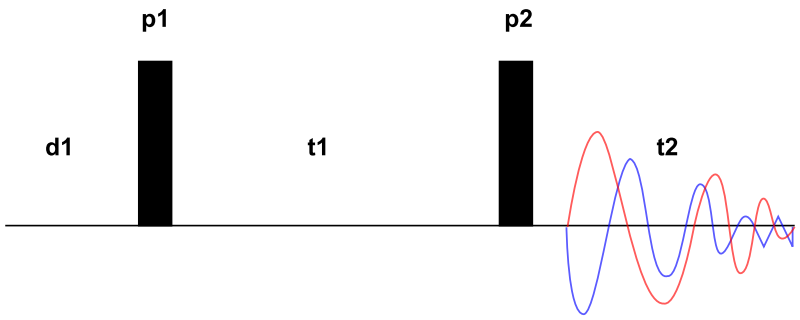


Figure 1.3. A simple 2D ^{1}H COSY experiment pulse program. The first pulse (p1) tips magnetization 90° into the transverse plane, where the chemical shift can evolve over an incremented time period (t1) for the indirect dimension. A second 90° pulse (p2) alternates the phase for the signal acquisition of the directly detected ^{1}H nucleus over a fixed time period (t2). The recycle delay, which allows for magnetization to return to the z-plane by relaxation, is given by d1.

The *J*-coupling also provides an efficient means to transfer magnetization between heteroatoms. Consider the classic Insensitive Nuclei Enhanced Polarization Transfer (INEPT) pulse sequence [14]. This experiment utilizes the known scalar coupling frequency between ^{1}H and an insensitive nucleus (X) to transfer magnetization, where X is ^{13}C or ^{15}N. The delays between pulses in the experiment are set to $1/4\ ^{1}J_{HX}$, which effectively transfer all of the transverse magnetization to the *J*-coupled X nucleus while placing all ^{1}H magnetization back on the z-axis. Phase cycling of the final X nucleus pulse results in a positive and negative peak for each X spin, separated by J_{HX}. The experiment has proven to be invaluable to the field, and the elements of the basic INEPT block are commonly found in many multidimensional heteronuclear NMR experiments.

1.2.2 Isotopic Labeling Strategies

With increasing protein size, homonuclear ^{1}H NMR spectra quickly become crowded with overlapping peaks that make interpretation of the data prohibitively complicated. In order to combat the resolution problem, protein deuterium labelling strategies were developed in the late 1960's by Katz and Jardetzky [15]–[18]. This advancement allowed for detection of specific interactions between protein residues and exogenous ligands, and it showed for the first time that isotopically labeling through modification of algae growth media was possible. With the advent of heteronuclear multidimensional NMR spectroscopy following almost three decades later, the field of protein NMR really began to evolve into a structural biology tool.

The first example of unambiguous protein backbone chemical shift assignment came from the group of Ad Bax in 1990 and involved the use of isotopically labeled amino acid types in bacterial growth media [19]. Like modern protein NMR, this approach requires prior knowledge of the primary sequence of amino-acids in the protein of interest. By taking

advantage of known dipeptide pairs in the sequence, one would be able to confidently assign specific magnetic resonances to those amino acid positions. Later that year, new experimental techniques were developed that would change the field dramatically and *Drosophila* calmodulin complexed with calcium became the first fully assigned, uniformly double isotope labeled protein [20]. The Bax group showed that *E. coli* could produce uniformly $^{13}C/^{15}N$-labeled proteins by supplementing $^{15}NH_4Cl$ and D-glucose-$^{13}C_6$ into M9 minimal growth media [21]. They had simultaneously introduced four new triple resonance three-dimensional experiments that allowed for the sequential assignment of the protein backbone. Both the sample preparation protocol and the experiments described in that report are commonly practiced today for proteins and protein complexes ranging up to ~30 kDa in size.

As the NMR structural biology community pushes forward to larger and more complicated protein systems, new isotopic labeling strategies have been developed to circumvent the challenges associated with size. One strategy attempted to expand the size constraints of homonuclear ^1H NMR methods and was conceived around the same time as the introduction of uniform double isotope labeling. This method, called random fractional deuteration, was developed by the group of David LeMaster in the late 1980's, and took advantage of the more favorable relaxation times of ^2H as compared to ^1H [22]. The concept utilized the random placement of ^2H throughout a protein that served to dilute ^1H-^1H dipolar interactions, which are chiefly responsible for the ^1H spectral line width in solution-state NMR. He was able to show that, with this method, proteins up to nearly 20 kDa could be assigned with traditional homonuclear ^1H NMR. Partial deuteration was then used in tandem with uniform double isotope labeling to tackle protein systems up to 40 kDa [23]–[25]. Since then there has been an explosion

of new NMR techniques in combination with selective labeling strategies to tackle unique challenges in protein systems all the way up to 200 kDa [26]–[30].

One of the challenges with isotopic labeling schemes is the production of milligram quantities of highly pure protein. Bacterial expression systems grown in minimal media tend to have markedly reduced protein yield compared to traditional growth media. A cost-effective substitution method has since been introduced to improve upon overall labeled protein yield [31] (**Figure 1.4**). The principle is to grow bacteria in 2-8 L of optimal, high yield natural abundance growth media, but harvest the cells right before protein expression is induced. Natural abundance expression of the protein of interest is inhibited by the T7 promoter of the plasmid. The cells are then resuspended in 1 L minimal growth media with isotopes, induced with isopropyl β-d-1-thiogalactopyranoside (IPTG), and allowed to express protein under normal conditions. The final overexpressed cell pellet contains protein with greater than 90% enriched isotope labeling efficiency, but often at a much larger quantity than normally obtained from 1 L of minimal media.

There is no single labeling scheme to satisfy all experimental goals. Requirements often depend heavily on the size and behavior of the protein of interest. More often it is a combinatorial approach of multiple labeling schemes required to effectively probe both protein structure and dynamics by NMR. Once the experiments and complimentary labeling schemes have been decided, one can begin the challenge of protein resonance assignment.

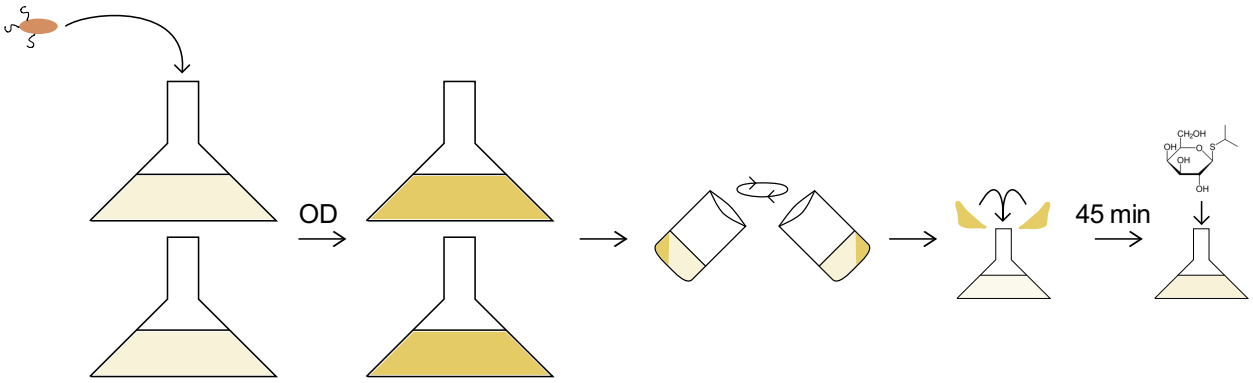


Figure 1.4. Diagram work-flow of a 4x natural abundance to 1L labeled growth media substitution for bacterial recombinant protein expression. Transformed *E. coli* cells are introduced to 4 L of sterile LB media and grown at 37 °C to OD_{600} ~0.7. Cells are harvested by centrifugation and resuspended in 1L isotope enriched minimal media. The cells are allowed to grow for 45 minutes at 37 °C to intake isotopes, before induction with IPTG and expression under normal conditions.

1.2.3 Protein Resonance Assignment

Nearly all NMR analyses require *a priori* knowledge of specific resonance assignment. Without this information, one would be unable to relate sequence specific structural and dynamics information back to the protein of interest. Backbone and side-chain resonance assignment can be somewhat labor intensive, and information on the sequence of amino-acids is critical for the assignment process. A variety of developments have been introduced to ease this process over the past two decades. In this subsection, I will describe the types of experiments for specific labeling strategies, the software tools that aid in the assignment process, and advances in the automation of this process.

The simplest means of resonance assignment can be performed with homonuclear NMR on natural abundance sample. Assignment can realistically be completed from just one sample by using two 2D experiments, the ^{1}H-^{1}H TOCSY and the ^{1}H-^{1}H NOESY. Total correlation spectroscopy (TOCSY) provides proton connectivities via spin-spin couplings [32]. In this experiment, magnetization is transferred between protons attached to a carbon atom, and isolated to individual residues by the backbone carbonyl. This allows for easy identification of the spin system, and thus the amino-acid type. In order to establish connectivity between residues, a ^{1}H-^{1}H NOESY is collected on the same sample and the spectra are overlaid (**Figure 1.5**). The NOESY spectrum will contain all of the same peaks from the TOCSY, in addition to neighboring resonances via dipolar relaxation mechanisms. The proximity of the neighboring resonances is encoded in the NOESY peak intensity and allows for the determination of neighboring residues (see **1.3.2** for more information). When all of this information is combined, a chemical shift map of the entire proton network can be produced for a peptide. Redundancy of the chemical shifts

for individual amino acids leads to ambiguous assignments due to resonance overlaps, therefore this simple approach to protein chemical shift assignment is restricted to only small proteins (< 10 kDa) and peptides.

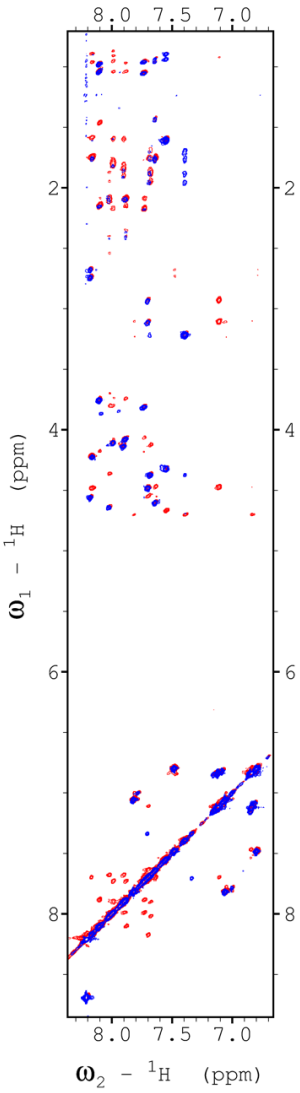


Figure 1.5. ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY spectral overlay of CproIAPP bound to DPC micelles. Overlapping TOCSY (blue) and NOESY (red) peaks at the same chemical shift on the x-axis belong to the same residue, while red peaks alone belong to neighboring residues. Spectra were collected on a 750 MHz Avance-III wide-bore spectrometer with a room temperature TXI probe. The TOCSY spectrum was collected with a 100 ms mixing time, and the NOESY spectrum was collected with a 300 ms mixing time. The sample contained 1.5 mM na-CproIAPP, 200 mM DPC, 30 mM sodium acetate, pH 4.5, 50 mM NaCl, 10% ^2H_2O, and 0.03% NaN_3.

The advent of isotopic enrichment and multidimensional heteronuclear NMR experiments has bolstered the use of NMR as a structural biology tool for medium-to-large protein systems. The additional ^{13}C and ^{15}N magnetically active isotopes provide additional resolution to separate the convoluted overlapping peaks in 2D homonuclear NMR. The most common approach to protein resonance assignment involves uniformly labeled $^{13}C/^{15}N$ protein, and a series of 3D experiments that connect the ^{1}H-^{15}N cross-peak of a ^{15}N-Heteronuclear Single Quantum Coherence (HSQC) spectrum to the intraresidue (i) ^{13}C or the preceding (i–1) ^{13}C resonance (**Figure 1.6**). The $^{13}C'$ resonance can be assigned through either an HNCO or HN(CA)CO, while $^{13}C^{\alpha}$ can be assigned by an HNCA or HN(CO)CA [20]. $^{13}C^{\beta}$ is then correlated to $^{13}C^{\alpha}$ by HN(CO)CACB or HNCACB experiments. When signal intensity is valued over resolution in the ^{13}C dimension, CBCA(CO)NH or CBCANH experiments can be used alternatively [33], [34]. These experiments provide the same correlations with fewer magnetization transfer steps, at the cost of a constant-time ^{13}C chemical shift evolution period with a maximum of ~6.6 ms (1/4 $^{1}J_{CC}$). Once $^{1}H^{N}$, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{15}N resonance have been assigned, the protein backbone can be unambiguously assigned to completion in most cases. The side-chain atoms can then be assigned by either a 3D or 4D HCCH-TOCSY for complete protein resonance assignment [35], [36]. Molecular size remains the pitfall of solution NMR, as these techniques only work on systems up to ~30 kDa.

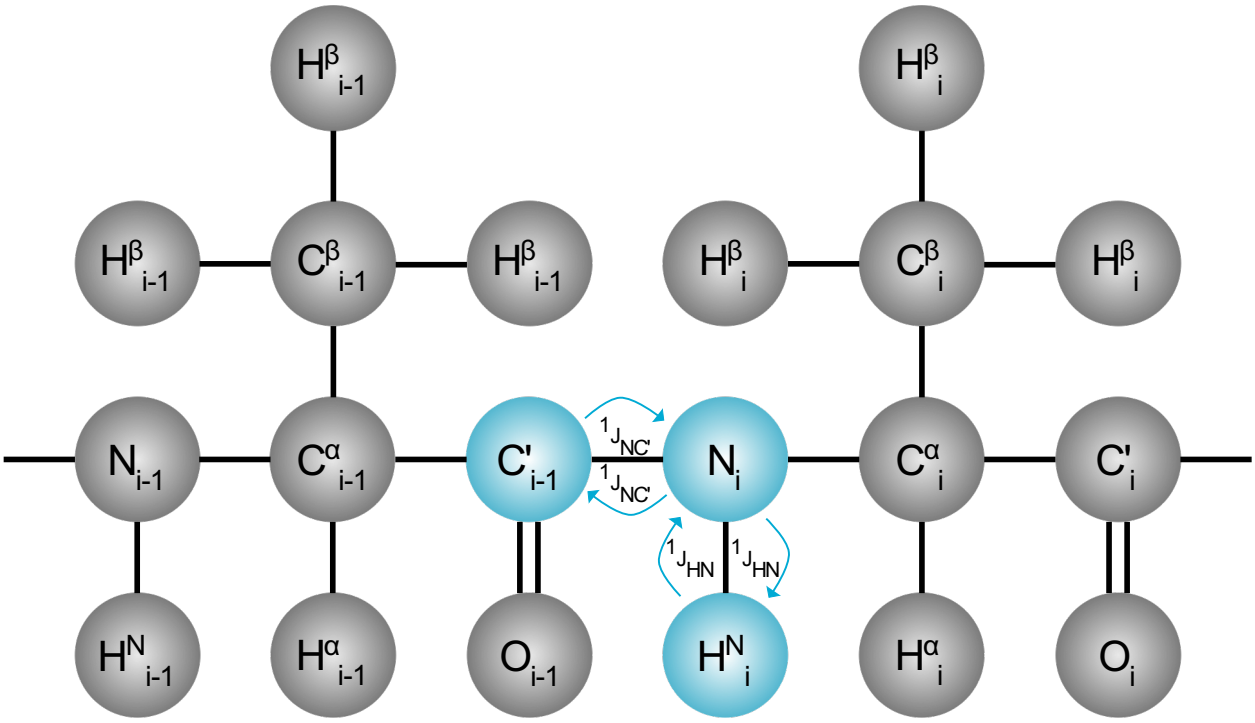


Figure 1.6. Experiment diagram of a 3D HNCO "out and back" triple resonance experiment. Magnetization is initially on H^N of the i^{th} residue, where it is transferred to the N of the i^{th} residue and then the C' of the preceding residue through 1-bond J-couplings. The chemical shift evolves on C' before magnetization is transferred back to N, where the chemical shift evolves once again before magnetization is transferred back to H^N.

For larger systems, perdeuteration triple isotope labeling ($^{2}H/^{13}C/^{15}N$) is used in tandem with transverse relaxation-optimized spectroscopy (TROSY) to produce sharp backbone resonances. TROSY is used to select the ^{1}H-^{15}N doublet peak with the longest T_2 relaxation time, as opposed to a conventional ^{15}N-HSQC which reports on the T_2 of the doublet [37]. Furthermore, selective deuteration strategies have been employed to produce only ^{1}H-^{13}C methyl groups, which allow for the detection of NOEs for structure determination [38]. This approach has allowed for detection of proteins and molecular complexes well above 50 kDa.

The development of software to ease in the assignment process of large biomolecules has been crucial to the use of NMR as a tool for structural biology. Processing of 1D and 2D spectra can often be performed within data acquisition software, but multidimensional experiments with greater than two dimensions require special handling. NMRpipe, developed by the Bax group in 1995, provides a UNIX based system for handling larger multidimensional NMR spectral processing [39]. With the advent of non-uniform sampling (NUS) a greater need for NMR data processing and spectral reconstruction has been required. NUS offers a significant reduction in the data collection time by breaking apart the indirect dimensions of a multidimensional experiment into non-uniformly sampled pieces. Putting those pieces back together correctly without loss of spectral information or addition of noise can be challenging. To overcome this, the Sparse Multidimensional Iterative Lineshape Enhanced (SMILE) NUS reconstruction plug-in was introduced into recent versions of the NMRpipe software [40]. Around the same time, a new GUI-based NMR data processor was introduced, NMRFx. NMRFx allows for processing of conventional and NUS NMR data while providing an interface conducive to beginners learning the basics of NMR processing [41]. Once the data has been processed, the assignment process can begin.

There are a few different biomolecular assignment software packages in use today. Two great free options are the Computer Aided Resonance Assignment (CARA) program from the Wütrich group and NMRFAM-SPARKY [42]–[44]. Cara provides a more rigid platform for the sequential assignment of biomolecules using established NMR experiments. The process is easy and intuitive, however only experiments found within CARA's database can be analyzed. Novel and unconventional experiments cannot be loaded into the spectra viewer without a template. Although it can be initially overwhelming to the NMR novice, NMRFAM-SPARKY offers a powerful alternative for NMR spectral analysis. The interface is very flexible, allowing for the analysis of less conventional experiments. The program also includes additional built-in software to analyze more than just resonance assignment, including relaxation dynamics and structure prediction.

These programs have changed the way NMR structural analysis is performed and made it significantly easier to assign protein systems. Once the protein resonances have been assigned, restraints can be defined for the generation of a structural model. The following section describes the process of generating experimental structural restraints from NMR data.

1.3 Structural Restraints

1.3.1 Angular Restraints

Martin Karplus was the first to introduce the idea that the angular geometry of a molecule was related to the 3-bond coupling constant between 2 protons separated by 3 bonds ($^3J_{HH}$) [45], [46]. In the Karplus equation, θ represents the dihedral angle associated with the bonds of which the J-coupling was calculated.

$$^3J_{HH} = A + Bcos(\theta) + C\cos^2(\theta) \quad (19)$$

With experimental advances like exclusive correlation spectroscopy (E. COSY) and quantitative J, determination of dihedral angle restraints for protein structures became commonplace [47], [48]. It was later shown that the Karplus equation coefficients are specific for the molecular parameters of the bonding nuclei. For example, measuring the $^{3}J_{HN,H\alpha}$ across the peptide backbone results in:

$$^{3}J_{HN,H\alpha} = 1.9 + 1.4cos(\theta) + 6.4\cos^{2}(\theta) \quad (20)$$

Evaluation across all of the peptide backbone J-couplings eventually led to the development of computer software, like SWEET J, to make dihedral angle determination even easier [49].

A strong dependence was established between backbone resonance chemical shifts and the protein secondary structure in the early 1990s [50]–[52]. This relationship lead to the development of the TALOS software, which made the determination of backbone dihedrals even easier than before [53]. The program made use of a database of 20 proteins with X-ray crystal structure resolution < 2.2 Å and fully assigned backbone ^{1}H, ^{13}C, and ^{15}N chemical shifts. Input experimental chemical shifts for $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N of an unknown protein were broken into tripeptide sequences and compared to the database proteins in order to generate a restraint window of possible ϕ and ψ dihedral angles. This software was further improved (TALOS+) with the incorporation of $^{1}H^{N}$ chemical shifts and the addition of an artificial neural network (ANN) component to better predict the likelihood of a secondary structural element of a given position, relative to adjacent neighboring residues [54]. The latest installment of this software (TALOS-N) goes a step further with the incorporation of a pentapeptide database mining approach, and the prediction of χ_1 side-chain dihedral angles [55].

A wealth of information is available using the TALOS family of software from just simple chemical shift assignment (**Figure 1.7**). A color-coordinated secondary structure

prediction histogram provides relative certainty encoded in the height of each bar from the ANN data mining, with red as helical and blue as sheet. A generalized order parameter for the backbone is also provided for each residue, correlated with the secondary structure analysis confidence. The software also assesses the likelihood of correct backbone assignment in a color mapped primary sequence figure, with green being likely, yellow being ambiguous, red being unlikely, and blue being dynamic or loop. Reference tripeptides with Ramachandran plots are provided as additional output. Overall, this software tool is invaluable in the determination of protein structure by solution NMR.

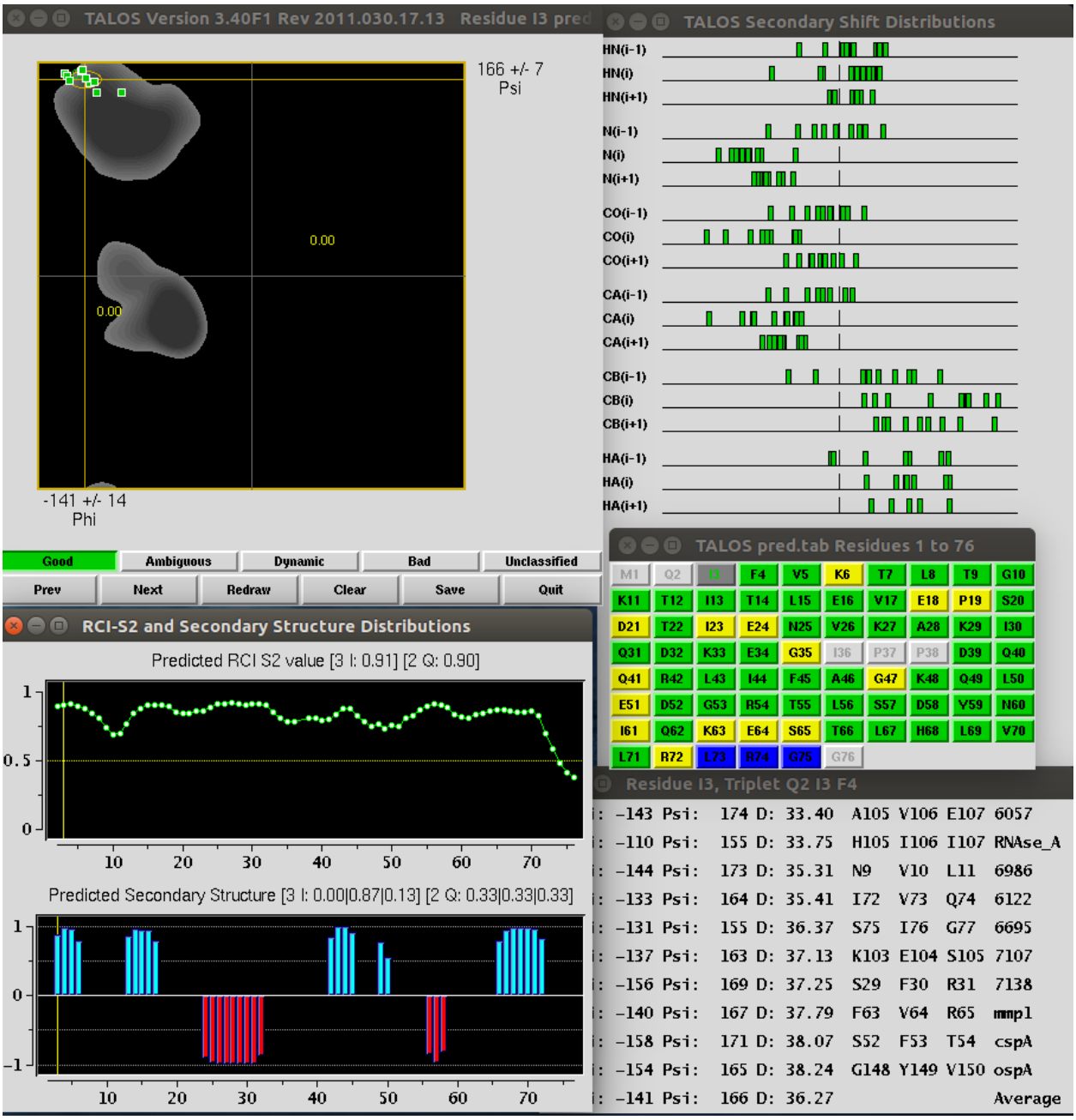


Figure 1.7. Example TALOS+ output for human $^{13}C,^{15}N$-His6-ubiquitin. The output was generated by entering a table of $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C'$, and ^{15}N chemical shifts that were assigned from ^{15}N-HSQC, HNCO, and HNCA spectra collected on a 500 MHz Bruker Avance-III spectrometer with a room temperature QXI probe.

1.3.2 NOESY Distance Restraints

Besides the NMR approaches that look at molecular connectivity using through-bond correlations (COSY, E. COSY, TOCSY, etc.), NMR can also be used to look at through-space molecular connectivity. Nuclear Overhauser effect spectroscopy (NOESY) is an unparalleled technique that gives exact interatomic distance information about a molecule or molecular complex. Only a brief introduction is given here, as a thorough explanation of NOE theory can be found in chapter 3.

NOESY relies on the principle that nearby dipoles can relax one another in solution. In a simple two-spin system, relaxation rates are comprised of auto-relaxation and cross-relaxation (NOE) components [56]. The NOE rate exhibits a strong distance (r) dependence, which can be used to back calculate interatomic distances from 1.8 to 6.0 Å with moderate accuracy.

$$NOE \propto \frac{1}{r^6} \tag{21}$$

The NOE also exhibits a dependence on time and molecular size. The period where dipolar relaxation is allowed to occur is referred to as the mixing time. Initially the NOE intensity builds up rapidly during the mixing time, before falling off and decaying due to relaxation processes. The linear component of the NOE build-up allows for the accurate quantitation of internuclear distance restraints (**Figure 1.8**). Distance restraints can be obtained manually from NOE signal intensities by correlating known secondary structure distances against experimental data, or in an automated fashion using software like CYANA or PONDEROSA [57]–[59].

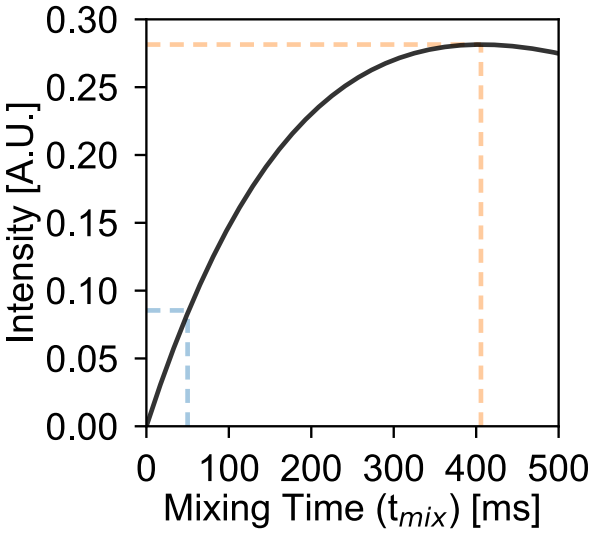


Figure 1.8. NOE build-up as function of mixing time. The linear dependence on mixing time is indicated by the blue dashed lines. The maximum NOE is given by orange dashed lines. After the maximum NOE has been reached, relaxation processes result in signal decay over time back to zero.

1.3.3 Residual Dipolar Couplings

As the protein size limitation is pushed further by NMR structural biologists, structural information is still limited to local spatial interactions via NOEs. Global structure or domain interactions are provided in relatively low resolution. NOEs provide an explicit distance restraint between two atoms in space, but residual dipolar couplings (RDCs) contain both distance and angular information in the form of an interatomic vector within a network of bond tensors. RDCs offer a complementary set of tertiary structural restraints to NOEs, encoding both long- and short-range bond orientation information. The RDC also contains slow dynamics information on the order of milliseconds and faster, as the vector orientation is spatially and temporally averaged over the course of the measurement [60], [61]. A brief introduction to the use of RDCs in protein structure determination will be described here, but a more complete description of RDC theory is described in many textbooks [62], [63].

All of the frequencies observed by NMR depend on the orientation of the molecule within the magnetic field. As previously mentioned, the *D*-coupling component of the nuclear magnetization is averaged to zero in an isotropic sample. Therefore, it is only possible to observe a fraction of the maximum anisotropic *D*-coupling when the analyte is in a state of partial alignment. The observed residual dipolar interaction between nuclei *A* and *B* (D^{AB}) can be described as [64]

$$D^{AB} = D^{AB}_{max}\langle P_2(\cos\theta)\rangle \tag{22}$$

where D^{AB}_{max} is the static dipolar coupling, P_2 is the second Legendre polynomial ($P_2(x) = \frac{1}{2}(3x^2 - 1)$), and θ is the angle between the internuclear vector and the magnetic field. Angular brackets denote the ensemble- and time-average value. The static dipolar coupling is then defined as:

$$D_{max}^{AB} = \frac{-\mu_0 \hbar \gamma_A \gamma_B}{4\pi^2 r_{AB}^3} \quad (23)$$

The constant $-\mu_0$ is the magnetic permittivity in a vacuum, $\hbar$ is the reduced Planck's constant, γ_A and γ_B are the gyromagnetic ratios of nucleus *A* and *B*, respectively, and *r* is the internuclear distance or bond length. In the case of a backbone amide H-N pair, the static coupling constant is equal to 12.5 kHz.

Since the protein is tumbling isotropically, the angle of the bond vector with respect to the magnetic field must be isolated from the ensemble averaged value of θ. For a rigid molecule, an arbitrary coordinate system can be defined with respect to the molecule by angles α_x, α_y, and α_z with respect to the Cartesian coordinate system [64]. The angles β_x, β_y, and β_z are used to relate the molecular axes to the magnetic field vector. Since θ is the scalar product between α and β, $\langle P_2(\cos\theta)\rangle$ can be redefined as

$$\langle P_2(\cos\theta)\rangle = \frac{3}{2}\langle(\cos\beta_x \cos\alpha_x + \cos\beta_y \cos\alpha_y + \cos\beta_z \cos\alpha_z)^2\rangle - \frac{1}{2} \quad (24)$$

The order matrix, proposed by Alfred Saupe in the 1960's, is the foundation for RDC analysis [65]–[67]. Using the previously defined coordinate system, the 3 × 3 matrix is comprised of elements

$$S_{pq} = \frac{3}{2}\langle\cos\beta_p \cos\beta_q\rangle - \delta_{pq} \quad (25)$$

β_p represents the angle of the p^{th} axis in the fixed Cartesian coordinate system with respect to an ordering director (parallel to B_0), and δ_{pq} is the Kronecker delta function. Angular brackets represent the time averaged ensemble. Substituting S_{pq} into equation 24 yields

$$\langle P_2(\cos\theta)\rangle = \sum_{pq=\{x,y,z\}} S_{pq} \cos\alpha_p \cos\alpha_q \quad (26)$$

Subsequent substitution into equation 22 defines the RDC as follows:

$$D^{AB} = D_{max}^{AB} \sum_{pq=\{x,y,z\}} S_{pq} \cos\alpha_p \cos\alpha_q \quad (27)$$

The Saupe order matrix is traceless, as $\langle \cos \beta_x \rangle^2 + \langle \cos \beta_y \rangle^2 + \langle \cos \beta_z \rangle^2 = 1$, and symmetric, as $\langle \cos \beta_p \cos \beta_q \rangle = \langle \cos \beta_q \cos \beta_p \rangle$, therefore only five independent elements of *S* are contained in the laboratory frame: S_{xy}, S_{xz}, S_{yz}, S_{zz}, and either S_{xx} or S_{yy} as these parameters are interdependent [64].

The interaction vector of the molecular coordinate frame defined by $\cos \alpha_p$ is fixed by the molecular geometry, allowing the independent elements of the matrix to be solved when defined by the diagonal. The symmetry of the Saupe matrix allows for diagonalization, where isotropic averaging ensures only differences in the diagonal values contribute to the RDC alignment

$$\langle \cos \beta_i \rangle^2 = {S_{ii}}^2 = \frac{1}{3} + A_{ii} \tag{28}$$

where A_{ii}, is the alignment tensor in the i^{th} direction with respect to the magnetic field.

$$D^{AB}(\alpha_x, \alpha_y, \alpha_z) = \frac{3}{2} D_{max}^{AB} \left[A_{xx} \cos^2 \alpha_x + A_{yy} \cos^2 \alpha_y + A_{zz} \cos^2 \alpha_z \right] \tag{29}$$

It is then possible to relate the molecule back to polar coordinates, as $\alpha_z = \theta$.

$$D^{AB}(\theta, \phi) = \frac{3}{2} D_{max}^{AB} \left[A_{xx} \sin^2 \theta \cos^2 \phi + A_{yy} \sin^2 \theta \sin^2 \phi + A_{zz} \cos^2 \theta \right] \tag{30}$$

The axial component of the alignment tensor is given by $A_a = \frac{3}{2} A_{zz}$, and the rhombic component is given by $A_r = A_{xx} - A_{yy}$. Because of the diagonalization of the matrix, the maximum value of $A_{zz} = \frac{2}{3}$ which makes the maximum value of $A_a = 1$ when the alignment tensor is completely parallel to the magnetic field. Substitution and rearrangement gives the common RDC expression

$$D^{AB}(\theta, \phi) = D_a \left((3 \cos^2 \theta - 1) + \frac{3}{2} R \sin^2 \theta \cos 2\phi \right) \tag{31}$$

where $D_a = \frac{D_{max}^{AB} A_a}{2}$ and reports on the degree of molecular alignment, and $R = A_r / A_a$ is the rhombicity or shape of molecular alignment.

RDCs can be used directly to refine an unknown structure, or they can be fit to a known structure by singular value decomposition (SVD) analysis [68]. RDCs manifest as a peak splitting with the *J*-coupling. In a practical sense, they can be measured by first measuring the ^{1}J bond coupling in an isotropic sample, followed by subsequent measurement of the $^{1}J+D$ coupling sample in a partial alignment media. The frequency difference between the couplings yields the RDC.

The availability of compatible partial alignment media for use with protein samples impeded the use of RDCs in structural refinement early on. Currently a variety of partial alignment media options exist, which offer diverse sample compatibilities and make the use of RDCs in the refinement of high-resolution structures much more commonplace [69]–[72]. However, there is no single alignment medium that can be applied to every biomolecular system of interest. The development and characterization of new alignment media is still an active area of research within the field of biomolecular NMR.

1.3.4 Hydrogen Bonds

Amide hydrogen bonds and hydrophobic group Van der Waals interactions are chiefly responsible for the global fold of a protein structure. Residue-specific hydrogen bond restraints can be included in a simulated annealing protocol of structure calculation software like XPLOR-NIH or CYANA [58], [73], [74]. It is not recommended to include these without multiple forms of supporting experimental evidence, as they can drastically impact the final protein structure. These restraints can be inferred from indirect evidence like amide chemical shifts and relative peak intensities, where resonances with restricted rotational motion (i.e. constrained by a hydrogen bond) show reduced peak intensity compared to random coil signals. The more generally accepted approach however is to directly measure scalar couplings across hydrogen

bonding pairs or to calculate the fraction of $^{1}H/^{2}H$ exchange. The magnitude of the measured $^{3}J_{NC'}$ coupling is directly proportional to the strength, therefore distance, of the hydrogen bond [75]. Similarly, ^{1}H atoms that are held tightly in a hydrogen bond are going to be less likely to exchange with ^{2}H from the bulk solvent [76].

If direct experimental evidence for hydrogen bonds is lacking or unavailable, XPLOR-NIH allows for the use of the hydrogen bonding database (HBDB) to generate hydrogen bonding pairs based on statistical likelihood [77]. It is also commonly accepted to include hydrogen bond restraints in later structural refinements when initial simulated annealing simulations consistently generate the same hydrogen bonding pairs.

1.3.5 Paramagnetic Relaxation Enhancement

Complementary to the NOE, paramagnetic relaxation enhancement (PRE) restraints have become increasingly popular in larger protein systems where perdeuteration prevents the measurement of the vast majority of NOE contacts [78]. PRE experiments involve the use of an inert paramagnetic probe which can be added to the solvent or chemically ligated to a protein or lipid directly [79]. The unpaired electrons in the probe create a local magnetic field parallel to B_0. Nuclei that are proximal to the probe experience increased R_1 and R_2 relaxation rates as a function of their distance from the probe. The range of effect for the paramagnetic probe is significantly larger than that of a dipolar relaxation interaction, and appropriate computational methodology did not exist to implement these restraints until 2004 [80], [81]. In the past decade, PRE-based structural restraints have proven to be an irreplaceable validation tool for protein structure models [82].

This allows for the generation of structural restraints based on paramagnetic probe location, similar to extrapolating interatomic distance information from NOE signals.

Although PRE restraints can be extrapolated from both R_1 and R_2 rates, R_2 provides a more reliable measurement [83], [84]. R_1 has greater susceptibility to cross-relaxation effects for ^{1}H spin, and internal motions, than R_2, and the measured R_1 rates are often considerably smaller than R_2 rates. PREs are most often quantitatively expressed as a ΔR_2 parameter in Hz, or a paramagnetic probe concentration normalized ΔR_2 parameter in Hz·mM^{-1}.

PREs follow the same distance dependency as NOEs ($NOE \propto \frac{1}{r^6}$ (21). A modified version of the distance (r) restraint from Battiste and Clore is as follows [78]:

$$r = \left[\frac{K}{\Delta R_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_H{}^2\tau_c{}^2}\right)\right]^{-6} \quad (32)$$

The value of $K = 1.23 \times 10^{-32}$ cm^6·s^{-2} and represents a field independent constant specific to ^{1}H R_2 measurements. The molecular correlation time is given by τ_c, and the ^{1}H Larmor frequency is given by ω_H. Note that the final restraint will have the units of Å.

Incorporation of the distance restraints generated by equation 32 are only directly applicable to structure calculation when derived from residue site-labeling. Lipid-labeling or solvent-based paramagnetic probes afford a more qualitative approach to protein structural model selection. Although PREs derived from these probes can be quantitated, they are more commonly reported by signal intensity retention. An increase in the transverse relaxation rate of ^{1}H results in a decreased signal intensity in the processed spectrum, thereby the disappearance of peaks can indicate the position of residues nearest to the probes. These analyses provide qualitative insight to corroborate protein structures.

The regular periodic nature of solvent-accessible α-helical PREs encode additional information for peripheral and transmembrane helical proteins. By fitting known regions of helical secondary structure, it is possible to determine the first residue depth of membrane

insertion (*A*), the tilt angle of the helix with respect to the membrane (τ), and the azimuth angle of the first residue (ρ). A modified version of the non-linear regression function from Respondek *et. al.* is as follows [85]:

$$\Delta R_2 = \frac{k\pi}{6d^3}$$

$$d = A + 1.5 \sin\tau \cdot (n-1) - \cos\tau \cdot B \cdot \cos(1.745 \cdot (n-1) \cdot \rho) \quad (33)$$

The constant *k* is a PRE factor for paramagnetic agent concentration correction, *d* is the backbone nucleus depth, *n* is the residue number, and *B* is the helical radius for the reporter nucleus. In the case of H^N or H^α, *B* is set to 1.95 or 3.25 Å, respectively. Membrane topology and helical depth can be easily determined for helical regions with greater than four measured PREs.

1.4 Structure Calculation

1.4.1 Simulated Annealing and Structural Refinement

Solution NMR structures are energetically favorable models that satisfy as many experimental restraints as possible. Therefore the most robust NMR structures are those determined with many consistent experimental restraints. A greater number of restraints will funnel the potential energy map for the folding protein towards the most likely structure, while inclusion of fundamental physical properties (bond lengths, VDW interactions, H-bonding geometry, etc.) restricts the model to the most energetically plausible.

There are quite a few NMR structural calculation programs available at present, with XPLOR-NIH and CYANA being two of the most popular [58], [73], [74]. Once experimental restraints have been obtained, they can be used in a simulated annealing procedure to generate a structure. A simulated annealing is basically a folding process of an extended structure of the protein primary sequence in a high temperature bath that ramps down to thermal equilibrium.

Experimentally derived restraints guide the folding process and can be introduced as small segments or all at once. For example, it is a relatively common practice to introduce low resolution or more global folding restraints first and then use the output model as the starting place for more stringent structural calculations. The greatest benefit to this approach is a large reduction in the computational time required to generate the final model. Statistical analysis and validation of the physical properties of the final structure models can be checked using software such as WHAT IF or PSVS [86], [87].

RDCs have been commonly used in the refinement of previous structures [88], [89]. A fully folded structure can be loaded as the starting point while an RDC forcefield is applied to internuclear bond vector orientations. The refined output structure then contains structural information weighted towards satisfying the RDC restraints. Cross-validation statistical models can be generated from the use of multiple alignment media and RDC nuclei pairs, adding to the validity of a protein structure [90]. Similarly a new set of experimental RDCs can be evaluated for agreement with an existing protein model using the DC module from NMRpipe [39]. The Saupe order matrix is used to fit the couplings to the structure via singular value decomposition (SVD) while simultaneously fitting the D_a and R parameters (equation 31).

1.4.2 Automation

Thanks to recent advancements, it is now possible to almost entirely automate the protein assignment process through to determination of the protein structure. Within NMRFAM-SPARKY backbone peak picking of processed spectra can be completely automated by the APES module and assigned in the Probabilistic Interaction Network of Evidence (PINE) module [91], [92]. If the peak list is generated elsewhere, a PINE client is now available through a web browser and accepts a variety of peak list formats [93]. Raw TOCSY and NOESY spectra can

then be included with the output backbone assignments for automated side-chain assignment and NOESY restraint generation in the PONDEROSA client [59], [94]. PONDEROSA incorporates external software packages like TALOS-N, CYANA, and XPLOR-NIH for a robust, iterative structure determination process [55], [58], [73], [74]. This process utilizes databases of existing structures, so new and challenging protein systems with sparse and incomplete data may not be accurately modeled through automation. A more conservative approach would be to use a combination of manual and automated procedures for assignment and restraint generation.

1.5 Relaxation

1.5.1 Dipolar Relaxation

A net magnetic dipole moment (magnetization) that has been oriented away from the applied B_0 field by radio frequency (rf) pulses will eventually return to its equilibrium state ($M_{z,eq}$) oriented along B_0. This process is known as relaxation and is chiefly governed by two first-order rate laws [12]. The first process is called spin-lattice relaxation, or R_1.

$$M_z(t) = M_{z,eq} - \left[M_{z,eq} - M_z(0)\right]e^{-t \cdot R_1} \tag{34}$$

R_1 is defined as the rate by which the z-axis component of the magnetization vector (M_z) recovers to 63% of its equilibrium state. This is an enthalpic process driven by couplings between the nuclear spin in a higher energy state and the spins of the lattice in a lower energy state. The energy obtained from the rf pulses disperses into the lattice, and the spin returns to its equilibrium state. The pulsing repetition rate is dictated by R_1, as a fast return to equilibrium allows for less down time between experiment scans.

The second process is called spin-spin relaxation, or R_2.

$$M_{xy}(t) = M_{xy}(0)e^{-t \cdot R_2} \tag{35}$$

R_2 is defined as the rate by which the transverse (detectable) component of the magnetization vector (M_{xy}) decays to 37% of its initial value following the rf pulse perturbation. This is a process where proximal spins dephase one another, resulting in a loss of coherence. The linewidth of the NMR signal is directly determined by R_2 ($\nu_{1/2} = R_2/\pi$).

1.5.2 Correlation and Spectral Density Functions

The rate law equations are useful in the determination of the observed relaxation rates, but they fail to report on the stochastic processes that create fluctuations within the local magnetic fields at a nucleus. Fluctuations can manifest from sources such as rotational or translational diffusion, vibrational motions, or conformational heterogeneity. Ultimately relaxation rates report on molecular motions, and the motional behaviors in solution can be described by a time-dependent correlation function, $C(t)$ [95]:

$$C(t) = \frac{1}{5}\langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t))\rangle \tag{36}$$

P_2 represents the second Legendre polynomial, $\hat{\mu}(0)$ is the orientation of the dipole at some arbitrary time $t = 0$, and $\hat{\mu}(\mathrm{t})$ is the orientation of the dipole at time *t*. Angular brackets indicate the time- and ensemble-average. The Fourier transform of the correlation function then yields the spectral density function, $J(\omega)$, which can be used to relate motions at specific frequencies to their contributions to observed molecular relaxation properties (R_1, R_2, NOE).

$$J(\omega) = 2\int_0^\infty (\cos \omega t) C(t)\, dt \tag{37}$$

Using ^{15}N-^{1}H bond vectors as an example, the relaxation properties can be defined with respect to $J(\omega)$ [96].

$$R_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \tag{38}$$

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

$$+\frac{1}{6}c^2[3J(\omega_N) + 4\mathrm{J}(0)] \quad (39)$$

$$NOE = 1 + (\frac{\gamma_H}{\gamma_N})\, d^2[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]\frac{1}{R_1} \quad (40)$$

The symbols ω_H and ω_N represent the Larmor frequencies for ^{1}H and ^{15}N, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, and both d^2 and c^2 are constants.

$$d^2 = 0.1(\frac{\hbar^2 {\gamma_N}^2 {\gamma_H}^2}{\langle {r_{NH}}^3 \rangle^2}) \quad (41)$$

$$c^2 = (\frac{2}{15})\, {\omega_N}^2(\sigma_\parallel - \sigma_\perp)^2 \quad (42)$$

The distance between ^{1}H and ^{15}N nuclei is given by r_{NH} and is approximately 1.02 Å, while $\sigma_\parallel$ and $\sigma_\perp$ represent the parallel and perpendicular components of the assumed axially symmetric ^{15}N chemical shift tensor.

1.5.3 Lipari-Szabo Formalism

The model-free formalism was introduced by Lipari and Szabo in 1982 as a minimalistic approach to the description of both global and internal motions of a bond vector [97], [98]. To apply the formalism, the overall molecular motions ($C_O(t)$) must be independent of the internal motions ($C_I(t)$).

$$C(t) = C_O(t) \cdot C_I(t) \quad (43)$$

The correlation function for the overall motion of a molecule can be defined by translational diffusion (D_t) and the rotational correlation time (τ_c).

$$C_O(t) = \frac{1}{5}e^{-6D_t \cdot t} = \frac{1}{5}e^{-t/\tau_c} \quad (44)$$

When $\hat{\mu}$ describes the orientation vector that is internally rigid, $C_I(t)$ can be described as

$$C_I(t) = \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \quad (45)$$

However, internal motions can be randomly diffusive and therefore explained by a series of exponentials.

$$C_I(t) = \sum_{i=0} a_i e^{-t/\tau_i} \tag{46}$$

The magnitude of the motion is represented by a_i, where $0 \leq a_i \leq 1$ for all *i*. The correlation time of the motion is given by τ_i, where $\tau_0 = \infty$, and $\tau_1 > \tau_2 > \tau_3$ ... for all *i*. This allows the correlation functions at $t = 0$ and $t = \infty$ to be determined exactly [97].

$$C_I(0) = \langle P_2\big(\hat{\mu}(0) \cdot \hat{\mu}(0)\big)\rangle = 1 \tag{47}$$

$$C_I(\infty) = \mathcal{S}^2 \tag{48}$$

Where $\mathcal{S}^2$ is the square of the order parameter, which reports on the rigidity of the bond vector without the use of a motional model. $\mathcal{S}^2$ is restricted to $0 \leq a_i \leq 1$, where a value of 0 indicates isotropic or large amplitude motions and a value of 1 indicates complete motional restriction.

The simplest exact approximation of $C_I(t)$ can then be reduced to

$$C_I{}^A(t) = \mathcal{S}^2 + (1 - \mathcal{S}^2)e^{-t/\tau_e} \tag{49}$$

with τ_e as the effective correlation time.

Equation 44 can now be explained with substitution of equations 45 and 50

$$C(t) = \frac{1}{5}\mathcal{S}^2 e^{-t/\tau_c} + \frac{1}{5}(1 - \mathcal{S}^2)e^{-t/\tau} \tag{50}$$

with $\tau^{-1} = \tau_c{}^{-1} + \tau_e{}^{-1}$. The resulting Fourier transform yields the spectral density as

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_c}{1+(\omega\tau_c)^2} + \frac{(1-\mathcal{S}^2)\tau}{1+(\omega\tau)^2}\right) \tag{51}$$

This is not a substitute of a discrete two-step motional model, but an approximation of a moment of this function.

The theory has since been expanded to incorporate distinct internal motions that differ by at least an order of magnitude with respect to timescale, and incorporates a chemical exchange

(R_{ex}) parameter to model R_2 [99], [100]. Computational advances have allowed for an automated approach to model fitting [101]–[103]. FAST-ModelFree in particular uses a GUI-based Monte Carlo fitting algorithm to fit the best motional model from the expanded model free formalism and determination of the corresponding order parameters, R_{ex} rates (if applicable), and τ_c for the system [102]. However, greater model complexity can result in greater errors in parameter estimation. A more reliable method would be to fit relaxation data from two different B_0 fields whenever possible.

1.6 Diffusion Ordered Spectroscopy

A very useful subset of NMR techniques involves tracking translational diffusion of a molecule. These are called pulse field gradient diffusion ordered spectroscopy (PFG-DOSY). One of the following studies takes advantage of the longitudinal eddy-current delay bipolar pulse-pair (LED-BPP) subset of DOSY techniques. The basic principle involves the application of a gradient pulse to encode the position of the molecule within the gradient axis [104]. The molecule can diffuse freely in solution over some time before an opposing gradient pulse decodes the new position of the molecule. Molecules that have changed position will fail to fully refocus, thus the rate of translational diffusion (D_t) will be encoded within the signal decay over the diffusion period.

DOSY provides valuable information on the size of the protein of interest. With the assumption of a generally spherical soluble protein, one can calculate the expected molecular weight of the system based on D_t [105]. However for non-spherical molecules, like IDPs, the assumption no longer holds. IDPs often possess transient structural characteristics in response to

pH, temperature, and ligands. Although DOSY cannot accurately determine the size of these systems, it can shed light on the degree of compaction under various conditions [106].

In later chapters, we apply DOSY measurements as a means to determine the bound percentage of an IDP to micelles at varying pH values. Determination of D_t of free lipid, free protein, and protein-micelle complex affords the free concentration of protein by

$$\frac{D_{t,protein}}{D_{t,micelle}} = \frac{[protein]-[protein]_{free}}{[protein]} + \frac{D_{t,protein-free}[protein]_{free}}{D_{t,micelle[protein]}} \quad (52)$$

This allows for the calculation of the concentration of micelle-complexed protein and the percentage bound [107]. Comparison across multiple conditions then provides information on relative affinities as a function of pH.

1.7 Chemical Shift Perturbation and Chemical Exchange

Chemical shifts are extremely sensitive probes of local micro-environment. In solution-state NMR, minute changes in pH, temperature, or local structure from ligand binding events are reflected in the ensemble average chemical shift. Tracking the chemical shift change with addition of a ligand is called the chemical shift perturbation (CSP) or the complexation-induced changes in shift (CIS). Using just the chemical shift, various inferences can be made about protein-ligand interactions. These include residue-specific binding site locations, binding kinetics, and binding modes. A detailed review of CSP applications can be found in [108].

Evidence of chemical exchange from ligand binding events can be used to infer the binding kinetics as well as the protein-ligand dissociation constant (K_D) [68], [109]. Additionally relaxation dispersion NMR methods can be used to determine the relative populations of exchanging species [110]–[113]. This type of methodology is incredibly useful for looking at transient states that would otherwise be rendered invisible by other structural biology techniques. A detailed review on practical applications of relaxation dispersion spectroscopy can be found in [114]. These unique real-time capabilities are what make NMR invaluable to the structural biology community.

1.8 Complimentary Biophysical Techniques

1.8.1 Size-exclusion Chromatography

A very useful complimentary method to structural biology by solution-state NMR is called size-exclusion chromatography (SEC), or sometimes gel filtration. This method allows for the estimation of the molecular weight (MW) of a protein or biomolecular complex through the assumption that the overall molecular shape is roughly spherical. Comparing a biomolecule of interest to a series of known standards yields an estimated hydrodynamic radius and thus molecular weight of the system [115]. Despite a few key disadvantages with this method, SEC can be a relatively simple way to confirm the molecular size of a protein system.

SEC is performed by slowly flowing a mobile phase containing the protein of interest over a stationary phase of an inert matrix of porous material with variable pore sizes [116]. The stationary phase commonly consists of silica beads coated with hydrophilic functional groups such as dextrose and agarose. The variable pore sizes produced from the gel bed packing limit the number of paths a molecule can take through the stationary phase based on molecular size.

Larger molecules have fewer routes of travel and elute first, while smaller molecules enter a greater number of pores and elute later. Simultaneous detection of protein in the eluent can be achieved through coupling of a UV light detector, common to most FPLC systems.

A series of globular protein standards with known MW are required to generate a standard log-linear regression curve for a given column. The gel phase distribution coefficient is determined for each protein standard (K_{av}).

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \tag{53}$$

V_e represents the elution volume of each protein, V_c represents the total column volume, and V_0 represents the column void volume. A linear regression analysis of K_{av} by $\log(MW)$ for each protein standard generates a working calibration curve, which allows for the size determination of the protein of interest. An example calibration curve is provided in **Figure 1.9**.

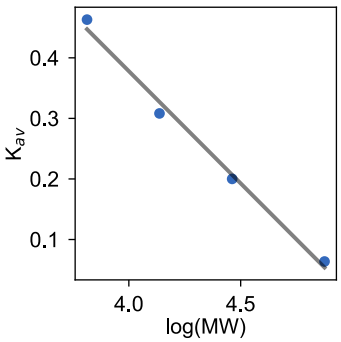


Figure 1.9. Example MW standard calibration curve for SEC. From left to right: conalbumin, carbonic anhydrase, ribonuclease, aprotinin. Theoretical relative MW can be calculated for an unknown protein K_{av} between the mass range of 75000 (conalbumin) and 6500 Da (aprotinin) using the slope and intercept of the linear regression.

SEC is a common biophysical technique due to its many advantages. The data collection and analysis are straightforward and easy to perform. Additionally, the inert properties of the stationary phase make it compatible with a wide range of physiologically relevant biochemical conditions. It can also be used to determine the size of protein-protein or protein-detergent-lipid complexes [117], [118]. The well-characterized sizes and shapes of detergent micelles and bicelles allow for extrapolation of bound oligomeric states in favorable cases [119], [120]. However, these results can also be misleading if the protein of interest changes the properties (e.g. aggregation number) of the micelle or bicelle.

Disadvantages to protein characterization by SEC chiefly revolve around the assumption of a spherical or globular protein system. This is due to the use of globular protein standards to determine the relative MW of an unknown protein. The linearity of denatured proteins, IDPs, and protein-DNA complexes prohibit reliable relative MW estimation by SEC. However, in principle one could analyze a set of linear molecular standards to compare with an unknown unfolded (linearized) protein. More recently, the coupling of multi-angle light scattering optics to SEC (SEC-MALS) has allowed for accurate size and shape determination of protein complexes [121]. With these advances, SEC remains a worthwhile technique in tandem with protein characterization by solution NMR.

1.8.2 Analytical Ultracentrifugation

One of the most versatile biophysical characterization techniques available is analytical ultracentrifugation (AUC). AUC uses the mass dependence of a centrifugal force, and it offers a broad range of applications to biomolecules in solution. Analytes can range from small peptides (< 1 kDa) to whole virus particles or cells [122], [123]. Samples are prepared non-invasively in biologically relevant, near-native conditions, without the use of matrices or surface

immobilization. AUC can provide information on molecular size, shape, mass, stoichiometry, and oligomeric or complex association constant through two complementary measurement techniques: sedimentation velocity and sedimentation equilibrium [124], [125]. For an excellent description of the principles and theory of both techniques, see Cole 2008 [126].

1.8.3 Circular Dichroism

Proteins are intrinsically chiral due to their being formed from chiral amino acids. Additionally their peptide backbones have a local handedness, sense of twist, making the amide transitions optically active and representative of the secondary structure. Protein secondary structures (helix, sheet, coil) produce unique fingerprint optical properties due to differences in the absorbance of left and right circularly polarized light [127]. The measurement and quantitation of a molecule's preference to absorb one sense of direction of circularly polarized light is called circular dichroism (CD), and it is widely used method for the characterization of protein secondary structure. A standard CD optical set-up is provided in **Figure 1.10**.

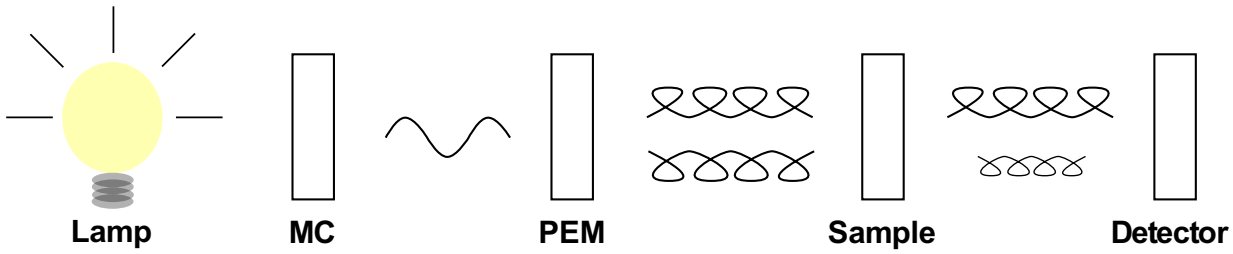


Figure 1.10. Simplified CD spectropolarimeter instrumentation diagram. High-intensity source (Lamp) emits light through a monochromator (MC) which also linearly polarizes light. Light is passed through the photoelastic modulator (PEM) and is converted into alternating left- and right-hand circularly polarized light. The light passes through the sample and the difference in absorbance is measured by the detector and converted into a CD spectrum as a function of wavelength.

A CD spectropolarimeter uses a high intensity light source, typically xenon gas discharge, with a range from deep UV to the visible spectrum. The light is first passed through a monochromator and linear polarizer. Linear polarized light is then converted into circularly polarized light by a photoelastic modulator. The modulator alternates between left-hand and right-hand circularly polarized light, which is passed through the sample solution to the detector. The difference in absorbance between left- and right-hand circularly polarized light is then recorded, where A_l is left-hand absorbance, A_r is right-hand absorbance, and ΔA is the difference between the two.

$$\Delta A = A_l - A_r \tag{54}$$

Rewriting in terms of Beer's Law gives the expression in terms of molar absorptivity (ε), sample concentration (c), and pathlength (l).

$$\Delta A = (\varepsilon_l - \varepsilon_r)cl \tag{55}$$

CD is commonly reported in degrees of ellipticity (θ), which can be directly related to the difference in absorbance ($\Delta A = A_l - A_r$) of the left and right circularly polarized components [128].

$$\theta = 32.98\Delta A \tag{56}$$

Substituting Beer's Law and normalizing the contribution of the sample concentration and pathlength yields molar ellipticity ([θ]).

$$[\theta] = 3298\Delta\varepsilon \tag{57}$$

The factor of 100 applied to 32.98 yields the traditional units of deg·cm^2·dmol^{-1}.

Protein samples notably contain characteristic minima and maxima features in the given in the far-UV spectra [127], [129], [130]. An example spectral overlay is provided in **Figure 1.11**. This property allows for rapid assessment of protein structure in a given experimental

condition. The technique is non-invasive and can be applied to a variety of conditions. Because CD is an electronic absorbance phenomenon, using relatively high frequency light, it reports on molecular changes on a sub-nanosecond timescale. This makes it a very attractive technique to measure the effect of various factors (temperature, pH, ligands) on protein folding events. Likewise, CD can be used to very quickly screen solution NMR sample conditions, as the technique requires substantially less protein and the samples can often be re-used if the system remains stable.

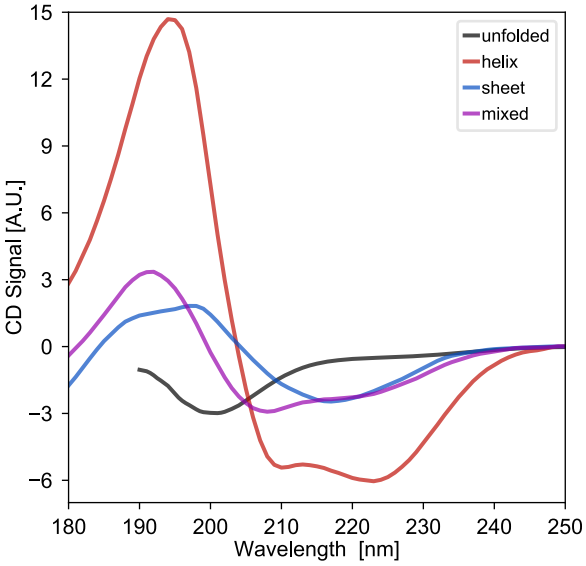


Figure 1.11. Example CD spectra for different secondary structure characteristics. Spectra include a completely unfolded IDP (human proIAPP, black), a purely helical protein (influenza virus HAfp, red), a purely beta sheet protein (rooster vitelline membrane outer layer protein I, blue), and a protein with mixed helix and sheet character (human ubiquitin, purple). The unfolded spectrum (black) was prepared and normalized from proIAPP experimental data collected on a Jasco J-810 spectropolarimeter. All other spectra were prepared by loading PDB structures 2KXA (red), 1VMO (blue), and 1UBQ (purple) into the PDB2CD converter [131].

CD is not without its disadvantages. Because this method relies on the absorbance of optically active species, sample components that absorb UV light can greatly impact the data quality. Specifically, halides like Cl^-, Br^-, and I^- cause large losses in signal near the shorter far-UV wavelengths. Compounds that contain carbon-carbon double bonds should also be avoided for this reason. Particles that scatter light (e.g. bicelles and vesicles) greatly reduce the reliability of experimental data, and often require cuvettes with shorter pathlengths that allow for greater protein concentrations to reduce the signal loss [132], [133]. Despite these problems, CD is a very useful technique for the rapid qualitative assessment of protein secondary structure in tandem with solution NMR studies.

1.9 Bibliography

[1] A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown, "Intrinsic protein disorder in complete genomes.," *Genome Inform. Ser. Workshop Genome Inform.*, vol. 11, pp. 161–71, 2000.

[2] M. E. Oates *et al.*, "D2P2: database of disordered protein predictions," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D508–D516, Nov. 2012.

[3] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd ed. Wiley, 2008.

[4] G. S. Rule and T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*. Dordrecht: Springer Netherlands, 2006.

[5] S. Forsén and R. A. Hoffman, "Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance," *J. Chem. Phys.*, vol. 39, no. 11, pp. 2892–2901, Dec. 1963.

[6] I. P. GEROTHANASSIS, A. TROGANIS, V. EXARCHOU, and K. BARBAROSSOU, "NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY: BASIC PRINCIPLES AND PHENOMENA, AND THEIR APPLICATIONS TO CHEMISTRY, BIOLOGY AND MEDICINE," *Chem. Educ. Res. Pr.*, vol. 3, no. 2, pp. 229–252, 2002.

[7] V. Sklenář and A. Bax, "Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra," *J. Magn. Reson.*, vol. 74, no. 3, pp. 469–479, Oct. 1987.

[8] M. Piotto, V. Saudek, and V. Sklenář, “Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions,” *J. Biomol. NMR*, vol. 2, no. 6, pp. 661–665, Nov. 1992.

[9] N. J. Baxter and M. P. Williamson, “Temperature dependence of 1H chemical shifts in proteins,” *J. Biomol. NMR*, vol. 9, no. 4, pp. 359–369, 1997.

[10] J. B. Matthew and F. M. Richards, “The pH dependence of hydrogen exchange in proteins.,” *J. Biol. Chem.*, vol. 258, no. 5, pp. 3039–44, Mar. 1983.

[11] R. Spadaccini, S. Leone, M. F. Rega, C. Richter, and D. Picone, “Influence of pH on the structure and stability of the sweet protein MNEI,” *FEBS Lett.*, vol. 590, no. 20, pp. 3681–3689, Oct. 2016.

[12] J. Cavanagh, W. Fairbrother, A. Palmer III, M. Rance, and N. Skelton, *Protein NMR Principles And Practice*, 2nd ed. Burlington, MA: Academic Press, 2007.

[13] J. Jeener and G. Alewaeters, “‘Pulse pair technique in high resolution NMR’ a reprint of the historical 1971 lecture notes on two-dimensional spectroscopy.,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 94–95, pp. 75–80, 2016.

[14] G. A. Morris and R. Freeman, “Enhancement of nuclear magnetic resonance signals by polarization transfer,” *J. Am. Chem. Soc.*, vol. 101, no. 3, pp. 760–762, Jan. 1979.

[15] H. L. Crespi, R. M. Rosenberg, and J. J. Katz, “Proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains.,” *Science*, vol. 161, no. 3843, pp. 795–6, Aug. 1968.

[16] H. L. CRESPI and J. J. KATZ, “High Resolution Proton Magnetic Resonance Studies of Fully Deuterated and Isotope Hybrid Proteins,” *Nature*, vol. 224, no. 5219, pp. 560–562, Nov. 1969.

[17] J. L. Markley, I. Putter, and O. Jardetzky, “High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease.,” *Science*, vol. 161, no. 3847, pp. 1249–51, Sep. 1968.

[18] I. Putter, A. Barreto, J. L. Markley, and O. Jardetzky, “Nuclear magnetic resonance studies of the structure and binding sites of enzymes. X. Preparation of selectively deuterated analogs of staphylococcal nuclease.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 64, no. 4, pp. 1396–403, Dec. 1969.

[19] M. Ikura, M. Krinks, D. A. Torchia, and A. Bax, “An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling,” *FEBS Lett.*, vol. 266, no. 1–2, pp. 155–158, Jun. 1990.

[20] M. Ikura, L. E. Kay, and A. Bax, “A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin,” *Biochemistry*, vol. 29,

no. 19, pp. 4659–4667, May 1990.

[21] F. C. Neidhardt, P. L. Bloch, and D. F. Smith, “Culture medium for enterobacteria.,” *J. Bacteriol.*, vol. 119, no. 3, pp. 736–47, Sep. 1974.

[22] D. M. LeMaster and F. M. Richards, “NMR Sequential Assignment of Escherichia coli Thioredoxin Utilizing Random Fractional Deuteriation,” *Biochemistry*, vol. 27, no. 1, pp. 142–150, 1988.

[23] S. Grzesiek, J. Anglister, H. Ren, and A. Bax, “Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of sequential amides,” *J. Am. Chem. Soc.*, vol. 115, no. 10, pp. 4369–4370, May 1993.

[24] T. Yamazaki *et al.*, “An HNCA Pulse Scheme for the Backbone Assignment of 15N,13C,2H-Labeled Proteins: Application to a 37-kDa Trp Repressor-DNA Complex,” *J. Am. Chem. Soc.*, vol. 116, no. 14, pp. 6464–6465, Jul. 1994.

[25] T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram, and L. E. Kay, “A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity,” *J. Am. Chem. Soc.*, vol. 116, no. 26, pp. 11655–11666, Dec. 1994.

[26] A. M. Ruschak and L. E. Kay, “Methyl groups as probes of supra-molecular structure, dynamics and function,” *J. Biomol. NMR*, vol. 46, no. 1, pp. 75–87, Jan. 2010.

[27] R. Verardi, N. J. Traaseth, L. R. Masterson, V. V Vostrikov, and G. Veglia, “Isotope Labeling for Solution and Solid-State NMR Spectroscopy of Membrane Proteins,” in *Advances in experimental medicine and biology*, vol. 992, NIH Public Access, 2012, pp. 35–62.

[28] M. K. Rosen, K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson, and L. E. Kay, “Selective Methyl Group Protonation of Perdeuterated Proteins,” *J. Mol. Biol.*, vol. 263, no. 5, pp. 627–636, Nov. 1996.

[29] L. Clark *et al.*, “Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris,” *J. Biomol. NMR*, vol. 62, no. 3, pp. 239–245, 2015.

[30] S. Rudiger, S. M. V Freund, D. B. Veprintsev, and A. R. Fersht, “CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90,” *Proc. Natl. Acad. Sci.*, vol. 99, no. 17, pp. 11085–11090, Aug. 2002.

[31] M. Cai, Y. Huang, K. Sakaguchi, G. M. Clore, A. M. Gronenborn, and R. Craigie, “An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli,” *J. Biomol. NMR*, vol. 11, no. 1, pp. 97–102, 1998.

[32] L. Braunschweiler and R. . Ernst, “Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy,” *J. Magn. Reson.*, vol. 53, no. 3, pp. 521–528, Jul. 1983.

[33] S. Grzesiek and A. Bax, “An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins,” *J. Magn. Reson.*, vol. 99, no. 1, pp. 201–207, Aug. 1992.

[34] S. Grzesiek and A. Bax, “Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR,” *J. Am. Chem. Soc.*, vol. 114, no. 16, pp. 6291–6293, Jul. 1992.

[35] A. Bax, G. M. Clore, and A. M. Gronenborn, “1H-1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins,” *J. Magn. Reson.*, vol. 88, no. 2, pp. 425–431, Jun. 1990.

[36] E. T. Olejniczak, R. X. Xu, and S. W. Fesik, “A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins,” *J. Biomol. NMR*, vol. 2, no. 6, pp. 655–659, Nov. 1992.

[37] K. Pervushin, R. Riek, G. Wider, and K. Wuthrich, “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.*, vol. 94, no. 23, pp. 12366–12371, Nov. 1997.

[38] J. E. Ollerenshaw, V. Tugarinov, and L. E. Kay, “Methyl TROSY: Explanation and experimental verification,” *Magn. Reson. Chem.*, vol. 41, no. 10, pp. 843–852, 2003.

[39] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, “NMRPipe: A multidimensional spectral processing system based on UNIX pipes,” *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[40] J. Ying, F. Delaglio, D. A. Torchia, and A. Bax, “Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data,” *J. Biomol. NMR*, vol. 68, no. 2, pp. 101–118, Jun. 2017.

[41] M. Norris, B. Fetler, J. Marchant, and B. A. Johnson, “NMRFx Processor: a cross-platform NMR data processing program,” *J. Biomol. NMR*, vol. 65, no. 3–4, pp. 205–216, Aug. 2016.

[42] R. L. J. Keller, “Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment,” ETH, 2005.

[43] T. D. Goddard and D. G. Kneller, “SPARKY 3,” *Univ. California, San Fransisco*, 2008.

[44] W. Lee, M. Tonelli, and J. L. Markley, “NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy,” *Bioinformatics*, vol. 31, no. 8, pp. 1325–1327, 2015.

[45] M. Karplus, “Contact Electron-Spin Coupling of Nuclear Magnetic Moments,” *J. Chem. Phys.*, vol. 30, no. 1, pp. 11–15, Jan. 1959.

[46] M. Karplus, “Vicinal Proton Coupling in Nuclear Magnetic Resonance,” *J. Am. Chem.*

Soc., vol. 85, no. 18, pp. 2870–2871, Sep. 1963.

[47] C. Griesinger, O. Sørensen, and R. Ernst, "Practical aspects of the E.COSY technique. Measurement of scalar spin-spin coupling constants in peptides," *J. Magn. Reson.*, vol. 75, no. 3, pp. 474–492, Dec. 1987.

[48] G. W. Vuister and A. Bax, "Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNHa) coupling constants in 15N-enriched proteins," *J. Am. Chem. Soc.*, vol. 115, no. 17, pp. 7772–7777, Aug. 1993.

[49] G. Balacco, "A Desktop Calculator for the Karplus Equation," *J. Chem. Inf. Comput. Sci.*, vol. 36, no. 4, pp. 885–887, Jan. 1996.

[50] D. S. Wishart, B. D. Sykes, and F. M. Richards, "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy," *Biochemistry*, vol. 31, no. 6, pp. 1647–1651, Feb. 1992.

[51] A. de Dios, J. Pearson, and E. Oldfield, "Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach," *Science (80-.).*, vol. 260, no. 5113, pp. 1491–1496, Jun. 1993.

[52] D. S. Wishart and B. D. Sykes, "The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data," *J. Biomol. NMR*, vol. 4, no. 2, pp. 171–180, Mar. 1994.

[53] G. Cornilescu, F. Delaglio, and A. Bax, "Protein backbone angle restraints from searching a database for chemical shift and sequence homology.," *J. Biomol. NMR*, vol. 13, no. 3, pp. 289–302, Mar. 1999.

[54] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, "TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts," *J. Biomol. NMR*, vol. 44, no. 4, pp. 213–223, 2009.

[55] Y. Shen and A. Bax, "Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks," *J. Biomol. NMR*, vol. 56, no. 3, pp. 227–241, Jul. 2013.

[56] I. Solomon, "Relaxation Processes in a System of Two Spins," *Phys. Rev.*, vol. 99, no. 2, pp. 559–565, Jul. 1955.

[57] K. Wüthrich, M. Billeter, and W. Braun, "Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances," *J. Mol. Biol.*, vol. 180, no. 3, pp. 715–740, 1984.

[58] P. Güntert, "Automated NMR Structure Calculation With CYANA," in *Protein NMR Techniques*, vol. 278, New Jersey: Humana Press, 2004, pp. 353–378.

[59] W. Lee, J. H. Kim, W. M. Westler, and J. L. Markley, "PONDEROSA, an automated 3D-

NOESY peak picking program, enables automated protein structure determination," *Bioinformatics*, vol. 27, no. 12, pp. 1727–1728, Jun. 2011.

[60] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard, "NMR evidence for slow collective motions in cyanometmyoglobin.," *Nat. Struct. Biol.*, vol. 4, no. 4, pp. 292–7, Apr. 1997.

[61] G. Bouvignies, P. Bernadó, and M. Blackledge, "Protein backbone dynamics from N-HN dipolar couplings in partially aligned systems: a comparison of motional models in the presence of structural noise.," *J. Magn. Reson.*, vol. 173, no. 2, pp. 328–38, Apr. 2005.

[62] N. R. Krishna and L. J. Berliner, *Structure Computation and Dynamics in Protein NMR*. New York: Kluwer Academic Publishers, 2002.

[63] K. Schmidt-Rohr and H. W. Spiess, *Multidimensional Solid-State NMR and Polymers*. Elsevier Ltd, 2012.

[64] A. Bax, G. Kontaxis, and N. Tjandra, "Dipolar couplings in macromolecular structure determination," in *Methods in Enzymology*, 2001.

[65] A. Saupe and G. Englert, "High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules," *Phys. Rev. Lett.*, vol. 11, no. 10, pp. 462–464, Nov. 1963.

[66] A. Saupe, "Kernresonanzen in kristallinen Flüssigkeiten und in kristallinflüssigen Lösungen. Teil I," *Zeitschrift für Naturforsch. A*, vol. 19, no. 2, pp. 161–171, Jan. 1964.

[67] A. Saupe, "Recent Results in the Field of Liquid Crystals," *Angew. Chemie Int. Ed. English*, vol. 7, no. 2, pp. 97–112, Feb. 1968.

[68] M. Arai, J. C. Ferreon, and P. E. Wright, "Quantitative Analysis of Multisite Protein–Ligand Interactions by NMR: Binding of Intrinsically Disordered p53 Transactivation Subdomains with the TAZ2 Domain of CBP," *J. Am. Chem. Soc.*, vol. 134, no. 8, pp. 3792–3803, Feb. 2012.

[69] R. Tycko, F. J. Blanco, and Y. Ishii, "Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole−Dipole Couplings in High-Resolution Biomolecular NMR," *J. Am. Chem. Soc.*, vol. 122, no. 38, pp. 9340–9341, Sep. 2000.

[70] H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, and S. Grzesiek, "Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes," *J. Biomol. NMR*, vol. 18, no. 4, pp. 303–309, 2000.

[71] J. Lorieau, L. Yao, and A. Bax, "Liquid Crystalline Phase of G-Tetrad DNA for NMR Study of Detergent-Solubilized Proteins," *J. Am. Chem. Soc.*, vol. 130, no. 24, pp. 7536–7537, Jun. 2008.

[72] P. Thiagarajan-Rosenkranz, A. W. Draney, S. T. Smrt, and J. L. Lorieau, "A Positively

Charged Liquid Crystalline Medium for Measuring Residual Dipolar Couplings in Membrane Proteins by NMR," *J. Am. Chem. Soc.*, vol. 137, no. 37, pp. 11932–11934, Sep. 2015.

[73] C. D. Schwieters, J. J. Kuszewski, and G. Marius Clore, "Using Xplor-NIH for NMR molecular structure determination," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 48, no. 1. pp. 47–62, Mar-2006.

[74] C. D. Schwieters, G. A. Bermejo, and G. M. Clore, "Xplor-NIH for molecular structure determination from NMR and other data sources," *Protein Sci.*, vol. 27, no. 1, pp. 26–40, 2018.

[75] G. Cornilescu, B. E. Ramirez, M. K. Frank, G. M. Clore, A. M. Gronenborn, and A. Bax, "Correlation between 3h J NC ' and Hydrogen Bond Length in Proteins," *J. Am. Chem. Soc.*, vol. 121, no. 26, pp. 6275–6279, Jul. 1999.

[76] S. N. Loh and J. L. Markley, "Hydrogen Bonding in Proteins As Studied by Amide Hydrogen D/H Fractionation Factors: Application to Staphylococcal Nuclease," *Biochemistry*, vol. 33, no. 4, pp. 1029–1036, Feb. 1994.

[77] A. Grishaev and A. Bax, "An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation," *J. Am. Chem. Soc.*, vol. 126, no. 23, pp. 7281–7292, 2004.

[78] J. L. Battiste and G. Wagner, "Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data †," *Biochemistry*, vol. 39, no. 18, pp. 5355–5365, May 2000.

[79] E. Schrank, G. Wagner, and K. Zangger, "Solution NMR Studies on the Orientation of Membrane-Bound Peptides and Proteins by Paramagnetic Probes," *Molecules*, vol. 18, no. 7, pp. 7407–7435, Jun. 2013.

[80] J. Iwahara, C. D. Schwieters, and G. M. Clore, "Ensemble Approach for NMR Structure Refinement against 1 H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule," *J. Am. Chem. Soc.*, vol. 126, no. 18, pp. 5879–5896, May 2004.

[81] L. Banci, I. Bertini, G. Cavallaro, A. Giachetti, C. Luchinat, and G. Parigi, "Paramagnetism-Based Restraints for Xplor-NIH," *J. Biomol. NMR*, vol. 28, no. 3, pp. 249–261, Mar. 2004.

[82] L. Shi, N. J. Traaseth, R. Verardi, M. Gustavsson, J. Gao, and G. Veglia, "Paramagnetic-Based NMR Restraints Lift Residual Dipolar Coupling Degeneracy in Multidomain Detergent-Solubilized Membrane Proteins," *J. Am. Chem. Soc.*, vol. 133, no. 7, pp. 2232–2241, Feb. 2011.

[83] G. M. Clore, C. Tang, and J. Iwahara, "Elucidating transient macromolecular interactions

using paramagnetic relaxation enhancement," *Curr. Opin. Struct. Biol.*, vol. 17, no. 5, pp. 603–616, Oct. 2007.

[84] N. L. Fawzi, J. Ying, D. A. Torchia, and G. M. Clore, "Kinetics of Amyloid β Monomer-to-Oligomer Exchange by NMR Relaxation," *J. Am. Chem. Soc.*, vol. 132, no. 29, pp. 9948–9951, Jul. 2010.

[85] M. Respondek, T. Madl, C. Göbl, R. Golser, and K. Zangger, "Mapping the Orientation of Helices in Micelle-Bound Peptides by Paramagnetic Relaxation Waves," *J. Am. Chem. Soc.*, vol. 129, no. 16, pp. 5228–5234, Apr. 2007.

[86] G. Vriend, "WHAT IF: A molecular modeling and drug design program," *J. Mol. Graph.*, vol. 8, no. 1, pp. 52–56, Mar. 1990.

[87] A. Bhattacharya, R. Tejero, and G. T. Montelione, "Evaluating protein structures determined by structural genomics consortia.," *Proteins*, vol. 66, no. 4, pp. 778–95, Mar. 2007.

[88] G. A. Bermejo, G. M. Clore, and C. D. Schwieters, "Improving NMR Structures of RNA," *Structure*, vol. 24, no. 5, pp. 806–815, May 2016.

[89] G. M. Clore and C. D. Schwieters, "How Much Backbone Motion in Ubiquitin Is Required To Account for Dipolar Coupling Data Measured in Multiple Alignment Media as Assessed by Independent Cross-Validation?," *J. Am. Chem. Soc.*, vol. 126, no. 9, pp. 2923–2938, Mar. 2004.

[90] A. S. Maltsev, A. Grishaev, J. Roche, M. Zasloff, and A. Bax, "Improved Cross Validation of a Static Ubiquitin Structure Derived from High Precision Residual Dipolar Couplings Measured in a Drug-Based Liquid Crystalline Phase," *J. Am. Chem. Soc.*, vol. 136, no. 10, pp. 3752–3755, Mar. 2014.

[91] J. Shin, W. Lee, and W. Lee, "Structural proteomics by NMR spectroscopy," *Expert Rev. Proteomics*, vol. 5, no. 4, pp. 589–601, Aug. 2008.

[92] A. Bahrami, A. H. Assadi, J. L. Markley, and H. R. Eghbalnia, "Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy," *PLoS Comput. Biol.*, vol. 5, no. 3, Mar. 2009.

[93] W. Lee *et al.*, "I-PINE web server: an integrative probabilistic NMR assignment system for proteins," *J. Biomol. NMR*, vol. 73, no. 5, pp. 213–222, May 2019.

[94] W. Lee, J. L. Stark, and J. L. Markley, "PONDEROSA-C/S: client–server based software package for automated protein 3D structure determination," *J. Biomol. NMR*, vol. 60, no. 2–3, pp. 73–75, Nov. 2014.

[95] A. Abragam, *The Principles of Nuclear Magnetism*. Oxford, England: Clarendon Press, 1961.

[96] N. A. Farrow *et al.*, "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by15N NMR Relaxation," *Biochemistry*, vol. 33, no. 19, pp. 5984–6003, 1994.

[97] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," *J. Am. Chem. Soc.*, vol. 104, no. 17, pp. 4546–4559, Aug. 1982.

[98] G. Lipari and A. Szabo, "Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results," *J. Am. Chem. Soc.*, vol. 104, no. 17, pp. 4559–4570, 1982.

[99] G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn, "Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins," *J. Am. Chem. Soc.*, vol. 112, no. 12, pp. 4989–4991, Jun. 1990.

[100] G. M. Clore, P. C. Driscoll, P. T. Wingfield, and A. M. Gronenborn, "Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy.," *Biochemistry*, vol. 29, no. 32, pp. 7387–401, Aug. 1990.

[101] V. Y. Orekhov, D. E. Nolde, A. P. Golovanov, D. M. Korzhnev, and A. S. Arseniev, "Processing of heteronuclear NMR relaxation data with the new software DASHA," *Appl. Magn. Reson.*, vol. 9, no. 4, pp. 581–588, Jul. 1995.

[102] R. Cole and J. P. Loria, "FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data," *J. Biomol. NMR*, vol. 26, no. 3, pp. 203–213, 2003.

[103] M. Bieri, E. J. D'Auvergne, and P. R. Gooley, "relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μs motion of proteins," *J. Biomol. NMR*, vol. 50, no. 2, pp. 147–155, Jun. 2011.

[104] D. H. Wu, A. D. Chen, and C. S. Johnson, "An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses," *J. Magn. Reson. Ser. A*, vol. 115, no. 2, pp. 260–264, Aug. 1995.

[105] J. J. Chou, J. L. Baber, and A. Bax, "Characterization of Phospholipid Mixed Micelles by Translational Diffusion," *J. Biomol. NMR*, vol. 29, no. 3, pp. 299–308, Jul. 2004.

[106] R. Konrat, "NMR contributions to structural dynamics studies of intrinsically disordered proteins," *J. Magn. Reson.*, vol. 241, no. 1, pp. 74–85, Apr. 2014.

[107] S. T. Smrt, A. W. Draney, and J. L. Lorieau, "The Influenza Hemagglutinin Fusion Domain Is an Amphipathic Helical Hairpin That Functions by Inducing Membrane Curvature," *J. Biol. Chem.*, vol. 290, no. 1, pp. 228–238, Jan. 2015.

[108] M. P. Williamson, "Using chemical shift perturbation to characterise ligand binding,"

Prog. Nucl. Magn. Reson. Spectrosc., vol. 73, pp. 1–16, Aug. 2013.

[109] R. W. Fessenden, "Introduction to magnetic resonance: with applications to chemistry and chemical physics (Carrington, Alan; McLachlan, Andre D.)," *J. Chem. Educ.*, vol. 44, no. 12, p. 772, Dec. 1967.

[110] S. Meiboom and D. Gill, "Modified Spin-Echo Method for Measuring Nuclear Relaxation Times," *Rev. Sci. Instrum.*, vol. 29, no. 8, pp. 688–691, Aug. 1958.

[111] H. Y. Carr and E. M. Purcell, "Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments," *Phys. Rev.*, vol. 94, no. 3, pp. 630–638, May 1954.

[112] P. Vallurupalli, D. F. Hansen, P. Lundström, and L. E. Kay, "CPMG relaxation dispersion NMR experiments measuring glycine 1Hα and 13Cα chemical shifts in the 'invisible' excited states of proteins," *J. Biomol. NMR*, vol. 45, no. 1–2, pp. 45–55, Sep. 2009.

[113] A. D. Bain, G. J. Duns, F. Rathgeb, and J. Vanderkloet, "A Study of Chemical Exchange in Unequally Populated Systems by Novel NMR Methodologies. Application to the Cis-Trans Isomerization in Furfural," 1995.

[114] G. M. Clore, "Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation," *Protein Sci.*, vol. 20, no. 2, pp. 229–246, Feb. 2011.

[115] Z. Grubisic, P. Rempp, and H. Benoit, "A universal calibration for gel permeation chromatography," *J. Polym. Sci. Part B Polym. Lett.*, vol. 5, no. 9, pp. 753–759, Sep. 1967.

[116] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 6th ed. New York: W. H. Freeman and Company, 2013.

[117] R. M. Wheaton and W. C. Bauman, "NON-IONIC SEPARATIONS WITH ION EXCHANGE RESINS," *Ann. N. Y. Acad. Sci.*, vol. 57, no. 3, pp. 159–176, Nov. 1953.

[118] M. le Maire, P. Champeil, and J. V Møller, "Interaction of membrane proteins and lipids with solubilizing detergents," *Biochim. Biophys. Acta - Biomembr.*, vol. 1508, no. 1–2, pp. 86–111, Nov. 2000.

[119] J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley, and S. Doniach, "Size and shape of detergent micelles determined by small-angle X-ray scattering," *J. Phys. Chem. B*, vol. 111, no. 43, pp. 12427–12438, 2007.

[120] R. C. Oliver, J. Lipfert, D. A. Fox, R. H. Lo, S. Doniach, and L. Columbus, "Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group," *PLoS One*, vol. 8, no. 5, p. 62488, 2013.

[121] E. Folta-Stogniew, "Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors," in *New and Emerging Proteomic Techniques*, New Jersey: Humana Press,

2006, pp. 97–112.

[122] X. Yang, S. Agarwala, S. Ravindran, and G. Vellekamp, “Determination of Particle Heterogeneity and Stability of Recombinant Adenovirus by Analytical Ultracentrifugation in CsCl Gradients,” *J. Pharm. Sci.*, vol. 97, no. 2, pp. 746–763, Feb. 2008.

[123] T. Arakawa, T. Niikura, Y. Kita, and F. Arisaka, “Structure analysis of short peptides by analytical ultracentrifugation: Review.,” *Drug Discov. Ther.*, vol. 3, no. 5, pp. 208–14, Oct. 2009.

[124] T. M. Laue, “Sedimentation equilibrium as thermodynamic tool,” in *Methods in Enzymology*, vol. 259, Academic Press, 1995, pp. 427–452.

[125] T. M. Laue and W. F. Stafford III, “MODERN APPLICATIONS OF ANALYTICAL ULTRACENTRIFUGATION,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 28, no. 1, pp. 75–100, Jun. 1999.

[126] J. L. Cole, J. W. Lary, T. P Moody, and T. M. Laue, “Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium.,” *Methods Cell Biol.*, vol. 84, pp. 143–79, 2008.

[127] N. Sreerama and R. W. Woody, “Protein Secondary Structure from Circular Dichroism Spectroscopy,” *J. Mol. Biol.*, vol. 242, no. 4, pp. 497–507, Sep. 1994.

[128] N. J. Greenfield, “Using circular dichroism spectra to estimate protein secondary structure,” *Nat. Protoc.*, vol. 1, no. 6, pp. 2876–2890, Dec. 2006.

[129] N. Sreerama and R. W. Woody, “On the analysis of membrane protein circular dichroism spectra,” *Protein Sci.*, vol. 13, no. 1, pp. 100–112, Jan. 2004.

[130] N. Sreerama and R. W. Woody, “Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set,” *Anal. Biochem.*, vol. 287, no. 2, pp. 252–260, Dec. 2000.

[131] L. Mavridis and R. W. Janes, “PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates,” *Bioinformatics*, vol. 33, no. 1, pp. 56–63, Jan. 2017.

[132] I. Tinoco, M. F. Maestre, and C. Bustamante, “Circular dichroism in samples which scatter light,” *Trends Biochem. Sci.*, vol. 8, no. 2, pp. 41–44, Feb. 1983.

[133] A. J. Miles and B. A. Wallace, “Circular dichroism spectroscopy of membrane proteins,” *Chem. Soc. Rev.*, vol. 45, no. 18, pp. 4859–4872, Sep. 2016.

Chapter 2

Pro-Islet Amyloid Polypeptide in Micelles Contains a Helical Prohormone Segment

This chapter has been submitted for publication to *The FEBS Journal* on October 12th, 2019, as "Micelle-bound pro-Islet Amyloid Polypeptide Contains a Carboxy-terminal Helical Segment", Charles F. DeLisle, Alexander L. Malooley, Indrani Banerjee, and Justin L. Lorieau, The FEBS Journal, 2019. *Under Revision*.

Pro-islet amyloid polypeptide (proIAPP) is the prohormone precursor molecule to IAPP, also known as amylin. IAPP is a calcitonin family peptide hormone that is co-secreted with insulin, and largely responsible for hunger satiation and metabolic homeostasis. Amyloid plaques containing mixtures of mature IAPP and misprocessed proIAPP deposit on, and destroy pancreatic β cell membranes, and they are recognized as a clinical hallmark of type 2 diabetes mellitus. In order to better understand the interaction with cellular membranes, we solved the solution NMR structure of proIAPP bound to dodecylphosphocholine micelles at pH 4.5. We show that proIAPP is a dynamic molecule with four α-helices. The first two helices are contained within the mature IAPP sequence, while the second two helices are part of the C-terminal prohormone segment (Cpro). We mapped the membrane topology of the amphipathic helices by paramagnetic relaxation enhancement, and we used circular dichroism and diffusion-

ordered spectroscopy to identify environmental factors that impact proIAPP membrane affinity. We discuss how our structural results relate to prohormone processing based on the varied pH environments and lipid compositions of organelle membranes within the regulated secretory pathway, and the likelihood of Cpro survival for co-secretion with IAPP.

2.1 Introduction

Human islet amyloid polypeptide (IAPP), or amylin, is a 37 amino-acid (aa) hormone peptide that is co-expressed with insulin in the β cells of the pancreas, and it is widely known to be associated with type 2 diabetes mellitus [1]–[3]. Although pancreatic amyloid deposits of this peptide are characteristic of the disease, the chief physiological role of amylin is to provide hunger satiation and maintain metabolic and bone homeostasis [4], [5]. IAPP belongs to the calcitonin family of hormone peptides, comprised of IAPP, calcitonin, calcitonin gene-related peptide, adrenomedullin, and intermedin [6]. This family shares a remarkable number of similarities. All members share the same two receptors, the calcitonin receptor and calcitonin-like receptor, which form heterodimers with one of the three isoforms of the receptor activity modifying protein. Common structural features include a tight peptide ring of 6-8 residues formed by a single intramolecular disulfide bond and the presence of a C-terminal amide, which are both necessary for complete biological activity [2], [7], [8].

To reach full maturity, the hormone peptides undergo a series of tightly regulated, tissue-specific processing events. IAPP is initially translated as an 89-aa pre-proIAPP protein in the β cells of the Langerhans islets [9]–[11]. The 22-aa N-terminal signaling peptide is cleaved and the intramolecular disulfide bond is formed in the endoplasmic reticulum, after which proIAPP is transferred to the Golgi network. The order of events that lead from proIAPP to the production

and storage of mature IAPP within the secretory granules of the β cells have been thoroughly characterized [12]–[14]. In the trans-Golgi network, prohormone convertase 1/3 (PC1/3) cleaves the 16-aa C-terminal prohormone (Cpro) segment, followed by PC2 cleavage of the 11-aa N-terminal prohormone (Npro) segment in the secretory granules along the regulated secretory pathway (RSP). PC2 can compensate in the absence of PC1/3 and cut at both termini [12], [15]. Carboxypeptidase E (CPE) removes the C-terminal dibasic PC cleavage recognition site, and the C-terminal glycine residue is converted into an amide by peptidyl glycine alpha-amidating monooxygenase (PAM). Mature IAPP is then stored with insulin inside the secretory granules for secretion in response to glycemic stimuli [16], [17].

It has been postulated that proIAPP processing intermediates seed amyloid formation *in vivo* [18], [19]. Research efforts have focused on the amyloidogenicity of IAPP precursor molecules as a means to better understand and prevent the formation of β cell plaques [20]–[24]. Initial steps in amyloid formation is thought to be a membrane-mediated event [25], and while studies have characterized proIAPP amyloid formation on the membrane [26], no membrane-bound proIAPP structure is available at present. Two NMR structures of the mature, membrane-bound human IAPP are available [27], [28]. However, both studies used non-native lipid headgroups with sodium dodecyl sulfate (SDS) micelles, and these studies lacked the additional structural information gained from ^{13}C and ^{15}N isotope labeling and residual dipolar couplings (RDCs) to accurately define the orientation and dynamics of the helices.

In this study, we expressed and purified recombinant human proIAPP with a -KKKKD solubility tail and included the ^{13}C and ^{15}N isotope labels needed for high-resolution solution NMR study. We define the first high-resolution ensemble structure of proIAPP in dodecylphosphocholine (DPC) micelles, using chemical shifts, nuclear Overhauser effect (NOE)

^{1}H distance restraints, RDCs, ^{15}N relaxation rates, and solvent-accessible paramagnetic relaxation enhancement rates (PRE). We used diffusion-ordered spectroscopy (DOSY) measurements to characterize the affinity of proIAPP to DPC micelles, and circular dichroism (CD) to probe membrane environment selectivity. We show that the solution NMR ensemble structure of the protein exchanges between multiple conformers, containing an unfolded Npro segment (residues 1-11) and a mostly helical IAPP region (residues 12-48). Surprisingly, we found that the Cpro sequence (residues 52-67) also adopts helical structure. Previous evidence suggests that Cpro is discarded after processing, and our results indicate a possible structural and functional role for this peptide.

2.2 Results

2.2.1 The Structure of DPC-bound proIAPP is Dynamic

We collected 2D ^{15}N-HSQC NMR spectra for proIAPP in DPC micelles over a pH range of 4.0-7.0. Data collection and analysis was performed at pH 4.5, which provides insight to the late stages of the RSP (pH 5.0-5.9) while minimizing changes in chemical shifts between samples with small variations in pH. The ^{15}N-HSQC showed a total of 68 of 68 observable backbone amide peaks at pH 4.5 (Error! Reference source not found.**A**) and 42 of 68 observable p eaks at pH 7.0. The absence of a significant number of peaks at neutral pH is indicative of elevated hydrogen exchange rates. For this reason, we focused our structural analysis on the low pH condition.

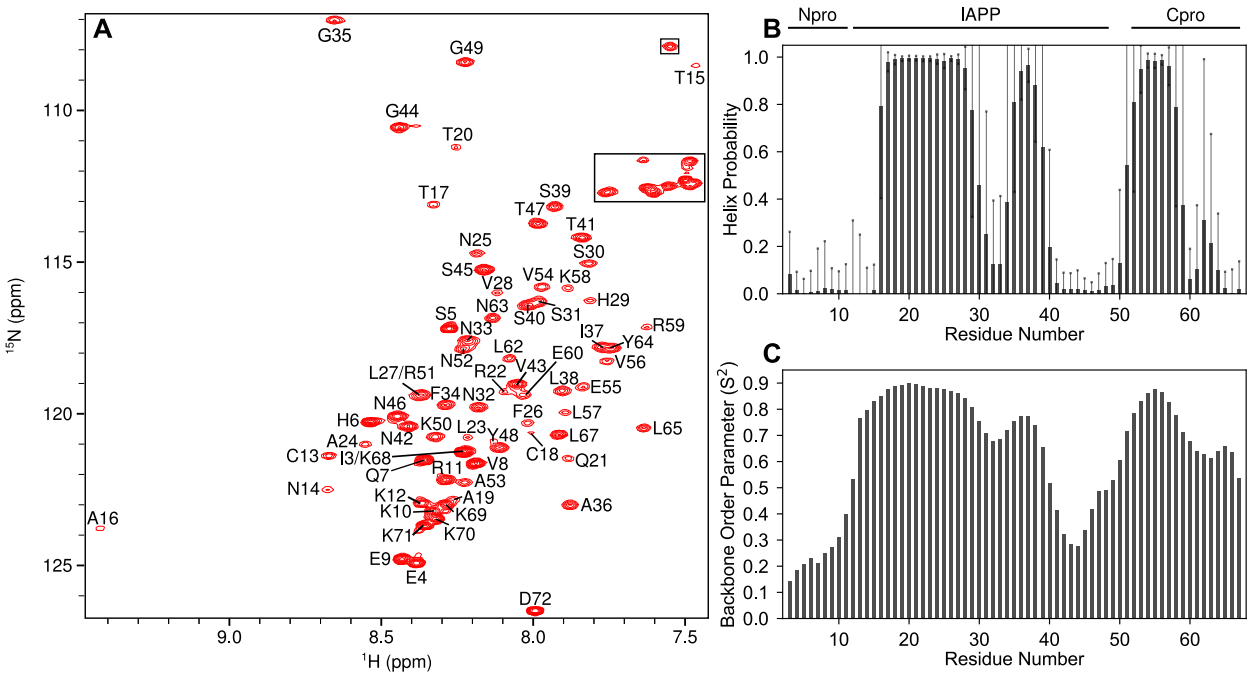


Figure 2.1. ^{15}N-HSQC NMR spectrum and secondary structure assignment from chemical shifts for proIAPP bound to DPC micelles at pH 4.5. (A) The assigned 500-MHz ^{15}N-HSQC spectrum is shown for 1.2 mM ^{13}C,^{15}N-proIAPP in 200 mM DPC, 50 mM NaCl, 30 mM sodium acetate at pH 4.5 with 10% $^{2}H_2O$, and 0.03% NaN_3. Visible Arg, Asn, and Gln side-chain resonances are indicated by the box. **(B)** The helical probability and **(C)** predicted backbone order parameters (S^2) are shown for each residue based on chemical shifts in TALOS+. The prohormone segments are indicated above.

Error! Reference source not found.**A** shows the assigned ^{15}N-HSQC spectrum for p roIAPP bound to DPC micelles at pH 4.5. Chemical shift assignments were identified for 98% of the ^{1}H, ^{13}C and ^{15}N backbone nuclei. ^{13}C resonances for residue T1 could not be obtained through traditional backbone assignment experiments, as it is followed by a proline. A NOESY-^{15}N-HSQC was collected and analyzed in conjunction with an aliphatic (H)CCH-TOCSY to assign 74% of the side-chain resonances. The presence of an oxidized disulfide bond between residues C13-C18 was identified by a downfield C^{β} chemical shift near 40 ppm. The oxidized state of this disulfide was confirmed by the absence of a reaction with Ellman's reagent [29]–[31].

The backbone chemical shift analysis using TALOS+ [32] indicates a mixture of unfolded and helical regions throughout the micelle bound structure (**Error! Reference source n ot found.B**). Residues A16-H29, F34-S39, and R51-K58 have a high probability (>50%) of helical secondary structure. Residues S30-N33, R59, L62, and N63 show some propensity for helical secondary structure (>20%) with a higher degree of uncertainty. The estimated backbone order parameters (S^2) from the chemical shifts follow a similar trend to the secondary structure prediction (**Error! Reference source not found.C**). The N-terminal propeptide segment (T1-R 11) is unfolded in DPC micelles, with average S^2 values below 0.4. The disulfide bond through the first helix, C13-H29, has a high average S^2 that begins to drop off at S30 for a lower, but still ordered, average S^2 for residues S30-S39. A disordered loop from S40-G49 connects the C-terminus of immature IAPP with the C-terminal propeptide segment of N52-L67, which appears fairly rigid.

We used ^{1}H-^{15}N and ^{1}H-^{13}C RDCs to confirm these secondary structure assignments and to accurately define contiguous α-helical segments [33], [34]. Error! Reference source not found. s

hows the ^{1}H-^{15}N RDCs for proIAPP in DPC at pH 4.5 in a 5.4 mm negatively charged stretched acrylamide gel (SAG). We found 4 segments (16-27, 32-39, 51-59 and 62-65) that could be accurately modeled by an α-helical dipolar wave with a period of 3.6 residues per turn [35], [36]. A reduced χ^2 was calculated for each helical dipolar wave to identify the residues included in each helix (see Materials and Methods). The first two helices are found in the amylin sequence of proIAPP. We fit helix 1 (α1, blue) with RDCs from residues 16-27. Inclusion of residues 28-29 more than doubled the reduced χ^2, indicating deviations from a regular α-helical pattern. We fit helix 2 (α2, red) with RDCs from residues 32-39, which gave a larger reduced χ^2 than the other three helices and is consistent with a distorted or dynamic helical structure. Inclusion of S31 in the dipolar wave of α2 doubled the reduced χ^2. The next two helices are found in the Cpro sequence, which is cleaved and removed during amylin processing. We fit helix 3 (α3, green) to residues 51-58. R59 is likely part of this helix as well, but spectral overlap with acrylamide resonances in the aligned IPAP-HSQC precluded the measurement of the ^{1}H-^{15}N RDC for that residue. We fit helix 4 (α4, yellow) to residues 62-65. E60 could not be fit to either the third or fourth helical wave functions. Altogether, proIAPP in DPC micelles is composed of a group of four helices. The first two helices are found in the amylin sequence and the second two helices are found in the Cpro sequence.

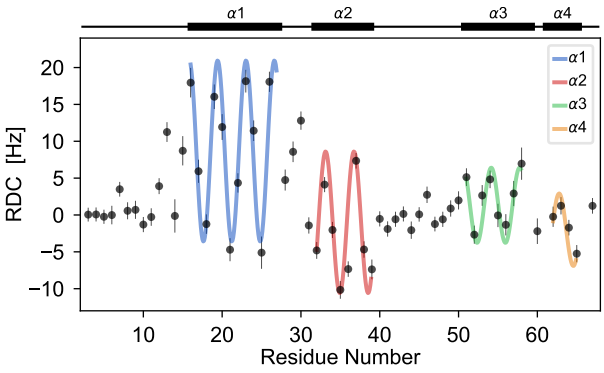


Figure 2.2. Observed ^{1}H-^{15}N RDCs for proIAPP plotted by residue number. The least-squares fits to dipolar wave functions are superimposed in different colors to indicate different α-helical segments [35]. Helices are labeled above the figure. RDC values were obtained from an ^{15}N-IPAP-HSQC [37] collected at 500-MHz on a partially aligned sample composed of 1.2 mM ^{13}C-^{15}N-proIAPP, 200 mM DPC, 30 mM sodium acetate at pH 4.5 with 50 mM NaCl, 10% $^{2}H_2O$, and 0.03% NaN_3 in a 5.4 mm negatively charged SAG [38].

To further investigate the dynamics of proIAPP, we collected ^{15}N relaxation data at 500- and 750-MHz (**Figure 2.3**). Residues N14-T15 show evidence of chemical exchange (R_{ex}) on an intermediate exchange timescale (**Figure 2.3B**), which is consistent with isomerization of the C13-C18 disulfide bond [39]. Fast-ModelFree [40] estimated the R_{ex} rates to be 9.0 and 7.4 s^{-1} for N14 and T15 H^N, respectively. Residues A53-K58 of α3 also show evidence of R_{ex} with rates ranging from 1.8 to 4.8 s^{-1}. α2 is more dynamic on a ps-ns timescale with a greater degree of motion than other regions with structure. The relaxation rates of the residues agree with the helical discontinuity between H29 and S30, with helical residues S30-S39 experiencing larger amplitude rotational motions. Likewise, the C-terminal α4 shows elevated motion in comparison to the helix that precedes it. Fast-ModelFree [40] from the NMRbox software suite [41] was used to estimate the backbone generalized order parameters (**Figure 2.3D**). With the exception of helix α3, the order parameters are fairly similar for residues within a given helix. Helix α3 displays a gradual increase in rigidity from the N-terminus to the C-terminus, consistent with partial fraying at the N-terminus of the helix.

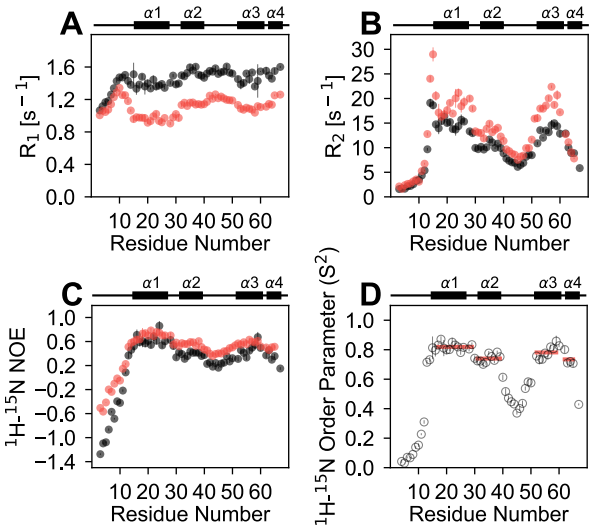


Figure 2.3. ^{15}N relaxation data for ^{13}C,^{15}N-proIAPP in DPC micelles. Data were collected at 500-MHz (black) and 750-MHz (red). The **(A)** R_1, **(B)** R_2 relaxation rates and the **(C)** ^{15}N-^{1}{H} heteronuclear NOE ratios are shown. **(D)** The fit model-free order parameters shown were derived from the Lipari-Szabo formalism using Fast-ModelFree [40], [42]. The average order parameters for each helix are indicated by red lines. The helices are labeled above each plot.

The ^{15}N relaxation analysis yielded a rotational correlation time (tumbling time) of 10.10 ± 0.08 ns. This tumbling time is consistent with a 29.7 kDa globular molecular weight, and it is equivalent to a DPC micelle with a single (monomeric) proIAPP [43]–[45]. No intermolecular NOEs were identified to indicate the presence of dimers or higher order oligomers, and the experimental protein:micelle ratio was 0.27 to minimize self-association of the protein over the course of data collection.

2.2.2 Membrane Insertion and Orientation

We collected $^{1}H^{N}$ PREs with gadodiamide to measure solvent accessibility of the proIAPP backbone and membrane insertion of the helices. PREs were measured using a {^{1}H-T_2}-^{15}N-HSQC experiment in the absence and presence of 2 mM gadodiamide. The PREs are plotted (**Figure 2.4**) from the difference in these relaxations rates. Residues I3-N14, comprised of Npro and the beginning of the disulfide bond loop region, have a high solvent accessibility, as demonstrated by the elevated PRE values. The large increase in the PRE for C18 indicates the disulfide bond is anterior to the lipid-binding interface. Amphipathic helical profiles are present from residues A16-V28 and N32-S39, with solvent exposed amides from H29-S31. The unstructured loop (S40-K50) that tethers IAPP to Cpro is highly accessible to solvent. Another amphipathic helix is observed from R51-R59, and the final helical turn contains equally solvent accessible L62-Y64 that, taken with relaxation dynamics information, indicates another mobile segment of this protein.

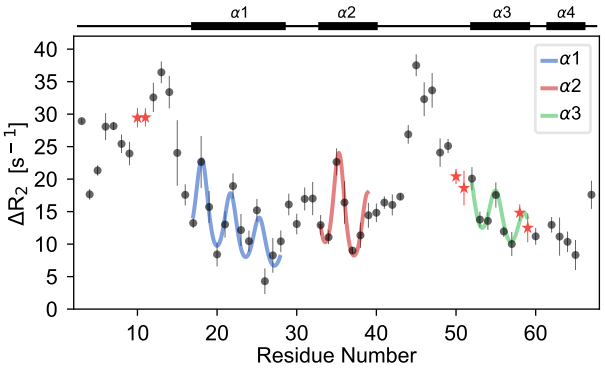


Figure 2.4. Solvent accessibility and membrane insertion profile for proIAPP in DPC micelles. Amide proton PREs (ΔR_2) for ^{15}N-proIAPP bound to DPC micelles at pH 4.5 are plotted as a function of residue number. The difference in R_2 rates were calculated by subtracting the reference ^{1}H R_2 rate from the rate in the presence of 2 mM gadodiamide. Helices are labeled above the plot, and the dibasic PC recognition sites are indicated by red stars. The colored lines represent the fits of helical segments to paramagnetic relaxation waves [46]. F26 exhibited the greatest deviation from the fit, with a much lower PRE than expected. α4 (P61-L65) had too few points in the fit to be accurately modeled. The NMR sample contained 1.0 mM ^{15}N-proIAPP, 200 mM DPC, 30 mM sodium acetate, pH 4.5, 50 mM NaCl, 10% $^{2}H_2O$, and 0.03% NaN_3.

Paramagnetic relaxation waves were fit to amphipathic helices by a least-squares regression analysis of the observed PREs (**Figure 2.4**), as previously described [46]. The fit parameters include the tilt angle of the helix (τ) with respect to the micelle, the insertion depth (A) and the azimuth angle (ρ) of the first residue in the helix. The best-fit values for these parameters represent average orientations and depths of each helix on the membrane. However, the protein likely adopts a range of orientations and depths in the micelle.

Values of τ that approach 0° or 180° indicate a helical orientation parallel to the micelle surface, whereas a τ of 90° indicates a transmembrane helix. The PRE wave fits show that α1 and α2 are tilted nearly in the plane of the micelle surface, with τ values of 13 ± 7° and 11 ± 9°, respectively. α3 lies tilted slightly into the membrane to a greater extent, with τ = 17 ± 7°. The positive value of τ indicates that the C-term of each helix is more micelle buried than the N-term.

The insertion depth and ρ of the first residue provides insight on the overall depth of each helix in the membrane. Values of ρ near 180° indicate the first residue in the helix is on the hydrophobic face, while values of ρ near 0° or 360° indicate a solvent facing orientation of the first residue. A membrane depth of 8.5 ± 2.1 Å and ρ of 87 ± 9° were determined for A16, which is buried into the headgroup region and pointing towards the interface between the micelle and solvent. N32 H^N has a depth of 2.2 ± 0.4 Å with a ρ of 167 ± 21°, indicating this residue is almost completely facing the micelle. When also comparing τ of α1 and α2, this information indicates that the second helix is slightly more solvent exposed on average than the first. The amide of the first residue of the third helix, R51, has a depth of <1 Å and a ρ of 359 ± 16°, indicating a completely solvent-facing orientation.

Fitting ρ to each of the PRE waves also maps the hydrophobic and hydrophilic faces of each amphipathic helix, where wave crests indicate a $^{1}H^{N}$ atom facing the solvent and troughs indicate a $^{1}H^{N}$ atom facing the micelle interior. The amide protons for T17, T20, L23, A24, F26, and L27 in α1 face the micelle interior. Likewise, the amide protons for N32, F34, I37, and L38 of helix α2, and A53, V54, V56 and L57 for helix α3 face the micelle interior.

2.2.3 proIAPP NMR Structures in DPC

A series of ensemble refinements were conducted on the proIAPP structure in DPC micelles: a 1-conformer, a 2-conformer, and a 3-conformer ensemble refinement. A statistical analysis was performed on the agreement of the experimental RDCs to the simulated RDCs of the lowest energy structure(s) (**Figure 2.5**). The RDC restraints are not consistent with a 1-conformer model, as indicated by a reduced χ^2 of 18.8. For this reason, we present the 2-conformer ensemble refinement, which produces good agreement to the experimental RDCs (reduced $\chi^2 \approx 1.0$; marked by an asterisk in **Figure 2.5**). A one-sided F-test at 95% confidence ($\alpha = 0.05$) passed for the 1- vs. 2-conformer refinements and failed for the 2- vs. 3-conformer refinements. Although proIAPP in DPC micelles likely adopts a much broader range of structures, the 2-conformer ensemble represents the minimum number of structures consistent with our data. This approach has been applied previously to systems with a distribution of structural conformations [47]–[49]. The complete 2-conformer ensemble refinement statistics are listed in **Table 2.1**.

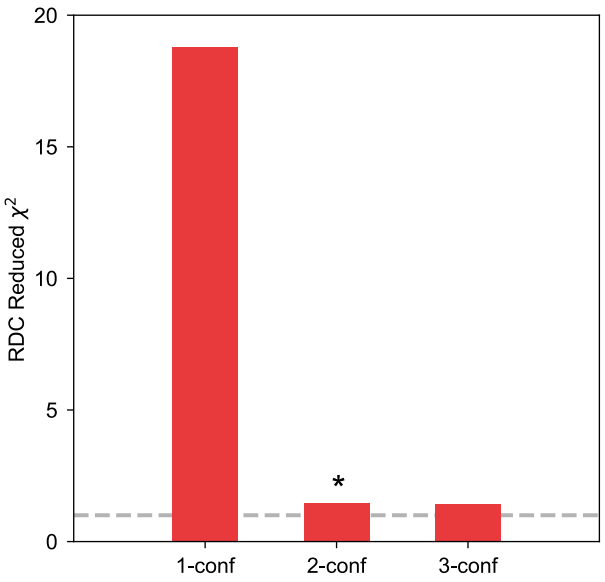


Figure 2.5. Reduced χ^2 statistical comparison of experimental and simulated RDCs for a 1-, 2-, and 3-member ensemble refinement. A χ^2 of 1.0 is shown by a broken gray line. Values significantly higher than 1.0 indicate a poor agreement between simulated RDCs and experimental RDCs relative to the experimental error. Values close to 1.0 indicate a good agreement, and values below 1.0 indicate overfitting of the experimental RDC restraints.

Table 2.1. Structural statistics for the 20 lowest-energy sets of structures of a 2-conformer refinement of proIAPP in DPC at pH 4.5.

	Open[a]	Bent[b]
Dihedral restraints		
Total	189	
Violations	0	
NOE distance restraints		
Total	457	
Intraresidue, $\|i = j\|$	116	
Sequential, $\|i - j\| = 1$	211	
Medium range, $1 < \|i - j\| < 5$	126	
Long range, $\|i - j\| \geq 5$	4	
Restraints per residue[c]	10.6	
Violations > 0.5 Å	0.8	
RDC restraints		
Total	149	
(−)SAG_{NH} experimental RMS, Hz (restraints)	1.6 (38)	
(−)SAG_{CH} experimental RMS, Hz (restraints)	9.7 (34)	
(∅)SAG_{NH} experimental RMS, Hz (restraints)	1.1 (30)	
(+)SAG_{NH} experimental RMS, Hz (restraints)	1.4 (28)	
(+)SAG_{CH} experimental RMS, Hz (restraints)	4.1 (19)	
Restraints per residue[c]	3.5	
(−)SAG_{NH} fit RMS, Hz (violations)	1.5 (0.2)	
(−)SAG_{CH} fit RMS, Hz (violations)	9.4 (0.1)	
(∅)SAG_{NH} fit RMS, Hz (violations)	0.9 (0)	
(+)SAG_{NH} fit RMS, Hz (violations)	1.4 (0.1)	
(+)SAG_{CH} fit RMS, Hz (violations)	3.9 (0)	
RMSD of 20 structures		
RMSD all backbone atoms, Å	3.6	2.6
RMSD backbone atom res. 13-39, 50-65 Å	1.7	1.3
Close contacts and deviations from ideal geometry		
Number of close contacts[d]	0	0
RMSD for bond angles	1.4 °	1.4 °
RMSD for bond lengths, Å	0.009	0.009
Ramachandran plot summary from Procheck		
Most favored regions	93.2%	92.3%
Additionally allowed regions	6.5%	7.7%
Generously allowed regions	0.4%	0%
Disallowed regions	0%	0%

[a]PDB ID: 6UCJ
[b]PDB ID: 6UCK
[c]NOE and RDC restraints included for 43/67 residues (order parameters > 0.6)
[d]Within 1.6 Å for H atoms, 2.2 Å for heavy atoms

Table 2.2 (continued). Structural statistics for the 20 lowest-energy sets of structures of a 2-conformer refinement of proIAPP in DPC at pH 4.5.

	Open[a]	Bent[b]
Ramachandran plot statistics from Richardson's laboratory Molprobity		
Most favored regions	98.3%	97.8%
Allowed regions	1.5%	1.6%
Disallowed regions	0.2%	0.6%
Structure quality factor Z-scores		
Verify3D	-8.03	-8.03
ProsaII	-4.38	-4.01
Procheck (ϕ-ψ)	1.65	1.10
Procheck (all)	0.24	-0.41
MolProbity clash score	1.22	0.97

[a]PDB ID: 6UCJ
[b]PDB ID: 6UCK

The NMR structures were solved with a combination of chemical shift dihedral restraints, NOEs, and RDCs. Dihedral angle restraints were obtained from backbone chemical shifts using TALOS+ [32]. NOE distance restraints were generated by calibrating H^N-H^N and H^N-H^α NOEs from residues A16-L27 to known distances of an idealized α-helix [50]. RDC restraints were used directly from measured experimental values. Simulated annealing starting from an extended structure of proIAPP was performed with the inclusion of dihedral, NOE, and RDC restraints in XPLOR-NIH [51], [52]. The final structures contain 3 distinct helices with a short helix (α4) at the C-terminus, following a proline residue (**Figure 2.6A**). The proIAPP structure contains an extended and dynamic Npro region, followed by helical regions of IAPP and Cpro that are separated by a 11-residue loop (S40-K50). Helical stretches are present from A16-V28 (α1), N32-S39 (α2), R51-R59 (α3), and P61-L65 (α4) for a total of 47.2% helical content in the protein.

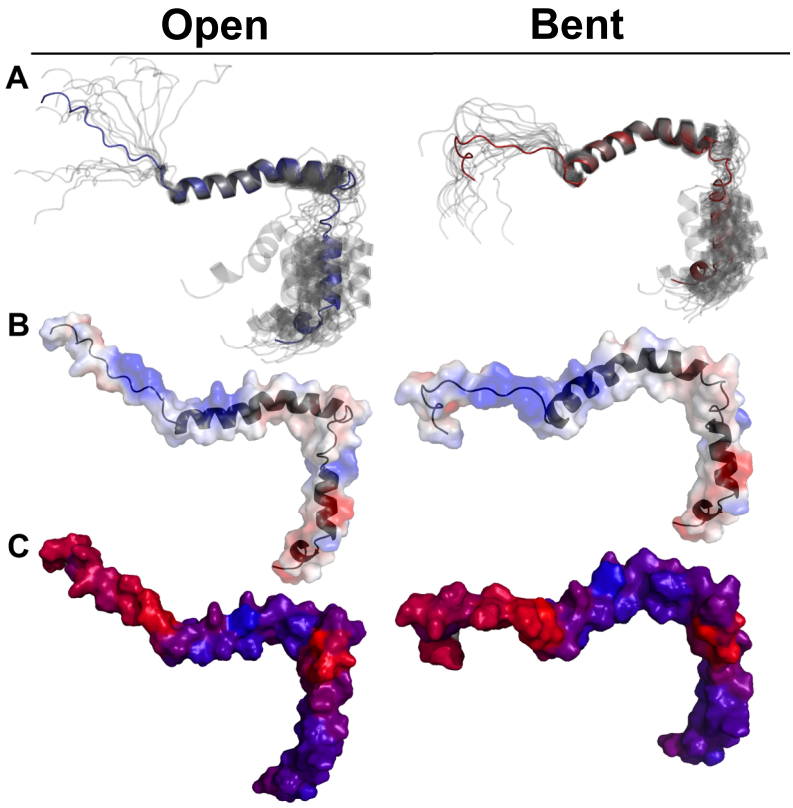


Figure 2.6. NMR structures for the 2-conformer ensemble refinement of proIAPP bound to DPC micelles at pH 4.5. (A) The lowest energy refined structure (opaque structure) is superimposed with the other 19 lowest energy structures (semi-transparent structures). **(B)** The electrostatic potential surface (semi-transparent) is shown with the lowest energy structure. Surfaces with a negative, positive or neutral charge are colored in red, blue or white, respectively. The potential was calculated using Adaptive Poisson-Boltzmann solver (APBS)[53] at a pH 4.5. **(C)** The PREs from Fig. 4 are plotted on the surface representation of the lowest energy structure. PREs range from low values in blue (PRE = 5 s^{-1}) to high values in red (PRE = 37 s^{-1}). Regions in blue represent surfaces occluded from solvent, and regions in red represent surfaces accessible to solvent.

The first two helices are similar, yet distinct, to the solution structure of IAPP in SDS micelles at pH 4.6 (**PDB ID: 2KB8**) [27]. A kink near H29-S31 connects the first, more rigid helix to the second, more dynamic helix. The average angle between α1 and α2 is defined by the RDCs, and it varied considerably between the 2 ensemble members. The first (open) conformer angle ranged from 162°-170°, while the second (bent) conformer angle ranged from 119°-127° for the 20 lowest energy structures. IAPP possesses both membrane curvature-sensing and curvature-inducing properties [54], and it is possible that the observed conformational exchange between these helices drives the binding of IAPP to positively curved membranes.

The hydrophobic faces of α1 and α2 are aligned towards each other (**Figure 2.6B**), lining the interior of the angle between them, indicating that the IAPP component of proIAPP forms an amphipathic interaction at the surface of the DPC micelle. This amphipathic alignment is notably absent from previous IAPP structures, which did not use RDCs to accurately define the orientation and phase of the 2 helices [27], [28]. The pH 7.3 IAPP structure (**PDB ID: 2L86**) from previous reports presents nearly a right angle between α1 and α2, which results in the Phe side-chain of either α1 or α2 pointing away from the membrane and into the solvent. This orientation is poorly defined, as it is held together by a single long-range NOE between the aromatic protons of F26 and N32 H^N. This NOE was not observed in our ^{15}N-filtered NOESY-HSQC of proIAPP in DPC micelles. Instead, we found an NOE cross-peak between the aromatic protons of F26 and S31 H^N. In addition to the inclusion of RDCs in our refinement, the IAPP segment of our structure may differ from previous structures due to the micelle environment. The membrane surface electrostatics of a DPC micelle are closer to a native membrane environment, which is significantly different from the negatively charged SDS micelles used in previous studies.

The C-terminal helices, α3 and α4, comprise most of the proIAPP Cpro segment, which is cleaved during the first step of IAPP processing. Interestingly, α3 begins at the dibasic PC1/3 cut site and spans the scissile peptide bond between R51-N52. The α3 helix terminates at a proline residue (P61), and a bifurcated hydrogen bond between P61 carbonyl and the H^N of both Y64 and L65 produces the single turn of the α4 helix. Another proline terminates the α4 helix from P66 to L67, leading to a short unstructured segment at the C-terminal end.

The entire proIAPP structure appears to wrap around the DPC micelle. In both structures, the hydrophobic face of α3 aligns with the hydrophobic faces of α1/α2. The distances measured between the most buried residues of α1/α2 and α3 ranged from 23.5-30.0 Å in both the open and the bent 2-conformer ensemble structures. A DPC micelle is a prolate structure with a long-axis and two short-axes [44], [45]. The range of distances observed in the 2-conformer structures is slightly smaller than the reported 32-33 Å short-axis diameter of a DPC micelle.

To better visualize the solvent accessibility profile of the protein structure in the presence of a membrane, we used PyMOL [55] to illustrate the measured PREs directly on the refined structures (**Figure 2.6C**). Regions with high solvent accessibility (high PREs) are labeled in red, and regions with low solvent accessibility (low PREs) are labeled in blue. The N-terminal segment (I3-R11) and random-coil region (S40-K50) connecting α2 and α3 have the highest solvent accessibility. The average amphipathic helical phase determined from the PRE data correlates very well with the helical pitch defined by the RDC restraints.

2.2.4 Membrane Charge and pH Mediate proIAPP Affinity

We used CD to identify changes in the proIAPP secondary structure in the presence of various membrane environments at acidic and neutral pH (**Figure 2.7**). In the absence of lipid, unfolded features dominate the CD spectra for proIAPP, and the protein is intrinsically

disordered. In the presence of detergent micelles and bicelles, the protein adopts a mostly α-helical structure at both pH values. The helical structure is indicated by the mean residue ellipticity ([θ]) minima near 208 and 222 nm (**Figure 2.7A-B**).

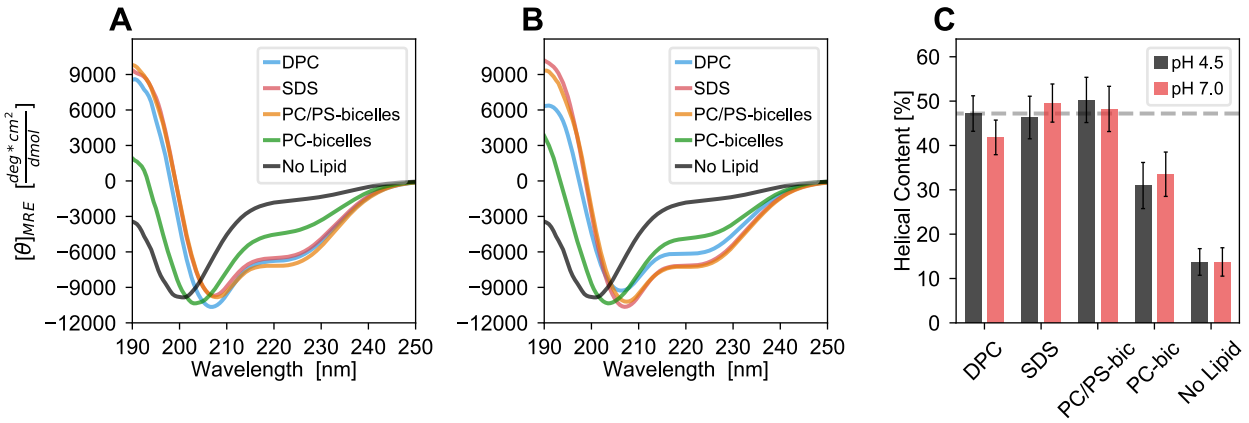


Figure 2.7. Circular dichroism analysis of proIAPP in the presence and absence of detergent micelles and bicelles at acidic and neutral pH. (A) Mean residue ellipticity curves of proIAPP in the absence of lipid (black) and in the presence DPC micelles (blue), SDS micelles (red), DMPS/DMPC/DHPC bicelles (orange, q=0.5 and [DMPS]:[DMPC]=1:3), and DMPC/DHPC bicelles (green, q=0.5), at pH 4.5 and **(B)** pH 7.0. **(C)** Helical content calculated from $[\theta]_{222}$ normalized to the NMR structure of proIAPP in DPC micelles, at pH 4.5 (black) and pH 7.0 (red). The NMR structure helical content is denoted by a dashed line at 47.2%. Error bars represent the propagated error of the standard deviations of sample measurements and protein concentrations.

Ellipticity minima at 208 and 222 nm can provide an excellent measure of relative helical secondary structure between samples [56], [57]. However, even in highly concentrated protein samples, these minima are obscured in bicelle samples because bicelles scatter light and reduce the signal of the far-UV range below 230 nm [58]–[60]. Scattering at lower wavelengths can introduce distortions that obfuscate the interpretation of the data at 208 nm. We therefore carried out our analysis by comparing [θ] at 222 nm ($[\theta]_{222}$) for each condition. Since our proIAPP structure showed a 47.2% helical content at pH 4.5, we calculated the helical content of other lipid environments relative to this sample. The calculated helicities for each membrane environment are presented in **Figure 2.7C**.

In the presence of membranes at pH 4.5, the proIAPP helicity is 47 ± 4%, 46 ± 5%, 50 ± 5%, and 31 ± 5% for DPC, SDS, 3:1 DMPC/DMPS-bicelles, and DMPC-bicelles, respectively. At pH 7.0, proIAPP helicity is 42 ± 4%, 50 ± 4%, 47 ± 5%, and 34 ± 5% for DPC, SDS, DMPC/DMPS-bicelles, and DMPC-bicelles respectively. The helical content of the protein in the absence of lipid is 14 ± 3% and 14 ± 3% at pH 4.5 at pH 7.0, respectively, which agrees with a previous report on the transiently helical α1 in IAPP and the presence of the C2-C7 (C13-C18 in proIAPP) disulfide bond [61].

At both pH values, proIAPP is less helical in zwitterionic (neutral) DMPC/DHPC bicelles than DPC micelles, which could be attributed to differences in lipid headgroup packing and in curvature of the lipid aggregate. Doping these bicelles with negative charge, using a 3:1 ratio of DMPC:DMPS in the bicelles, recovers the helical content. Altogether, these results indicate that proIAPP adopts a maximum *ca.* 50% helical secondary structure in the presence of positive curvature lipid aggregates (DPC, SDS), lipid aggregates with negative charge (DMPC/DMPS/DHPC bicelles and SDS) or both factors together.

With the exception of DPC samples, the proIAPP structure remains mostly unchanged between low and neutral pH. In DPC, the helicity of 47% at low pH decreased to 42% at neutral pH. The reduced helicity and elevated hydrogen exchange rates from the ^{15}N-HSQC spectra suggest that proIAPP is partially unbound from DPC micelles at neutral pH.

We measured the population of proIAPP molecules bound to DPC micelles at neutral and low pH using NMR DOSY experiments (**Figure 2.8**). In this experiment, we measured the translational diffusion rate of free proIAPP in solution, of the free DPC micelle, and of the proIAPP-micelle complex at pH 4.5 and 7.0. The aqueous proIAPP sample has a low solubility without the presence of micelles, so we used His_6-proIAPP to conduct these experiments. The CD spectra and the chemical shifts in DPC micelles are identical for His_6-proIAPP and proIAPP [data not shown]. We quantitated the concentration of free His_6-proIAPP using NMR [62] to calculate the bound percentage.

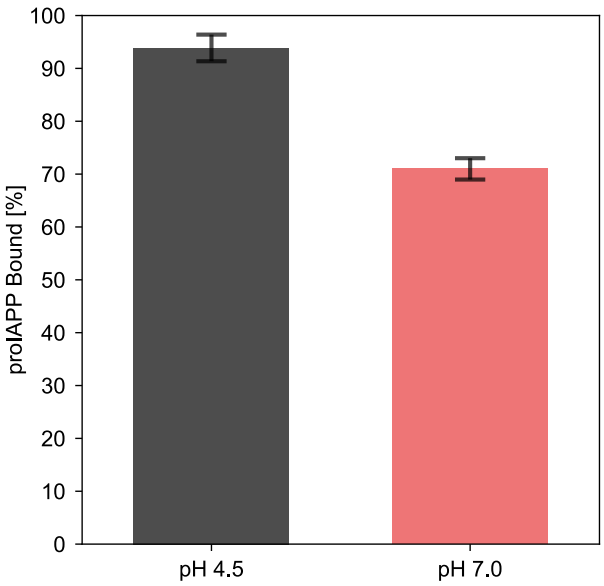


Figure 2.8. Micelle-bound population of proIAPP at pH 4.5 and 7.0 as determined by ^{1}H DOSY. DOSY samples were prepared with 0.7 mM natural abundance His_6-proIAPP in the absence of DPC, and 1.5 mM natural abundance His_6-proIAPP in the presence of DPC micelles. Samples contained 200 mM DPC, 50 mM NaCl, 10% $^{2}H_2O$, 0.03% NaN_3, and 30 mM sodium acetate. The pH 7.0 sample was prepared by titrating the pH 4.5 sample with Tris base.

At pH 4.5, His_6-proIAPP has a bound population of 94 ± 3%, whereas at pH 7.0, His_6-proIAPP has a bound population of 71 ± 2% to DPC micelles. These results indicate that proIAPP binds less tightly to membranes at neutral pH, and it is consistent with the reduced helicity and elevated hydrogen exchange observed in the ^{15}N-HSQC spectra.

Even though hydrogen exchange precludes a detailed structural analysis for proIAPP in DPC at pH 7.0, we measured chemical shifts for non-exchangeable groups and visible peaks for this sample. Backbone shifts could only be assigned to 62% completion for this condition. A comparison of the TALOS+[32] helical probabilities for proIAPP in DPC at pH 7.0 and pH 4.5 is shown in **Figure 2.9**. At neutral pH, the backbone helicity increases for residues H29-N32, which comprise the turn between α1 and α2. There is a slight decrease in predicted helicity for α2 overall, but the largest change at neutral pH appears to stem from a loss of helicity at the N-terminus of the α3 helix. The helical probability for α4 is nearly identical at both pH values, indicating that this short helical stretch likely remains folded.

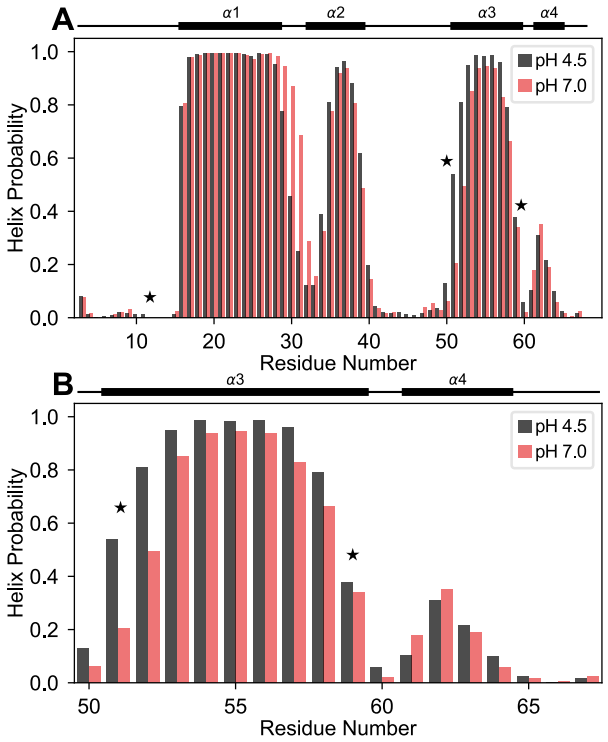


Figure 2.9. Chemical shift helix probabilities between pH 4.5 and pH 7.0 for proIAPP in DPC micelles. (A) Helix probability comparison for all proIAPP residues. An increase in predicted helicity is observed for residues between α1 and α2, and a decrease in predicted helicity is observed for the residues in α2 and α3. **(B)** Helix probability comparison for the Cpro residues from (A). The greatest change occurs near the putative PC1/3 cleavage site, indicated by the first star. Helices from the pH 4.5 structure are shown above the plots, and PC cleavage sites are indicated with stars. Helix predictions were estimated from backbone chemical shifts using TALOS+ [32].

2.2.5 Cpro folds independently of proIAPP

We used CD to conduct a secondary structure analysis of the Cpro peptide at pH 4.5 and 6.1, in the presence and absence of DPC micelles (**Figure 2.10A**). In the absence of lipid, Cpro displays unfolded or coil characteristics at both pH values. At pH 4.5 in DPC micelles, Cpro appears predominantly helical. However at pH 6.1 in DPC micelles, near the pH of PC1/3 cleavage, the peptide has a mixture of helical and unfolded characteristics. A ^{1}H chemical shift index analysis on Cpro in DPC micelles at pH 4.5 suggests that the peptide adopts the same fold after separation from IAPP, despite a slight truncation of the N-terminus of α3 (**Figure 2.10B**). Collectively these results indicate that Cpro can bind to membranes and fold independently of full-length proIAPP.

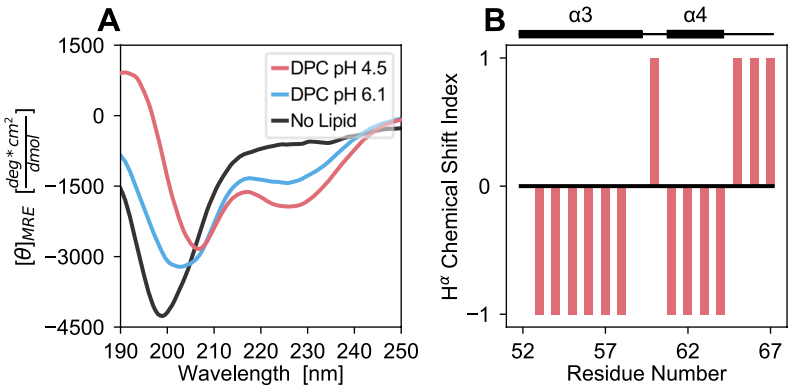


Figure 2.10. Secondary structure analysis of Cpro in DPC micelles. (A) Circular dichroism mean residue ellipticity curves of Cpro in the absence of lipid (black) and the presence of DPC micelles at pH 4.5 (red) and pH 6.1 (blue). **(B)** $^{1}H^{\alpha}$ chemical shift index of Cpro at pH 4.5 in DPC micelles. Index values were calculated by subtracting the index chemical shift from the $^{1}H^{\alpha}$ chemical shift for each residue, as previously described [63]. Continuous stretches of values of -1 indicate α-helical secondary structure, while stretches of +1 suggest β-sheet secondary structure. Values of 0 and isolated values of -1 or +1 suggest coiled regions. Helix and residue numbering are in the context of full-length proIAPP, and helices are indicated above the plot. The NMR sample contained 1.0 mM natural abundance Cpro, 200 mM ^{2}H-DPC, 30 mM sodium acetate, pH 4.5, 50 mM NaCl, 10% $^{2}H_2O$, and 0.03% NaN_3.

2.3 Discussion

2.3.1 Ensemble structures, dynamics and measurement precision

We refined the structures of proIAPP in a micelle environment and found that a minimum of 2 conformers are consistent with our data. The ensemble represents the dynamic distribution of structures, and proIAPP likely visits a larger number of structures in DPC micelles. The dynamic nature of this system and similar systems, like the previously reported IAPP structures, can be concealed without a comprehensive structural dataset including NOEs, dihedral restrains and RDCs from multiple alignments.

In the context of our other measurements, including the relaxation, PRE and CD data, the reported values are *averages* of the ensemble structures. The uncertainties associated with these measurements represent the precision of the measurement rather than the distribution of their values for the different structures. For example, our PRE data fits give a precision of 7-9º for the average membrane insertion angles of helices. The helices likely adopt a larger range of membranes orientations than the uncertainty of our fits. In a similar system, the influenza hemagglutinin fusion peptide (HAfp), we previously showed than the helices can 'wobble' by up to 20º within a membrane environment [64].

2.3.2 Implications for proIAPP Processing

The distribution of structures does not appear to have functional significance in IAPP receptor binding. Rather the significance of the structural fluctuations between the helices lies in the interactions with lipids during proIAPP processing. Our findings provide insight to the role of membranes in the proper processing and storing of IAPP in the RSP.

IAPP secretion is a tightly regulated process [65]. The mature peptide is produced from proIAPP by a PC1/3 cleavage of the R51-N52 peptide bond, a subsequent cut by PC2 between R11-K12, removal of K50 and R51 by CPE, and conversion of the C-terminal G49-COOH to Y48-$CONH_2$ by PAM [12], [13], [15]. PC1/3 activity takes place in the trans-Golgi apparatus between pH 6.0-6.4, while the subsequent steps occur in the secretory granules between pH 5.7-5.0 [14], [66]. Although proIAPP contains three dibasic PC1/3 and PC2 recognition sites (**Figure 2.11A**), cleavage only occurs at two sites in cells: the R11-K12 and R51-N52 peptide bonds. The two cleavages produce three products: Npro, immature IAPP, and Cpro [12], [14], [15], [67].

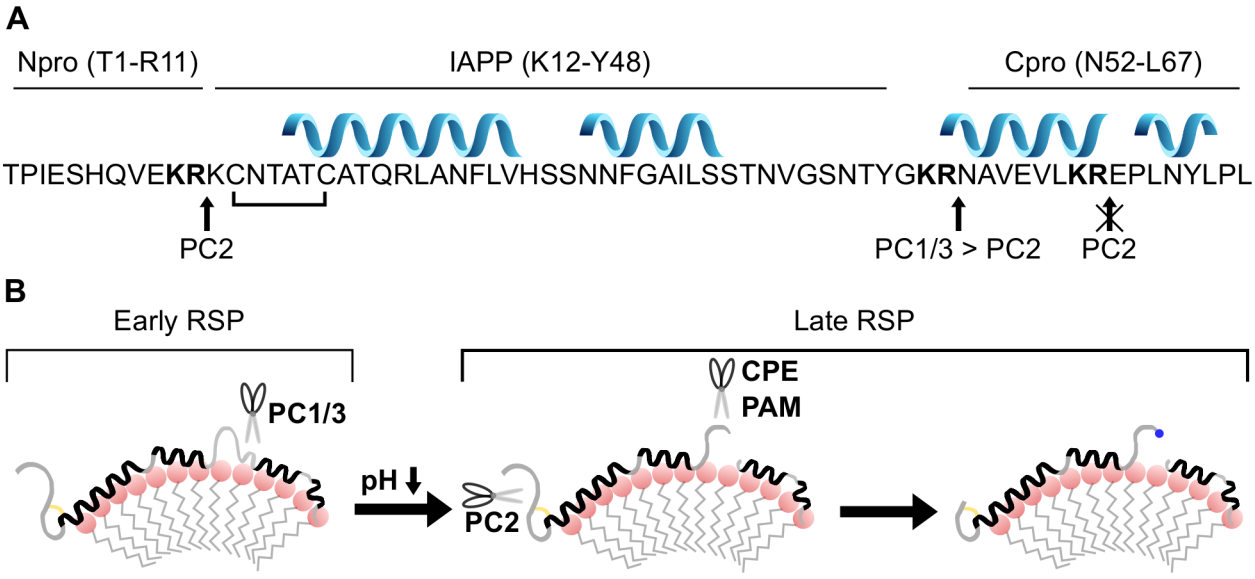


Figure 2.11. Structural implications of proIAPP processing. (A) Primary and secondary structure diagram of proIAPP bound to DPC micelles at pH 4.5. The C13-C18 disulfide bond is indicated by a bracket. Prohormone Convertase 1/3 and 2 dibasic enzymatic recognition sites are boldened with cleavage positions indicated by arrows and proIAPP processing products indicated above the sequence. The discontinuity between the first and second amylin helix is shown as a break between H29 and S31. The R51-Y64 C-terminal helical segment is shown with loss of helicity at E60 due to the proline break at P61. **(B)** Membrane-mediated proIAPP processing diagram. The secondary structure is indicated with helices drawn in black and random coil or loop regions drawn in gray. The C13-C18 disulfide bond is drawn in yellow. Enzymes are represented by scissors, and the C-terminal amide is represented as a blue circle. Stages of the regulated secretory pathway are indicated above the plot.

Based on studies with proinsulin, PC1/3 and PC2 cleave substrate peptides with a high propensity for random-coil structure at the dibasic cut sites [68][69]. Proinsulin has a random-coil structure at the PC1/3 site [70], similar to the proIAPP PC2 site (R11-K12). Likewise, the proinsulin PC2 junction has a random-coil structure with transient helical states, reminiscent of the proIAPP PC1/3 site (R51-N52). These results indicate that PC1/3 and PC2 recognize cleavage sites that are flexible.

Our chemical shift data indicates that the proIAPP sequence immediately before and after the canonical PC2 cut site (R11-K12) is always in a random-coil configuration, regardless of pH or the presence of lipids. Additionally, our PRE data show the N-terminal dibasic site as the most solvent accessible of the three possible sites (**Figure 2.4**, red stars), with the disulfide bond geometry of α1 presenting this recognition site away from the membrane interface. The flexibility and accessibility of this site are consistent with its cleavage *in vivo*.

We found the putative PC1/3 site (R51-N52) exists primarily as a membrane-bound, rigid α-helical conformation in our pH 4.5 structure. Fast-ModelFree [40] analysis of our ^{15}N relaxation data points to chemical exchange within α3 and fraying of the PC1/3 cut site compared to the rest of the helix (**Figure 2.3D**). Our chemical shift data (**Figure 2.9**) show that the population of the unfolded species at this site increases appreciably at neutral pH. Consequently, this site is amenable to cleavage at neutral pH, which is consistent with its observed *in vivo* cleavage in the early RSP [14]. The early RSP has a pH range of 6.0-7.2, in contrast to the pH of 5.0-5.9 for the late RSP [66].

The third and non-canonical PC cleavage site (R59-E60) is located in the middle of the Cpro sequence. This site is not cleaved *in vivo*. Our structure shows that this site terminates α3, and the α-helical structure at this site would preclude its cleavage in cells. To our knowledge,

only one research report described modest *in vitro* cleavage at the R59-E60 junction [71]. This study used either CHAPS with a lipid concentration below the critical micelle concentration or Brij-35 micelles which, according to our CD measurements, do not interact with and fold proIAPP (data not shown). Consequently, cleavage at the E59-E60 junction appears to occur only when proIAPP behaves as an intrinsically disordered protein in solution. When interacting with membranes, this site has a secondary structure, and it is unavailable for cleavage. Altogether, our structure is measured in non-native DPC micelles, yet it shows how secondary structure changes could mediate proIAPP cleavage and processing *in vivo*.

A summary diagram of the potential membrane-mediated IAPP maturation in the RSP is provided in **Figure 2.11B**. In the early RSP, proIAPP binds to membranes as four helical units separated by loops and coils. Helical fraying at the N-terminus of α3 allows for cleavage by PC1/3 at the R51-N52 peptide bond in the trans-Golgi network. Immature IAPP and Cpro are packaged into secretory vesicles where PC2 cleaves the unstructured R11-K12 peptide bond, CPE removes K50 and R51 from the C-term of immature IAPP, and PAM converts G49 into a C-terminal amide. Acidification within the secretory granules keeps IAPP and Cpro in a mostly membrane-bound conformation, thus protecting Cpro from PC2 cleavage at the R59-E60 peptide bond. The peptides remain membrane-bound until a hormone secretion event raises the local pH and rearranges the membrane lipid composition, promoting disassociation and release into the extracellular space.

2.3.3 Similarities to other calcitonin family hormones

Members of the calcitonin family of hormones share numerous traits. Each is expressed as a prohormone and later converted into the bioactive form, containing a short intramolecular disulfide loop of either 4 or 5 residues, and a C-terminal amide. Interestingly, two of the five prohormones in this family, pro-calcitonin and pro-adrenomedullin, contain a second bioactive peptide hormone that is released during normal processing, called katacalcin and pro-adrenomedullin N-terminal 20 peptide (PAMP), respectively [72]–[74]. In both systems, the processing of the prohormone molecule ensures co-secretion of two bioactive peptides. As discussed in the previous section, our structural findings indicate a likelihood for Cpro protection from proteolytic cleavage in the RSP. If the human genome designed this propeptide to survive for secretion, then future studies on the physiological relevance of Cpro may be warranted.

2.4 Concluding Remarks

In summary, we solved the solution NMR structure and molecular dynamics of proIAPP bound to DPC micelles. We showed that the protein interacts with the micelle as four helical segments without tertiary structure: α1-2 from mature IAPP, and α3-4 in the Cpro segment. We mapped the proIAPP solvent accessibility in DPC micelles using aqueous paramagnetic probes, and we identified an amphipathic profile for these helices to shed light on the accessibility of the PC1/3 and PC2 recognition sites in the presence of membranes. CD and DOSY experiments collectively indicate a heightened membrane affinity with positive membrane curvature, negative surface charge, and acidic pH. Our results support a relationship between proIAPP processing and its structure in a membrane environment, and the possibility that the Cpro peptide survives for co-secretion with IAPP.

2.5 Materials and Methods

2.5.1 Protein Expression

The proIAPP construct was subcloned in the pET15b vector with NdeI and BamHI sites. The cDNA was synthesized (Genscript) with an N-terminal Factor Xa (FXa) cleavage site, human proIAPP (UniProt Protein Database P10997), and a C-terminal tetra-lysine aspartate (K_4D) solubility tail. The resulting plasmid encodes a 97-residue protein (10.77 kDa) that includes an N-terminal hexa-histidine (His_6) tag for immobilized metal affinity chromatography. Natural abundance protein samples were prepared from transformed *E. coli* BL21 (DE3) competent cells in standard Luria-Bertani (LB) media (MP Biomedicals) at 37 °C and 200 rpm agitation until an optical density of 0.70-0.80 at 600 nm was reached. Induction of protein expression was achieved with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; GoldBio) for a period of 4 hours at 37 °C and 200 rpm. To prepare isotopically labeled samples for NMR experiments, BL21 (DE3) cells were grown in LB media and transferred to M9 minimal media supplemented with $^{15}NH_4Cl$, ^{13}C-glucose (Cambridge Isotope Laboratories), or both components together, prior to induction [75]. Cells were harvested by centrifugation at 6 000 x *g* at 4 °C for 30 minutes, flash frozen with liquid nitrogen, and stored at -80 °C until purification.

2.5.2 Protein Purification

Cells were lysed by sonication in buffer A (6 M GnHCl (Chem-Impex International), 40 mM Tris pH 8.6). The lysate was centrifuged until clear at 60 000 x *g* at 4 °C for 30 min. An AKTA Start FPLC system (GE Healthcare) was used to equilibrate a 5 mL HisTrap column (GE Healthcare) to buffer A, prior to injection at a rate of 2 mL/min. Column-bound proteins were washed with 2% buffer B (6 M GnHCl, 40 mM Tris pH 8.6, 500 mM imidazole) to baseline

absorbance. The protein was then eluted with 50% buffer B and dialyzed for 10-12 hours at 4 °C against 5 L of Milli-Q H_2O (Millipore Sigma) four times prior to lyophilization. The lyophilized powder stocks were stored at -20 °C.

For NMR resonance assignment, ^{15}N relaxation, and PRE experiments, the His_6 purification tag was removed by Factor Xa (Haematologic Technologies) enzymatic cleavage at a 1:100 enzyme:protein mass ratio in the presence of 50 mM Tris pH 8.0, 50 mM NaCl, 4 mM $CaCl_2$, and 50 mM DPC (Anatrace). The 72-aa (8.07 kDa) cleavage product was subsequently concentrated and buffer exchanged in a Vivaspin-6 10 kDa MWCO centrifugal concentrator (GE Healthcare) to remove the 25-aa N-terminal purification tag. The final sample was prepared in a buffer suitable for NMR studies, including 30 mM sodium acetate at pH 4.5, 50 mM NaCl, 200 mM DPC, 10% ^2H_2O, and 0.03% NaN_3. Experiments carried out at pH 7.0 used 30 mM Tris-HCl instead of sodium acetate buffer. The presence of the intramolecular disulfide bond was confirmed by the absence of a spectrophotometric response from Ellman's reagent [30], [31] (Thermo Scientific), the presence of a single identical monomer band by SDS-PAGE performed in the presence and absence of reducing agent, and a single population of peptide backbone peaks in a ^{15}N-HSQC.

For the NMR DOSY measurements, we needed to measure the translational diffusion of proIAPP in the presence and absence of DPC micelles. However, the cleaved proIAPP without the N-terminal His_6-tag precipitated from solution and would form dense aggregates. For this reason, the DOSY measurements were conducted on the pre-cleavage His_6-proIAPP. The DPC micelle-bound chemical shifts of His_6-proIAPP were compared to those of proIAPP and found to be identical for all residues except for a small deviation in the H^N and N shifts of I3 (data not

shown) which indicates that the additional 26 unstructured residues on the N-terminus have no effect on the folded protein.

2.5.3 NMR Resonance Assignments

Backbone resonance assignments were acquired using the standard Bruker pulse program library (Topspin 3.2, Bruker, Karlsruhe, Germany). Experiments at pH 4.5 included a ^{15}N-HSQC, HNCO, HNCA, CBCA(CO)NH, CBCANH, and ^{15}N-filtered NOESY-HSQC, collected on a Bruker Avance III HD 500 MHz spectrometer with a ^{1}H/^{13}C/^{15}N/^{31}P QXI room-temperature probe. The NMR buffer for the NOESY experiment had ^{2}H-labeled DPC to eliminate protein-lipid NOE peaks. The side-chain assignments were confirmed with an (H)CCH-TOCSY 3D on a Bruker Avance III HD 750 MHz wide-bore spectrometer with a ^{1}H/^{13}C/^{15}N TXI room-temperature probe. Experiments at pH 7.0 included a ^{15}N-HSQC, HNCO, HNCA, and CBCA(CO)NH, collected on a Bruker Avance III HD 500 MHz spectrometer with a ^{1}H/^{13}C/^{15}N/^{31}P QXI room-temperature probe. All proIAPP chemical shift assignments were performed using CARA [76]. Assigned chemical shifts have been deposited in the BMRB with accession number 50007.

Synthesized natural abundance Cpro was purchased from the Protein Core of the University of Illinois at Chicago Research Resources Center. ^{1}H resonance assignment of the peptide was achieved through a ^{1}H TOCSY 2D and a ^{1}H NOESY 2D collected on a Bruker Avance III HD 750 MHz wide-bore spectrometer with a ^{1}H/^{13}C/^{15}N TXI room-temperature probe. The NMR sample contained 1.0 mM Cpro, 200 mM ^{2}H-DPC, 30 mM sodium acetate at pH 4.5, 50 mM NaCl, 10% $^{2}H_2O$, and 0.03% NaN_3. Chemical shift assignment was initially performed manually in Sparky [77], [78] and completed by automation with CYANA [79]. Assigned chemical shifts have been deposited in the BMRB with accession number 50019.

2.5.4 Residual Dipolar Couplings

^{1}H-^{15}N RDCs were collected with an IPAP-HSQC [37] by subtracting measured isotropic J-couplings from J+D-couplings of a partially aligned sample in 6.0 mm charge-neutral, 5.4 mm diameter positively charged, and 5.4 mm diameter negatively charged SAGs [38]. $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ RDCs were similarly obtained on the 5.4 mm SAG sample from a modified HNCOCA pulse program [80] with the ^{1}H composite pulse decoupling removed during the $^{13}C^{\alpha}$ chemical shift evolution period in t_1. All RDC analysis was performed using Sparky [77], [78]. Slight over-alignment of the positive SAG sample resulted in a loss of some of the $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ RDCs in α1. Dipolar waves were fit to experimental data using the sliding window method to the equation $RDC = A\sin\left(\frac{2\pi n}{3.6} - \rho_0\right) + K$, where A is the difference between the largest and smallest RDC in the window, n is the residue number, ρ_0 is the pitch of the first residue in the helix, and K is the average value of the sinusoid in the window [33]. A reduced χ^2 value was calculated for each dipolar wave function to determine which residue numbers to include in the window. χ^2 values of 3.05, 9.24, 1.82, and 2.90 were calculated for α1-4, respectively.

2.5.5 Relaxation Measurements

^{15}N T_1, $T_{1\rho}$, and ^{15}N-{^{1}H} steady-state heteronuclear NOE (hetNOE) experiments [81], [82] were collected for uniformly labeled ^{15}N-proIAPP in NMR buffer at 500 and 750 MHz. The data were processed in NMRPipe [83], and analyzed in Sparky [77], [78] to determine the ^{15}N R_1 and $R_{1\rho}$ relaxation rates and hetNOE values. To obtain T_1 and $T_{1\rho}$ values, pseudo-3D ^{15}N-HSQC experiments were collected with 7 variable delays ranging from 0.1 to 1.6 s for T_1, and 4 to 140 ms with a 1500 Hz spin-lock field for $T_{1\rho}$. R_1 and $R_{1\rho}$ were calculated by fitting a mono-exponential to the signal decay profiles of each experiment for each resonance. R_2 rates were

calculated from $R_{1\rho}$ by removing the contribution from the effective spin-lock field and R_1 rate [84]. The ^{15}N-{^{1}H} hetNOE rates were obtained as previously described [82], using a recycle delay of 9 s. The amide order parameters (S^2) and rotational correlation time (τ_c) were calculated from the Lipari-Szabo formalism using FAST-ModelFree [40], [42].

2.5.6 PRE Experiments

PRE experiments were performed to assess the solvent accessibility of micelle-bound proIAPP. To obtain ^{1}H R_2 rates, pseudo-3D ^{15}N-HSQC experiments were collected for uniformly labeled ^{15}N-proIAPP in NMR buffer with 5 variable delays ranging from 7 ms to 45 ms and a recycle delay of 5 s, in the absence and presence of 2 mM gadodiamide (Omniscan, GE Healthcare). All experiments were collected on a Bruker Avance III 500 MHz wide-bore spectrometer with a ^{1}H/^{13}C/^{15}N TXI room-temperature probe, processed using NMRPipe [83], and analyzed in Sparky [77], [78].

R_2 rates were calculated by fitting a mono-exponential to the signal decay profiles of each experiment for each resonance. PRE rates were then determined by subtracting ^{1}H R_2 rates in the absence of gadodiamide from those determined in the presence of the paramagnetic agent. PRE wave functions for amphipathic helices in a micelle were fit as previously described [46]. Reduced χ^2 values of 1.74, 0.23, and 0.60 were calculated for α1-3, respectively. We were unable to fit a PRE wave function to α4, as there were too few data points in the fit.

2.5.7 Structure Calculation

The ensemble structure of proIAPP bound to DPC micelles was simulated using a simulated annealing protocol with experimental dihedral, NOE, and RDC restraints in an XPLOR-NIH [51], [52]. Dihedral restraints were obtained from chemical shift data using TALOS+ [32]. Interatomic ^{1}H distance restraints were obtained from the ^{15}N-filtered NOESY-

HSQC experiment by calibrating $^{1}H^{N}$ and $^{1}H^{\alpha}$ signal intensities against known distances of an idealized α-helix [50]. ^{1}H-^{15}N and $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ RDC restraints were used directly as measured, with incorporation of an ^{1}H-^{15}N sign flip for the negative gyromagnetic ratio of ^{15}N. All dihedral restraints from TALOS+ were used in the refinement, but NOE and RDC restraints were only included for residues with a general order parameter > 0.6 (**Figure 2.1C**).

Simulated annealing of the proIAPP structure was achieved using the XPLOR-NIH [51], [52] version 2.48 variable-step internal variable module (IVM) procedure [85]. An extended protein molecule was generated from the proIAPP sequence and used as the starting point for all calculations. An initial temperature of 1000 K was used with a linear 10 K step-wise cooldown to 30 K. All experimental and non-experimental restraints were applied equally to both conformers in the ensemble refinement. A statistical torsion angle database (TorsionDB) potential [86] was ramped multiplicatively from 0.001 to 0.25. A purely repulsive nonbonded energy function, REPEL, and the repel14 module from TorsionDB were ramped multiplicatively from 0.004 to 4.0. Backbone amide hydrogen bonds were included as a fixed list using the hydrogen bonding database (HBDB) potential [87] for α1-4 derived from the fitted RDC dipolar wave functions (**Figure 2.2**).

Dihedral restraints were applied with a constant 300 kcal/mol·rad^2 energy term. NOE restraints were applied with a multiplicatively ramped force constant from 0.2 to 30 kcal/Å^2 using a soft square potential. For the 2- and 3-conformer ensemble refinements, NOE restraints were refined against an equally weighted average of the NOE rate (r^{-6} averaging) for each conformer. A total of 457 NOE restraints (327 short-range, 126 medium-range, and 4 long-range) were included in the refinement. A total of 88 ^{1}H-^{15}N and 53 $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ RDC restraints were included in the refinement with a force constant ramped multiplicatively from 0.001

kcal/Hz^2 to 0.09 kcal/Hz^2 for the 1-conformer, and 0.001 kcal/Hz^2 to 0.075 kcal/Hz^2 for the 2- and 3-conformer ensembles. The alignment tensor D_a, rhombicity, and orientation parameters were set as floating variables [88].

Goodness-of-fit validation of the RDC restraints was conducted by a reduced χ^2 analysis. The simulated RDCs from the lowest-energy structure(s) of each ensemble refinement were fit against all experimental RDC restraints by the XPLOR-NIH calcTensor helper program to determine the fit residuals. The reduced χ^2 was calculated from the residuals and the experimental error. We subsequence conducted an F-test ($\alpha = 0.05$) using the degrees of freedom (df) for the number of observations minus the number of Saupe matrix solutions for each ensemble refinement [89].

A total of 400 structures were calculated in the 1-conformer refinement, and a total of 800 sets of structures were calculated in the 2- and 3-conformer refinements. The 20 lowest-energy structures were retained for analysis in the Protein Structure Validation Software suite (PSVS)[90] and deposition in the PDB (**PDB IDs: 6UCJ** and **6UCK**), with structural statistics reported in **Table 2.1**.

2.5.8 Circular dichroism

CD spectra were collected in a variety of lipid environments at two pH values. Samples had a 100-175 μM final protein concentration. Protein concentrations of proIAPP were measured on a NanoDrop 2000c (Thermo Scientific) using an estimated extinction coefficient of 2600 $M^{-1}cm^{-1}$ (CLC Workbench, QIAGEN Bioinformatics). The Cpro peptide samples were prepared from a lyophilized powder stock and diluted to a 100 μM final concentration. The sample buffer contained either 10 mM sodium acetate or 10 mM Tris base, titrated with acetic acid to pH 4.5

and 7.0, respectively. Detergent concentrations were kept well above the CMC at 50 mM for both DPC (Anatrace) and SDS (VWR). Bicelles had a total lipid concentration of 30 mM and a molar q-ratio of 0.5, which consisted of 20 mM DHPC (Avanti Polar Lipids) detergent capping lipid and 10 mM DMPC (Anatrace) or DMPS (Avanti Polar Lipids) planar lipid. Five freeze-thaw cycles were performed to ensure bicelle stock homogeneity. Bicelles containing 25% DMPS were then prepared by mixing DMPC bicelles with DMPS bicelles in a 3:1 ratio, followed by three additional freeze-thaw cycles.

CD experiments were performed on a JASCO J-810 spectropolarimeter. The far-UV spectra from 250 to 190 nm were obtained using a quartz cuvette with a path length of 0.2 mm to circumvent some of the signal loss from bicelle particle scattering [58]–[60]. Five spectra were obtained and averaged with 0.5 nm resolution, 1 nm bandwidth, and a scan speed of 50 nm/min. Buffered blanks containing only lipid were collected with the same parameters as the protein samples and subtracted off the final averaged spectrum. Resulting spectra were smoothed using a five-point means-mode algorithm within the spectral analysis software of the instrument. The analysis was performed by comparing the θ_{MRE} at 222 nm for each membrane environment and pH value to the θ_{MRE} at 222 nm for DPC micelles at pH 4.5, since the NMR structural data was known for that condition.

2.5.9 DOSY Measurements

To calculate the percentage of protein bound to DPC micelles, translational diffusion rates of His_6-proIAPP were measured with and without DPC micelles at pH 4.5 and 7.0 by diffusion-ordered spectroscopy (DOSY). Experiments were performed in triplicate on a Bruker Avance III HD 500 MHz spectrometer with a $^{1}H/^{13}C/^{15}N/^{31}P$ QXI room-temperature probe using a ^{1}H LED-BPP-WATERGATE [91] with experimental parameters as previously described [62].

The rate of translational diffusion (D_t) was extracted from a Gaussian fit of the signal intensity decay profile using NMRPipe [83]. The D_t values were corrected for temperature and viscosity at 32 ºC and a composition of 10% ^2H_2O and 90% H_2O. Diffusion of DPC was followed by the resolved terminal ^1H_3C resonances, while protein diffusion was monitored using $^1H^N$ and $^1H^\alpha$ resonances.

The percentage bound of protein was calculated as described previously [62]. The micelle-bound sample consisted of 1.5 mM natural abundance His_6-proIAPP in NMR buffer at pH 4.5. After the data were collected, the same sample was titrated with Tris base to pH 7.0 and remeasured. The D_t of the free protein was calculated using a sample of 0.7 mM natural abundance His_6-proIAPP in NMR buffer without DPC at pH 4.5. His_6-proIAPP without DPC micelles is insoluble at pH 7.0, and the D_t for the free protein at pH 4.5 was used in the analysis of the binding affinity at pH 7.0.

2.5.10 Bibliography

[1] P. Westermark, C. Wernstedt, E. Wilander, and K. Sletten, "A novel peptide in the calcitonin gene related peptide family as an amyloid fibril protein in the endocrine pancreas," *Biochem. Biophys. Res. Commun.*, vol. 140, no. 3, pp. 827–831, Nov. 1986.

[2] A. N. Roberts *et al.*, "Molecular and functional characterization of amylin, a peptide associated with type 2 diabetes mellitus.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 24, pp. 9662–9666, Dec. 1989.

[3] J. W. M. Höppener, B. Ahrén, and C. J. M. Lips, "Islet Amyloid and Type 2 Diabetes Mellitus," *N. Engl. J. Med.*, vol. 343, no. 6, pp. 411–419, Aug. 2000.

[4] D. L. Hay, S. Chen, T. A. Lutz, D. G. Parkes, and J. D. Roth, "Amylin: Pharmacology, Physiology, and Clinical Potential," *Pharmacol. Rev.*, vol. 67, no. 3, pp. 564–600, Jun.

2015.

[5] D. Naot, D. S. Musson, and J. Cornish, “The Activity of Peptides of the Calcitonin Family in Bone,” *Physiol Rev*, vol. 99, pp. 781–805, 2019.

[6] D. Naot and J. Cornish, “The role of peptides and receptors of the calcitonin family in the regulation of bone metabolism,” *Bone*, vol. 43, no. 5, pp. 813–818, Nov. 2008.

[7] M. Lang, S. De Pol, C. Baldauf, H.-J. Hofmann, O. Reiser, and A. G. Beck-Sickinger, “Identification of the Key Residue of Calcitonin Gene Related Peptide (CGRP) 27−37 to Obtain Antagonists with Picomolar Affinity at the CGRP Receptor,” *J. Med. Chem.*, vol. 49, no. 2, pp. 616–624, Jan. 2006.

[8] J. Barwell, J. J. Gingell, H. A. Watkins, J. K. Archbold, D. R. Poyner, and D. L. Hay, “Calcitonin and calcitonin receptor-like receptors: common themes with family B GPCRs?,” *Br. J. Pharmacol.*, vol. 166, no. 1, pp. 51–65, May 2012.

[9] T. Sanke, G. I. Bell, C. Sample, A. H. Rubenstein, and D. F. Steiner, “An islet amyloid peptide is derived from an 89-amino acid precursor by proteolytic processing.,” *J. Biol. Chem.*, vol. 263, no. 33, pp. 17243–6, Nov. 1988.

[10] S. Mosselman, J. W. M. Höppener, C. J. M. Lips, and H. S. Jansz, “The complete islet amyloid polypeptide precursor is encoded by two exons,” *FEBS Lett.*, vol. 247, no. 1, pp. 154–158, Apr. 1989.

[11] J. W. M. Höppener *et al.*, “Molecular physiology of the islet amyloid polypeptide (IAPP)/amylin gene in man, rat, and transgenic mice,” *J. Cell. Biochem.*, vol. 55, no. S1994A, pp. 39–53, 1994.

[12] L. Marzban *et al.*, “Role of -Cell Prohormone Convertase (PC)1/3 in Processing of Pro-Islet Amyloid Polypeptide,” *Diabetes*, vol. 53, no. 1, pp. 141–148, Jan. 2004.

[13] L. Marzban, G. Soukhatcheva, and C. B. Verchere, “Role of carboxypeptidase E in processing of pro-islet amyloid polypeptide in β-cells,” *Endocrinology*, vol. 146, no. 4, pp. 1808–1817, Apr. 2005.

[14] L. Marzban, G. Trigo-Gonzalez, and C. B. Verchere, “Processing of Pro-Islet Amyloid Polypeptide in the Constitutive and Regulated Secretory Pathways of β Cells,” *Mol. Endocrinol.*, vol. 19, no. 8, pp. 2154–2163, 2005.

[15] J. Wang, J. Xu, J. Finnerty, M. Furuta, D. F. Steiner, and C. B. Verchere, “The Prohormone Convertase Enzyme 2 (PC2) Is Essential for Processing Pro-Islet Amyloid Polypeptide at the NH2-Terminal Cleavage Site,” *Diabetes*, vol. 50, no. 3, pp. 534–539, Mar. 2001.

[16] A. Lukinius, E. Wilander, G. T. Westermark, U. Engström, and P. Westermark, “Co-localization of islet amyloid polypeptide and insulin in the B cell secretory granules of the human pancreatic islets,” *Diabetologia*, vol. 32, no. 4, pp. 240–244, 1989.

[17] M. Stridsberg, S. Sandler, and E. Wilander, “Cosecretion of islet amyloid polypeptide (IAPP) and insulin from isolated rat pancreatic islets following stimulation or inhibition of beta-cell function.,” *Regul. Pept.*, vol. 45, no. 3, pp. 363–70, Jun. 1993.

[18] J. F. Paulsson, A. Andersson, P. Westermark, and G. T. Westermark, “Intracellular amyloid-like deposits contain unprocessed pro-islet amyloid polypeptide (proIAPP) in beta cells of transgenic mice overexpressing the gene for human IAPP and transplanted human islets,” *Diabetologia*, vol. 49, no. 6, pp. 1237–1246, Jun. 2006.

[19] L. Marzban, C. J. Rhodes, D. F. Steiner, L. Haataja, P. A. Halban, and C. B. Verchere, “Impaired NH2-Terminal Processing of Human Proislet Amyloid Polypeptide by the Prohormone Convertase PC2 Leads to Amyloid Formation and Cell Death,” *Diabetes*, vol.

55, no. 8, pp. 2192–2201, Aug. 2006.

[20] I. T. Yonemoto, G. J. A. Kroon, H. J. Dyson, W. E. Balch, and J. W. Kelly, "Amylin proprotein processing generates progressively more amyloidogenic peptides that initially sample the helical state.," *Biochemistry*, vol. 47, no. 37, pp. 9900–10, Sep. 2008.

[21] F. Meng and D. P. Raleigh, "Inhibition of Glycosaminoglycan-Mediated Amyloid Formation by Islet Amyloid Polypeptide and proIAPP Processing Intermediates," *J. Mol. Biol.*, vol. 406, no. 3, pp. 491–502, Feb. 2011.

[22] H. Wang and D. P. Raleigh, "The ability of insulin to inhibit the formation of amyloid by pro-islet amyloid polypeptide processing intermediates is significantly reduced in the presence of sulfated glycosaminoglycans," *Biochemistry*, vol. 53, pp. 2605–2614, 2014.

[23] J. A. Courtade *et al.*, "Measurement of Pro-Islet Amyloid Polypeptide (1–48) in Diabetes and Islet Transplants," *J. Clin. Endocrinol. Metab.*, vol. 102, no. 7, pp. 2595–2603, Jul. 2017.

[24] E. Shardlow, C. Rao, R. Sattarov, L. Wu, P. E. Fraser, and C. Exley, "Aggregation of the diabetes-related peptide ProIAPP1-48 measured by dynamic light scattering," *J. Trace Elem. Med. Biol.*, vol. 51, pp. 1–8, Jan. 2019.

[25] J. D. Knight and A. D. Miranker, "Phospholipid Catalysis of Diabetic Amyloid Assembly," *J. Mol. Biol.*, vol. 341, no. 5, pp. 1175–1187, Aug. 2004.

[26] S. Jha, D. Sellin, R. Seidel, and R. Winter, "Amyloidogenic Propensities and Conformational Properties of ProIAPP and IAPP in the Presence of Lipid Bilayer Membranes," *J. Mol. Biol.*, vol. 389, no. 5, pp. 907–920, 2009.

[27] S. M. Patil, S. Xu, S. R. Sheftic, and A. T. Alexandrescu, "Dynamic α-helix structure of micelle-bound human amylin," *J. Biol. Chem.*, vol. 284, no. 18, pp. 11982–11991, 2009.

[28] R. P. R. Nanga, J. R. Brender, S. Vivekanandan, and A. Ramamoorthy, “Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1808, no. 10, pp. 2337–2342, 2011.

[29] D. Sharma and K. Rajarathnam, “13C NMR chemical shifts can predict disulfide bond formation,” *J. Biomol. NMR*, vol. 18, no. 2, pp. 165–71, Oct. 2000.

[30] G. L. Ellman, “Tissue sulfhydryl groups,” *Arch. Biochem. Biophys.*, vol. 82, no. 1, pp. 70–77, May 1959.

[31] P. W. Riddles, R. L. Blakeley, and B. Zerner, “Ellman’s reagent: 5,5′-dithiobis(2-nitrobenzoic acid)—a reexamination,” *Anal. Biochem.*, vol. 94, no. 1, pp. 75–81, Apr. 1979.

[32] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, “TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts,” *J. Biomol. NMR*, vol. 44, no. 4, pp. 213–223, 2009.

[33] M. F. Mesleh, G. Veglia, T. M. DeSilva, F. M. Marassi, and S. J. Opella, “Dipolar Waves as NMR Maps of Protein Structure,” *J. Am. Chem. Soc.*, vol. 124, no. 16, pp. 4206–4207, Apr. 2002.

[34] G. A. Cook, L. A. Dawson, Y. Tian, and S. J. Opella, “Three-Dimensional Structure and Interaction Studies of Hepatitis C Virus p7 in 1,2-Dihexanoyl- sn -glycero-3-phosphocholine by Solution Nuclear Magnetic Resonance,” *Biochemistry*, vol. 52, no. 31, pp. 5295–5303, Aug. 2013.

[35] M. F. Mesleh and S. J. Opella, “Dipolar Waves as NMR maps of helices in proteins,” *J. Magn. Reson.*, vol. 163, no. 2, pp. 288–299, Aug. 2003.

[36] G. A. Cook and S. J. Opella, “Secondary structure, dynamics, and architecture of the p7 membrane protein from hepatitis C virus by NMR spectroscopy.,” *Biochim. Biophys. Acta*, vol. 1808, no. 6, pp. 1448–53, Jun. 2011.

[37] M. Ottiger, F. Delaglio, and A. Bax, “Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra,” *J. Magn. Reson.*, vol. 131, no. 2, pp. 373–378, Apr. 1998.

[38] J. J. Chou, S. Gaemers, B. Howder, J. M. Louis, and A. Bax, “A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles,” *J. Biomol. NMR*, vol. 21, no. 4, pp. 377–382, 2001.

[39] B. Schmidt and P. J. Hogg, “Search for allosteric disulfide bonds in NMR structures,” *BMC Struct. Biol.*, vol. 7, no. 1, p. 49, 2007.

[40] R. Cole and J. P. Loria, “FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data,” *J. Biomol. NMR*, vol. 26, no. 3, pp. 203–213, 2003.

[41] M. W. Maciejewski *et al.*, “NMRbox: A Resource for Biomolecular NMR Computation,” *Biophys. J.*, vol. 112, no. 8, pp. 1529–1534, 2017.

[42] G. Lipari and A. Szabo, “Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results,” *J. Am. Chem. Soc.*, vol. 104, no. 17, pp. 4559–4570, 1982.

[43] J. Cavanagh, W. Fairbrother, A. Palmer III, M. Rance, and N. Skelton, *Protein NMR Principles And Practice*, 2nd ed. Burlington, MA: Academic Press, 2007.

[44] J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley, and S. Doniach, “Size and shape of detergent micelles determined by small-angle X-ray scattering,” *J. Phys. Chem. B*, vol. 111, no. 43, pp. 12427–12438, 2007.

[45] R. C. Oliver, J. Lipfert, D. A. Fox, R. H. Lo, S. Doniach, and L. Columbus, "Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group," *PLoS One*, vol. 8, no. 5, p. 62488, 2013.

[46] M. Respondek, T. Madl, C. Göbl, R. Golser, and K. Zangger, "Mapping the Orientation of Helices in Micelle-Bound Peptides by Paramagnetic Relaxation Waves," *J. Am. Chem. Soc.*, vol. 129, no. 16, pp. 5228–5234, Apr. 2007.

[47] J. Iwahara, C. D. Schwieters, and G. M. Clore, "Ensemble Approach for NMR Structure Refinement against 1 H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule," *J. Am. Chem. Soc.*, vol. 126, no. 18, pp. 5879–5896, May 2004.

[48] J. Huang and S. Grzesiek, "Ensemble Calculations of Unstructured Proteins Constrained by RDC and PRE Data: A Case Study of Urea-Denatured Ubiquitin," *J. Am. Chem. Soc.*, vol. 132, no. 2, pp. 694–705, Jan. 2010.

[49] J. L. Lorieau, J. M. Louis, C. D. Schwieters, and A. Bax, "pH-triggered, activated-state conformations of the influenza hemagglutinin fusion peptide revealed by NMR," *Proc. Natl. Acad. Sci.*, vol. 109, no. 49, pp. 19994–19999, Dec. 2012.

[50] K. Wüthrich, M. Billeter, and W. Braun, "Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances," *J. Mol. Biol.*, vol. 180, no. 3, pp. 715–740, 1984.

[51] C. D. Schwieters, J. J. Kuszewski, and G. Marius Clore, "Using Xplor-NIH for NMR molecular structure determination," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 48, no. 1. pp. 47–62, Mar-2006.

[52] C. D. Schwieters, G. A. Bermejo, and G. M. Clore, "Xplor-NIH for molecular structure

determination from NMR and other data sources," *Protein Sci.*, vol. 27, no. 1, pp. 26–40, 2018.

[53] E. Jurrus *et al.*, "Improvements to the APBS biomolecular solvation software suite," *Protein Sci.*, vol. 27, no. 1, pp. 112–128, Jan. 2018.

[54] N. C. Kegulian *et al.*, "Membrane Curvature-sensing and Curvature-inducing Activity of Islet Amyloid Polypeptide and Its Implications for Membrane Disruption," *J. Biol. Chem.*, vol. 290, no. 43, pp. 25782–25793, 2015.

[55] W. L. DeLano, "PyMOL: An Open-Source Molecular Graphics Tool," *CCP4 Newsl. Protein Crystallogr.*, vol. 40, pp. 82–92, 2002.

[56] N. E. Shepherd, H. N. Hoang, G. Abbenante, and D. P. Fairlie, "Single Turn Peptide Alpha Helices with Exceptional Stability in Water," *J. Am. Chem. Soc.*, vol. 127, no. 9, pp. 2974–2983, Mar. 2005.

[57] S. T. Joy and P. S. Arora, "An optimal hydrogen-bond surrogate for α-helices," *Chem. Commun.*, vol. 52, no. 33, pp. 5738–5741, Apr. 2016.

[58] B. A. Wallace and D. Mao, "Circular dichroism analyses of membrane proteins: An examination of differential light scattering and absorption flattening effects in large membrane vesicles and membrane sheets," *Anal. Biochem.*, vol. 142, no. 2, pp. 317–328, Nov. 1984.

[59] A. J. Miles and B. A. Wallace, "Circular dichroism spectroscopy of membrane proteins," *Chem. Soc. Rev.*, vol. 45, no. 18, pp. 4859–4872, Sep. 2016.

[60] I. Tinoco, M. F. Maestre, and C. Bustamante, "Circular dichroism in samples which scatter light," *Trends Biochem. Sci.*, vol. 8, no. 2, pp. 41–44, Feb. 1983.

[61] D. C. Rodriguez Camargo *et al.*, "The redox environment triggers conformational changes

and aggregation of hIAPP in Type II Diabetes," *Sci. Rep.*, vol. 7, no. 1, p. 44041, Apr. 2017.

[62] S. T. Smrt, A. W. Draney, and J. L. Lorieau, "The Influenza Hemagglutinin Fusion Domain Is an Amphipathic Helical Hairpin That Functions by Inducing Membrane Curvature," *J. Biol. Chem.*, vol. 290, no. 1, pp. 228–238, Jan. 2015.

[63] D. S. Wishart, B. D. Sykes, and F. M. Richards, "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy," *Biochemistry*, vol. 31, no. 6, pp. 1647–1651, Feb. 1992.

[64] J. L. Lorieau, J. M. Louis, and A. Bax, "Whole-Body Rocking Motion of a Fusion Peptide in Lipid Bilayers from Size-Dispersed 15 N NMR Relaxation," *J. Am. Chem. Soc.*, vol. 133, no. 36, pp. 14184–14187, Sep. 2011.

[65] M. Cluck, C. Chan, and T. Adrian, "The regulation of amylin and insulin gene expression and secretion," *Pancreas*, vol. 30, no. 1, pp. 1–14, Jan. 2005.

[66] P. Paroutis, N. Touret, and S. Grinstein, "The pH of the Secretory Pathway: Measurement, Determinants, and Regulation," *Physiology*, vol. 19, no. 4, pp. 207–215, Aug. 2004.

[67] L. Marzban, K. Park, and C. B. Verchere, "Islet amyloid polypeptide and type 2 diabetes," *Exp. Gerontol.*, vol. 38, pp. 347–351, 2003.

[68] S. P. Smeekens *et al.*, "Proinsulin processing by the subtilisin-related proprotein convertases furin, PC2, and PC3.," *Proc. Natl. Acad. Sci.*, vol. 89, no. 18, pp. 8822–8826, Sep. 1992.

[69] G. Lipkind and D. F. Steiner, "Predicted Structural Alterations in Proinsulin during Its Interactions with Prohormone Convertases †," *Biochemistry*, vol. 38, no. 3, pp. 890–896, Jan. 1999.

[70] Y. Yang *et al.*, "Solution structure of proinsulin: connecting domain flexibility and prohormone processing.," *J. Biol. Chem.*, vol. 285, no. 11, pp. 7847–51, Mar. 2010.

[71] C. E. Higham *et al.*, "Processing of synthetic pro-islet amyloid polypeptide (proIAPP) 'amylin' by recombinant prohormone convertase enzymes, PC2 and PC3, in vitro," *Eur. J. Biochem.*, vol. 267, no. 16, pp. 4998–5004, Aug. 2000.

[72] C. Hillyard, G. Abeyasekera, R. Craig, C. Myers, J. Stevenson, and I. Macintyre, "KATACALCIN: A NEW PLASMA CALCIUM-LOWERING HORMONE," *Lancet*, vol. 321, no. 8329, pp. 846–848, Apr. 1983.

[73] K. Kitamura, J. Sakata, K. Kangawa, M. Kojima, H. Matsuo, and T. Eto, "Cloning and Characterization of cDNA Encoding a Precursor for Human Adrenomedullin," *Biochem. Biophys. Res. Commun.*, vol. 194, no. 2, pp. 720–725, Jul. 1993.

[74] T. Shimosawa, Y. Ito, K. Ando, K. Kitamura, K. Kangawa, and T. Fujita, "Proadrenomedullin NH(2)-terminal 20 peptide, a new product of the adrenomedullin gene, inhibits norepinephrine overflow from nerve endings.," *J. Clin. Invest.*, vol. 96, no. 3, pp. 1672–1676, Sep. 1995.

[75] M. Cai, Y. Huang, K. Sakaguchi, G. M. Clore, A. M. Gronenborn, and R. Craigie, "An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli," *J. Biomol. NMR*, vol. 11, no. 1, pp. 97–102, 1998.

[76] R. L. J. Keller, "Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment," ETH, 2005.

[77] T. D. Goddard and D. G. Kneller, "SPARKY 3," *Univ. California, San Fransisco*, 2008.

[78] W. Lee, M. Tonelli, and J. L. Markley, "NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy," *Bioinformatics*, vol. 31, no. 8, pp. 1325–1327, 2015.

[79] P. Güntert, "Automated NMR Structure Calculation With CYANA," in *Protein NMR Techniques*, vol. 278, New Jersey: Humana Press, 2004, pp. 353–378.

[80] A. Bax and M. Ikura, "An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the α-carbon of the preceding residue in uniformly 15N/13C enriched proteins," *J. Biomol. NMR*, vol. 1, no. 1, pp. 99–104, May 1991.

[81] L. E. Kay, D. A. Torchia, and A. Bax, "Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease," *Biochemistry*, vol. 28, no. 23, pp. 8972–8979, Nov. 1989.

[82] N. A. Farrow *et al.*, "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by15N NMR Relaxation," *Biochemistry*, vol. 33, no. 19, pp. 5984–6003, 1994.

[83] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "NMRPipe: A multidimensional spectral processing system based on UNIX pipes," *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[84] A. G. Palmer and F. Massi, "Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy," *Chem. Rev.*, vol. 106, no. 5, pp. 1700–1719, May 2006.

[85] C. D. Schwieters and G. M. Clore, "Internal coordinates for molecular dynamics and minimization in structure determination and refinement," *J. Magn. Reson.*, vol. 152, no. 2, pp. 288–302, Oct. 2001.

[86] G. A. Bermejo, G. M. Clore, and C. D. Schwieters, "Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures," *Protein Sci.*, vol. 21, no. 12,

pp. 1824–1836, Dec. 2012.

[87] A. Grishaev and A. Bax, "An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation," *J. Am. Chem. Soc.*, vol. 126, no. 23, pp. 7281–7292, 2004.

[88] G. M. Clore and C. D. Schwieters, "How Much Backbone Motion in Ubiquitin Is Required To Account for Dipolar Coupling Data Measured in Multiple Alignment Media as Assessed by Independent Cross-Validation?," *J. Am. Chem. Soc.*, vol. 126, no. 9, pp. 2923–2938, Mar. 2004.

[89] S. Kokoska, *Introductory Statistics: A Problem Solving Approach*, 2nd ed. W. H. Freeman, 2015.

[90] A. Bhattacharya, R. Tejero, and G. T. Montelione, "Evaluating protein structures determined by structural genomics consortia.," *Proteins*, vol. 66, no. 4, pp. 778–95, Mar. 2007.

[91] J. J. Chou, J. L. Baber, and A. Bax, "Characterization of Phospholipid Mixed Micelles by Translational Diffusion," *J. Biomol. NMR*, vol. 29, no. 3, pp. 299–308, Jul. 2004.

Chapter 3

The Development of Super Resolution Nuclear Overhauser Effect Spectroscopy

A significant portion of this chapter has been reprinted and adapted with permission from Springer Nature: “Super resolution NOESY spectra of proteins”, Charles F. DeLisle, H. Bhagya Mendis, and Justin L. Lorieau, Journal of Biomolecular Nuclear Magnetic Resonance, 2019. 73(3): 105-116. Supplemental figures and tables can be found in Appendix A.

Spectral resolution remains one of the most significant limitations in the NMR study of biomolecules. We present the srNOESY (super resolution nuclear Overhauser effect spectroscopy) experiment, which enhances the resolution of NOESY cross-peaks at the expense of the diagonal peak line-width. We studied 2 proteins, ubiquitin and the influenza hemagglutinin fusion peptide in bicelles, and we achieved average resolution enhancements of 21-47% and individual peak enhancements as large as *ca.* 450%. New peaks were observed over the conventional NOESY experiment in both proteins as a result of these improvements, and the final structures generated from the calculated restraints matched published models. We discuss the impact of experimental parameters, spin diffusion and the information content of the srNOESY lineshape.

3.1 Introduction

NMR spectroscopy is a valuable tool in the elucidation of bonding topology, molecular structure and molecular dynamics in chemistry and structural biology. Yet, the complexity and size of interesting biomolecules still limits the utility of NMR in structural studies. As molecules become larger, the number of resonances grows with the number of atoms and tumbling (rotational) times increase, thereby producing lower resolution spectra with broadened and more numerous peaks. Consequently, the study of large biomolecules is prohibitive by NMR. Different approaches mitigate these problems, including methyl spectroscopy[1]–[4], transverse relaxation optimized spectroscopy (TROSY)[5], [6], partial deuteration and site-specific labeling,[7]–[9] and resolution enhancement with constant-time evolution [10]. These approaches often require specialized samples to study only a subset of atoms in a molecule, and may require *a priori* knowledge of the molecular structure. Importantly, they may preclude essential structural information, such as the distance restraints of perdeuterated molecules collected through NOESY spectra,[11], [12] which benefits from extensive ^{1}H spin labeling.

In this article, we present an easily implemented approach to the resolution enhancement of fully protonated molecules with NOESY spectroscopy. We achieve average linewidth reductions of 16-32% (resolution enhancements of 21-47%) and linewidth reductions as large as 78% to resolve new peaks not observed in conventional NOESY experiments. These enhancements represent apparent linewidths that are narrower than the natural linewidth for a given resonance.

Our approach uses a time-dependent NOE mix period (t_{mix}) that increments with the evolution period in the indirect ^{1}H dimension, t_1. A NOESY experiment with a linear time-dependent mixing time was previously reported for the suppression of J-couplings in spectra [13], [14]. However, the aim of those studies was to minimally change the NOE mix time (*ca.*

20% of t_{mix}) to maintain a relatively constant NOE cross-peak intensity while suppressing coherent J-coupling transfers. Our approach incorporates significant changes in the NOE mix time (*ca.* 300-500% of t_{mix}) to utilize the signal buildup of a cross-peak to enhance its resolution.

The srNOESY uses two concepts to improve the quality of spectra. At short t_1 values in the indirect dimension, the t_{mix} is small and close to its initial value. The intensity of a peak is emphasized in the initial evolution of a time-dependent signal, and the resolution is emphasized in the later portions. Shorter t_{mix} periods are desirable for accurate NOESY distance restraints because spin diffusion pathways are minimally expressed in this regime [15], [16]. Second, at longer t_1 values, the t_{mix} is significantly increased from its initial value to utilize the signal buildup of the cross-peak to enhance resolution. At longer t_{mix} periods, direct transfer and spin diffusion pathways continue to transfer magnetization between spins to increase the intensity of cross-peaks, yet this additional intensity is typically discarded in favor of greater accuracy for NOE distance restraints. The srNOESY experiment utilizes this buildup period to enhance the resolution of peaks while having a minimal impact on the accuracy of cross-peak intensities.

3.2 srNOESY Materials and Methods

3.2.1 Sample Preparation

A DNA sequence for ubiquitin was subcloned in pET-15b (Genscript), and the plasmid was transformed into *E. coli* BL21(DE3) cells (Fisher). Expression of uniformly ^{15}N-labeled ubiquitin was achieved as previously described [17]. Bacterial cells were lysed by sonication in a pH 7.8 buffer containing 50 mM Tris, 500 mM NaCl and 20 mM imidazole. The resulting slurry was centrifuged at 70,000 x *g* for 30 min at 4 °C. The protein solution was passed through a His-Trap column (GE) with an AKTA Start FPLC (GE) and eluted with the lysis buffer containing an

additional 250 mM imidazole. The purity of the sample was confirmed by SDS-PAGE. The folded structure of a 0.8 mM ^{15}N-labeled ubiquitin sample in 10 mM sodium phosphate, 0.03% NaN_3 7% D_2O at pH 6.6 and 25 ºC was confirmed with an ^{15}N-HSQC spectrum compared to published chemical shifts [18].

Influenza HAfp expression and purification was achieved as previously described,[19], [20] with a final size-exclusion chromatography step using a Superdex 75 26/600 PG column (GE). The sample purity was confirmed with SDS-PAGE and MALDI-TOF mass spectrometry. The folded structure of 1.0 mM ^{15}N-labeled HAfp in 30 mM Tris-d_{11} with q=0.44 bicelles with 44 mM ^{2}H-dimyristoyl phosphatidylcholine (^{2}H-DMPC) and 100 mM dihexanoyl phosphatidylcholine (DHPC), 0.03% NaN_3, 10% D_2O at pH 7.2 was confirmed with published ^{15}N-HSQC chemical shifts [19].

3.2.2 NMR Experiments

NMR spectra were recorded for uniformly ^{15}N-labeled ubiquitin (10 mM sodium phosphate, 93% H_2O/7% D_2O at pH 6.6 and a temperature of 25°C) and uniformly ^{15}N-labeled HAfp bound to q = 0.44 ^{2}H-DMPC/DHPC bicelles (30 mM Tris-d_{11}, 44 mM DMPC-d_{54}, 100 mM DHPC, 90% H_2O/10% D_2O at pH 7.2 and a temperature of 32°C). A Bruker AVIII-750 wide-bore spectrometer equipped with a ^{1}H/^{13}C/^{15}N TXI 5 mm room temperature probe was used for all NOESY-HSQC and HAfp-bicelle ^{15}N relaxation experiments. A Bruker AVIII-500 wide-bore spectrometer equipped with a ^{1}H/^{13}C/^{15}N TXI 5 mm room temperature probe was used for ubiquitin ^{15}N relaxation experiments. Spectra were apodized using a sine-bell function with an initial value of 0.45π and a final value of 0.90π, and processed with zero-filling to 1024 points in

the indirect ^{1}H dimension, 512 points in the ^{15}N indirect dimension, and 2048 points in the direct ^{1}H dimension.

3.2.3 Data Analysis

NMRPipe [21] was used for processing all NMR data, and Sparky [22], [23] was selected for NOESY peak shape analysis due to its peak deconvolution features. The correlation times of each system were determined by fitting the ^{15}N relaxation rates using FAST-Modelfree within the NMRbox software suite [24]–[26]. The ^{15}N R_1, $R_{1\rho}$, R_2, and ^{15}N-{^{1}H} NOE rates were determined for both systems as previously described [20].

A structure refinement using Xplor-NIH[27] v2.47 with simulated annealing was performed with the inclusion of published NOESY distance restraints replaced with data from the srNOESY-HSQC spectra in this study. The simulations matched the published protocols for the structures. Data for ubiquitin were obtained from PDB ID 1D3Z [18] and data for HAfp were obtained from 2KXA [19].

3.3 srNOESY Theory

Cross-relaxation in a homonuclear [^{1}H,^{1}H]-NOESY experiment is governed by large relaxation matrices that couple the direct transfer of magnetization between two spins as well as numerous indirect spin diffusion pathways through local and intermediary spins [15]. In the simplest case, the direct cross-relaxation between two spins, ‘i’ and ‘s,’ is represented by two coupled different equations and a 2x2 relaxation matrix [15], [28], [29].

$$\begin{aligned}\frac{d\Delta I_z(t)}{dt} &= -R_{1,i}\Delta I_z(t) - \sigma_{is}\Delta S_z(t)\\ \frac{d\Delta S_z(t)}{dt} &= -\sigma_{is}\Delta I_z(t) - R_{1,s}\Delta S_z(t)\end{aligned} \tag{1}$$

$\Delta I_z(t)$ is the deviation from the Boltzmann equilibrium magnetization for the diagonal peak and $\Delta S_z(t)$ is the magnetization for the cross-peak. $R_{1,i}$ and $R_{1,s}$ are the auto-relaxation rates for spins 'i' and 's', and σ_{is} is the cross-relaxation (NOE) rate between spins 'i' and 's.' These can be calculated using semi-classical theory and spectral density functions (J(ω)) at spectrometer frequencies 'ω' [28], [30].

$$R_{1,i} = \frac{\gamma_H^4 \hbar^2}{10} \sum_{\substack{s=1 \\ s \neq i}}^{spins} \frac{1}{\left\langle r_{is}^6 \right\rangle} \left(1J(0) + 3J(\omega_H) + 6J(2\omega_H) \right)$$
$$\sigma_{is} = \frac{\gamma_H^4 \hbar^2}{10} \frac{1}{\left\langle r_{is}^6 \right\rangle} \left(6J(2\omega_H) - 1J(0) \right) \tag{2}$$

The gyromagnetic ratio of ^{1}H is represented by γ_H, $\hbar$ is the Planck constant. The $\langle r_{is}^6 \rangle$ term is the internuclear distance to the 6th power for spins 'i' and 's.' The angle brackets denotes a motionally averaged value over r_{is}^6 from internal motions on a picosecond timescale.

For simplicity, we have neglected the contribution of the ^{1}H chemical shift anisotropy (CSA) since the ^{1}H-^{1}H dipolar interaction dominates ^{1}H relaxation at the NMR fields measured (500- and 600-MHz). The spectral density function, *J(nω)*, may adopt a variety of functional forms depending on the type and timescale of internal and overall motions for a molecule [24], [31]. The simplest form is represented by a molecule in solution that is internally rigid and that tumbles (rotates) isotropically with a correlation time τ_c, where *n* represents the zero, single, and double quantum contributions to the Larmor frequency.

$$J(n\omega) = \frac{\tau_c}{1 + (n\omega\tau_c)^2} \tag{3}$$

In a 600-MHz NMR spectrometer, the *J(0)* term dominates the auto- and cross-relaxation rates when the τ_c is above 0.3 ns, representing an intrinsically ordered molecule larger than approximately 700 Daltons [32]. This condition is known as the 'spin-diffusion' regime, and it

will be used in the analysis of the large biomolecules in this study. The general principles of the srNOESY are nevertheless applicable to smaller molecules.

In the spin-diffusion limit, the auto- and cross-relaxation rates can be simplified.

$$\sigma_{is} = -\frac{\gamma_H^4 \hbar^2}{10} \tau_c \frac{1}{\left\langle r_{is}^6 \right\rangle}$$
$$R_{1,i} = \frac{\gamma_H^4 \hbar^2}{10} \tau_c \sum_{\substack{s=1 \\ s \neq i}}^{spins} \frac{1}{\left\langle r_{is}^6 \right\rangle} = -\sum_{\substack{s=1 \\ s \neq i}}^{spins} \sigma_{is} \tag{4}$$

For a large molecule with no internal motion, σ_{is} is negative and the R_1 rates are positive. These rates represent the maximum magnitudes for the rates at a given τ_c, and fast internal motions will reduce their magnitude. Additionally, the R_{2H} rate, or rate of transverse relaxation of ^{1}H spin 'i', is evaluated as follows [30]:

$$R_{2,i} = \frac{\gamma_H^4 \hbar^2}{20} \sum_{\substack{s=1 \\ s \neq i}}^{spins} \frac{1}{\left\langle r_{is}^6 \right\rangle} \left(9J(\omega_H) + 5J(0) + 6J(2\omega_H) \right)$$
$$= -\frac{5}{2} \sum_{\substack{s=1 \\ s \neq i}}^{spins} \sigma_{is} \tag{5}$$

The R_2 rates are also positive.

In 2-dimensional NOESY experiments, or NOESY experiments with higher dimensionalities, the contribution of cross-relaxation and auto-relaxation can be resolved from cross-peaks, $\Delta S_z(t)$, and diagonal peaks, $\Delta I_z(t)$, respectively. The coupled differential equation (1) can be solved analytically [29]. For simplicity, we present a solution for the buildup of the cross-peak magnetization, $\Delta S_z(t)$, when the auto-relaxation rates, $R_{1,i}$ and $R_{1,s}$, are approximately equal ($R_{1,i} = R_{1,s} = R_1$).

$$\Delta I_z(t) = \cosh(\sigma_{is} t) \cdot \exp(-R_1 t) \cdot \Delta I_z(0)$$
$$\Delta S_z(t) = \sinh(\sigma_{is} t) \cdot \exp(-R_1 t) \cdot \Delta I_z(0) \tag{6}$$

Note that equation (2.1) of Vögeli et al. [29] should read:

$$\frac{\Delta I_z(t)}{\Delta I_z(0)} = \frac{1}{2}\left[\left(1+\frac{\rho_I-\rho_S}{\lambda_+-\lambda_-}\right)e^{-\lambda_- t}+\left(1-\frac{\rho_I-\rho_S}{\lambda_+-\lambda_-}\right)e^{-\lambda_+ t}\right] \tag{7}$$

Initially, the system has an initial magnetization for the donor spin ($\Delta I_z\,(0) \neq 0$) and the cross-peak magnetization is zero ($\Delta S_z\,(0) \neq 0$). In the limit that the auto-relaxation rate is small or the mix time is short (i.e. $R_1 t = 0$), the cross-peak follows a linear buildup, as previously described [16]. At long mixing times, the cross-peak and diagonal peak magnetizations reach their Boltzmann equilibrium values due to the the exponential decay, exp $(-R_1 t)$.

For a 2-dimensional, or higher dimensional NOESY experiment, the initial magnetization is modulated by the chemical shift (ω_{cs}) and ^{1}H R_2 relaxation in the evolution period prior to the NOE mixing block.

$$\begin{aligned}\mathrm{Re}\{\Delta S_z(t_{mix},t_1)\} &= \cos(\omega_{CS}t_1)\cdot\exp(-R_2t_1)\cdot\Delta S_z(t_{mix})\\ \mathrm{Im}\{\Delta S_z(t_{mix},t_1)\} &= \sin(\omega_{CS}t_1)\cdot\exp(-R_2t_1)\cdot\Delta S_z(t_{mix})\end{aligned} \tag{8}$$

With TPPI or States acquisition modes, the real and imaginary components are collected by incrementing the phase of the first or second NOESY pulse by 90º.

In a normal NOESY experiment, the intensity of the NOE cross-peak, $S(t_{mix})$, can be evaluated after the first dimension and the fixed mix t_{mix} period.

$$\begin{aligned}\mathrm{Re}\{\Delta S(t_1)\} &= \cos(\omega_{CS}t_1)\cdot\exp(-R_2t_1)\cdot\Delta S_{noe}\\ \mathrm{Im}\{\Delta S(t_1)\} &= \sin(\omega_{CS}t_1)\cdot\exp(-R_2t_1)\cdot\Delta S_{noe}\\ \Delta S_{noe} &= \sinh\left(\sigma_{is}t_{mix}\right)\cdot\exp(-R_1t_{mix})\cdot\Delta I_z(0)\end{aligned} \tag{9}$$

In the srNOESY experiment, the t_{mix} period depends on the t_1 time period.

$$t_{mix} = \sum_{i=0}^{N} a_i t_1^i \tag{10}$$

In the conventional NOESY experiment, only the *N=0* term is present. In this study, the srNOESY experiment utilizes mixing times with an *N=1* (linear) time dependency.

$$t_{mix} = a_0 + a_1 t_1 \tag{11}$$

The srNOESY cross-peak likewise evolves for a two-spin system by combining equations (6) and (8).

$$\begin{aligned}\mathrm{Re}\{\Delta S(t_1)\} &= \cos(\omega_{CS} t_1)\cdot\sinh(\sigma_{is} a_0 + \sigma_{is} a_1 t_1)\cdot\exp(-R_2 t_1 - R_1 a_0 - R_1 a_1 t_1)\cdot\Delta I_z(0)\\ \mathrm{Im}\{\Delta S(t_1)\} &= \sin(\omega_{CS} t_1)\cdot\sinh(\sigma_{is} a_0 + \sigma_{is} a_1 t_1)\cdot\exp(-R_2 t_1 - R_1 a_0 - R_1 a_1 t_1)\cdot\Delta I_z(0)\end{aligned} \tag{12}$$

To simplify the analytic expression, equation (6) can be expressed in the linear build-up regime for the cross-peak ($\Delta S(t)$) using a Taylor expansion for the sinh ($\sigma_{is} t$) function.

$$\Delta S(t) = \sigma_{is} t \cdot \exp(-R_1 t) \tag{13}$$

This function has the same initial time dependence as the cross-peak function in equation (6), yet the function decays more quickly at longer mixing times. Consequently, the predicted srNOESY resolution enhancement will be smaller than the experimental enhancement, using this approximation.

We then evaluate the behavior of the cross-peak in the srNOESY.

$$\begin{aligned}\mathrm{Re}\{I_{ij}(t_1)\} &= \cos(\omega_{CS} t_1)\cdot\sigma_{ij}(a_0 + a_1 t_1)\cdot\exp(-R_{2,i} t_1 - R_1 a_0 - R_1 a_1 t_1)\cdot\Delta I_z(0)\\ &= \cos(\omega_{CS,i} t_1)\left(1 + \frac{a_1 t_1}{a_0}\right)\cdot\exp(-R_{2,i} t_1 - R_1 a_1 t_1)\cdot\Delta S_{noe}\\ \mathrm{Im}\{I_{ij}(t_1)\} &= \sin(\omega_{CS} t_1)\cdot\sigma_{ij}(a_0 + a_1 t_1)\cdot\exp(-R_{2,i} t_1 - R_1 a_0 - R_1 a_1 t_1)\cdot\Delta I_z(0)\\ &= \sin(\omega_{CS,i} t_1)\left(1 + \frac{a_1 t_1}{a_0}\right)\cdot\exp(-R_{2,i} t_1 - R_1 a_1 t_1)\cdot\Delta S_{noe}\\ \Delta S_{noe} &= \sigma_{is} a_0 \cdot\exp(-R_1 a_0)\cdot\Delta I_z(0)\end{aligned} \tag{14}$$

In a conventional NOESY, the cross-peak evolves during t_1 and decays exponentially by $R_2 t_1$. The decay component can be isolated from equation (9).

$$D_{\mathrm{NOESY}}(t_1) = \exp(-R_2 t_1) \tag{15}$$

This function gives the cross-peak a Lorentzian shape.

In the srNOESY, the cross-peak decay includes the contribution from the NOE build-up. Its contribution to the shape of the peak and width can also be isolated from equation (14).

$$D_{\text{srNOESY}}(t_1) = \left(1 + \frac{a_1 t_1}{a_0}\right) \cdot \exp(-R_{2,i} t_1 - R_1 a_1 t_1) \tag{16}$$

The function increases in intensity before a more rapid exponential decay. The resulting spectrum (**Figure 3.1**) has a non-Lorentzian lineshape closer to a Lorentz-to-Gaussian lineshape or the first lobe of an offset sine-bell function. Differences between the srNOESY decay function and apodization functions are described in the Discussion.

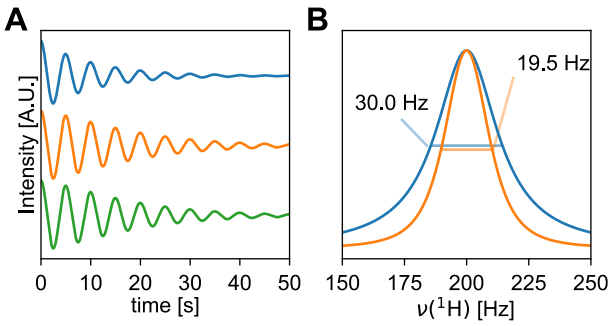


Figure 3.1. Simulated effect of the srNOESY pulse program on NMR signal. Simulation of the (**A**) free induction decays (FIDs) and (**B**) Fourier transformed spectra for a reference NOESY cross-peak (blue), an srNOESY cross-peak modeled with equation (12) and a $\sinh(\sigma_{is}t)e^{-\rho t}$ function (orange), and an srNOESY cross-peak modeled with equation (14) and a $\sigma_{is}t \cdot e^{-\rho t}$ function (green). The FIDs and spectra were simulated using a 2-spin system with the following parameters for spin 'i': $\nu = 200$ Hz, $\tau_c = 4.9$ ns, $R_2 = 30\ \pi\ s^{-1}$, $R_1 = 1.0\ s^{-1}$ and a $\sigma_{is} = 2.0\ s^{-1}$ ($r_{HH} = 2.2$ Å). The reference FID was simulated with a 50 ms mixing time, and the srNOESY FIDs were simulated with $a_0 = 50$ ms and $a_1 = 4$. The full-widths at half-height (FWHH) are labeled for each peak in the FT spectra in panel (B). The FT spectra for the two srNOESY cross-peaks (green and orange) are identical at this magnification.

The diagonal peaks adopt the following form in the srNOESY experiment, using equations (6) and (8).

$$\begin{aligned}\mathrm{Re}\{\Delta S(t_1)\} &= \cos(\omega_{CS}t_1)\cdot\cosh(\sigma_{is}a_0+\sigma_{is}a_1t_1)\cdot\exp(-R_2t_1-R_1a_0-R_1a_1t_1)\cdot\Delta I_z(0)\\ \mathrm{Im}\{\Delta S(t_1)\} &= \sin(\omega_{CS}t_1)\cdot\cosh(\sigma_{is}a_0+\sigma_{is}a_1t_1)\cdot\exp(-R_2t_1-R_1a_0-R_1a_1t_1)\cdot\Delta I_z(0)\end{aligned} \tag{17}$$

A Taylor expansion for the $\cosh(\sigma_{is}t)$ function shows that the diagonal decays initially with an exponential function.

$$\Delta I(t) = \exp(-R_1t) \tag{18}$$

In the srNOESY, the diagonal peak decay includes the additional intensity decay from the NOE transfer.

$$D_{\mathrm{srNOESY}}(t_1) = \exp(-R_{2,i}t_1 - R_1a_1t_1) \tag{19}$$

Altogether, the srNOESY produces sharper cross-peaks but the diagonal peaks are broader.

3.4 srNOESY Results

The srNOESY-HSQC pulse sequence (**Figure 3.2**) was modified from a conventional ^{1}H-NOESY-^{15}N-HSQC (NOESY-HSQC) to include a t_{mix} that increases with the ^{1}H evolution period (t_1) according to equation (11). The conventional (reference) NOESY-HSQC used a constant t_{mix}, and it was collected with the same parameters.

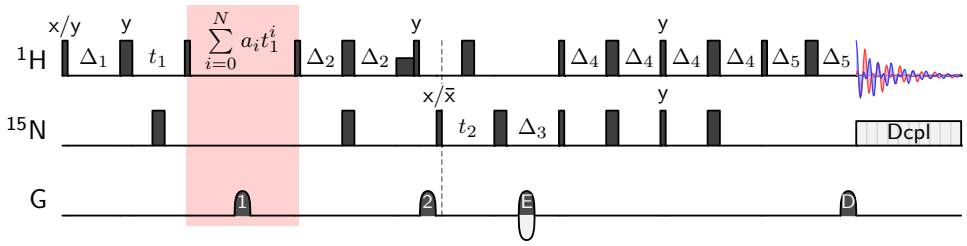


Figure 3.2. The srNOESY-HSQC pulse sequence. The pulse sequence is based on the Bruker NOESY-HSQC pulse sequence (noesyhsqcf3gpsi3d) with the NOE mixing time modified to increment with the t_1 evolution delay in the F_1 dimension. Thin lines represent hard 90º pulses and thick lines represent hard 180º pulses. States-TPPI phase discrimination in F_1 was achieved by incrementing the phase of the ^{1}H 'x/y' 90º pulse by 90º. Echo-antiecho phase discrimination in F_2 was achieved by incrementing the phase of the ^{15}N 'x/-x' pulse by 180º and inverting the sign of the encoding gradient, labeled 'E'. A WALTZ-16 decoupling scheme was used in collecting the direct dimension FID (F_3) [33]. A minimum of 8-steps in the phase cycle were collected. See the pulse program for phase cycle and delay details [34], [35].

Two well-characterized protein systems of very different molecular size were selected to validate this method: ubiquitin (98 amino acids, aa, 10.9 kDa, including a 22 aa N-terminal His_6 tag) and the influenza hemagglutinin fusion peptide domain bound to large, isotropically tumbling bicelles (HAfp-bicelles, 30 aa with ^{2}H-DMPC/DHPC bicelles at a molar ratio, q, of 0.44). Based on ^{15}N relaxation experiments (see **Figure 3.S1**, Appendix B), the τ_c for ubiquitin is 4.25 ± 0.03 ns at 25ºC and the τ_c for HAfp-bicelles is 18.95 ± 0.14 ns at 32ºC. The calculated equivalent globular protein molecular weight for the HAfp-bicelle system is *ca.* 49.5 kDa, consistent with a 450 aa globular protein [36], [37]. For the enhanced srNOESY-HSQC of ubiquitin, an a_1 of 4 with a base NOE mix time (a_0) of 50 ms was used, whereas an a_1 of 2 and a_0 of 25 ms was employed for HAfp-bicelles. These parameters were found to be optimal (see **Figure 3.S2-3.S3**, Appendix B), given the difference in ^{1}H R_1 between the two systems. For example, an a_0 of 100 ms for ubiquitin and 75 ms for HAfp-bicelles only produced modest improvements in resolution (*ca.* 1-9%) since the cross-peak intensity is nearer to the decay portion of the evolution, rather than the buildup. We also tested quadratic functions for equation (10), but we found that linear functions produced the best resolution enhancements (see **Figure 3.S3A**, Appendix B).

In comparison to a conventional (reference) NOESY-HSQC, the srNOESY-HSQC produces average cross-peak linewidth reductions of 31 ± 1% for ubiquitin and 18 ± 2% for HAfp-bicelles in the indirect ^{1}H dimension (**Figure 3.3A, C**). The corresponding resolution enhancement is 45 ± 2% for ubiquitin and 22 ± 1% for HAfp-bicelles. A comparison of the linewidths for the ^{15}N indirect and ^{1}H direct dimensions (see **Figure 3.S4**, Appendix B) shows that both experiments have the same resolution in the other dimensions. In contrast, the diagonal peaks receive a reduction in resolution in the indirect ^{1}H dimension with an average linewidth

increase of 24 ± 2% for ubiquitin and 26 ± 3% for HAfp-bicelles (**Figure 3.3B, D**). In principle, the broadened diagonal peaks could pose a problem for cross-peaks that are very close to the diagonal, although this was not a problem in resolving any of the $^{1}H^{N}$-$^{1}H^{N}$ cross-peaks in ubiquitin and HAfp-bicelles.

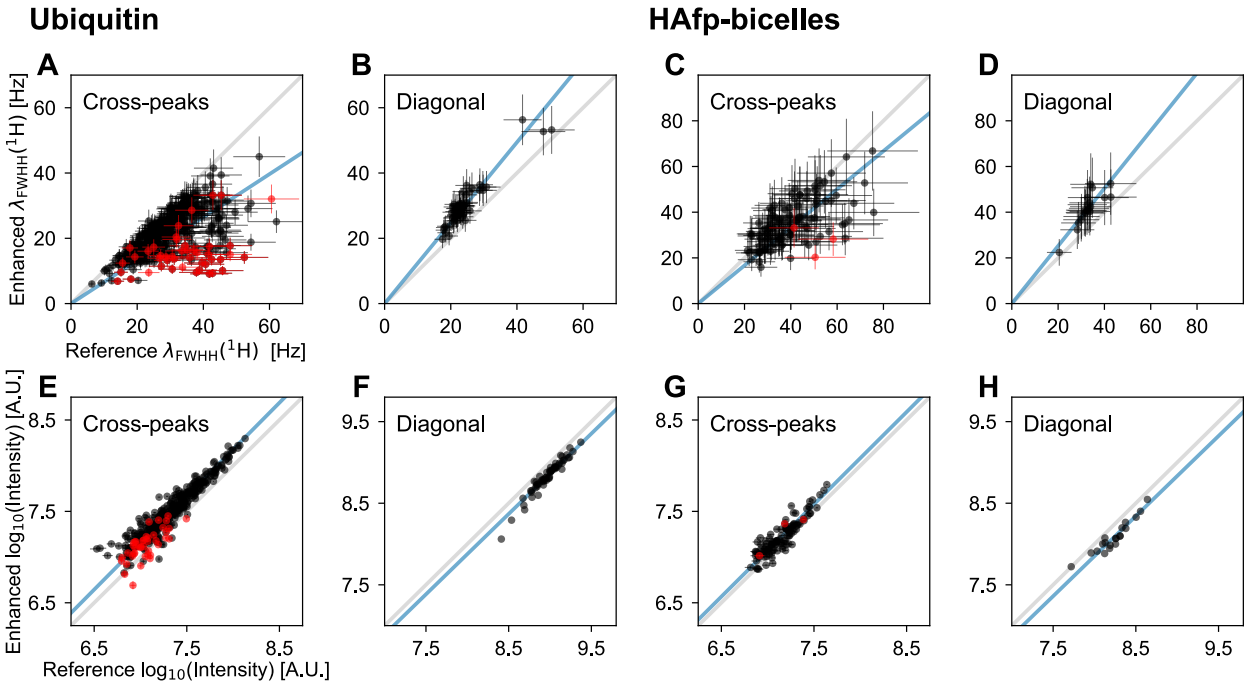


Figure 3.3. Comparison plots of the (A-D) ^{1}H indirect dimension FWHH linewidths and (E-H) log intensity correlations between a reference NOESY-HSQC and an enhanced srNOESY-HSQC. An a_0 of 50 ms and an a_1 of 4 were used in the ubiquitin experiments, while an a_0 of 25 ms and an a_1 of 2 were used in the HAfp-bicelles experiments. Visual guide 1:1 lines are shown in gray, and the linear regression lines (intercept = 0) are shown in blue. The red circles represent single peaks in the reference experiment that are resolved into two or more peaks in the enhanced experiments. For these peaks, average values were calculated in the enhanced experiment. The 1-σ (68.8%) confidence interval of the ^{1}H direct dimensions for these two experiment pairs was calculated and used to estimate the error in **(A-D)**, using estimated linewidth fit errors (see **Figure 3.S4-3.S5**, Appendix B). The error bars are smaller than the markers in **(E-H)**.

An average increase in signal intensity was also observed for the cross-peaks in the srNOESY-HSQC spectra (**Figure 3.3E, G**). Similar to the resolution enhancements, the opposite effect is observed for the peaks along the diagonal (**Figure 3.3F, H**). The average intensity enhancement is 44 ± 1% for ubiquitin and 20 ± 2% for HAfp-bicelles, while the diagonal peaks display an average intensity reduction 26 ± 2% for ubiquitin and 29 ± 2% for HAfp-bicelles. The srNOESY-HSQC intensities are highly correlated to the reference NOESY-HSQC ($R^2 > 0.97$). The improved sensitivity of the srNOESY-HSQC revealed many new peaks in comparison to the reference NOESY-HSQC. The reference NOESY-HSQC had 428 cross-peaks for ubiquitin, whereas the srNOESY-HSQC had 635 cross-peaks—a total of 207 (48%) more peaks. The reference NOESY-HSQC for HAfp-bicelles had 119 cross-peaks, while the srNOESY-HSQC had 136 cross-peaks peaks—a total of 17 (14%) more peaks.

Most of the new peaks in the srNOESY-HSQC (162 for ubiquitin and 14 for HAfp-bicelles) can be identified from a conventional NOESY-HSQC with a longer mixing time. The remaining 45 new peaks for ubiquitin and 3 new peaks for HAfp-bicelles were resolved by the enhanced resolution of the srNOESY-HSQC (red points in **Figure 3.3**). The fewer number of new peak assignments is expected for HAfp-bicelles because HAfp is a small molecule with nearly all of its resonances already resolved, even though the HAfp complex with bicelles is a significantly larger system. New peaks represent either new assignments, or J_{HH}-couplings in the case of ubiquitin. Visible examples of the resolution enhancements for both cases can be seen in the matched contour plots for each protein (**Figure 3.4**). For ubiquitin, new assignments were found for V5 $H^{\gamma 1}$ and $H^{\gamma 2}$, G10 $H^{\alpha 2}$ and $H^{\alpha 3}$, L15 H^{α} and V17 H^{α}, K27 $H^{\delta 2}$ and $H^{\delta 3}$, and Q41 $H^{\beta 2}$ and $H^{\beta 3}$. Resolved peaks in the HAfp-bicelle system include G8 $H^{\alpha 2}$ and $H^{\alpha 3}$, I6 H^{β} and $H^{\gamma 12}$, M17 $H^{\gamma 2}$ and I18 $H^{\gamma 12}$. Matched contour plots, with accompanying ^1H cross sections, for all

occurrences of resolved peaks in the srNOESY-HSQC can be found in the SI (see **Figure 3.S6-3.S9**, Appendix B). Cross-peaks were assigned from literature chemical shifts and strips from the srNOESY-HSQC spectra [18], [19].

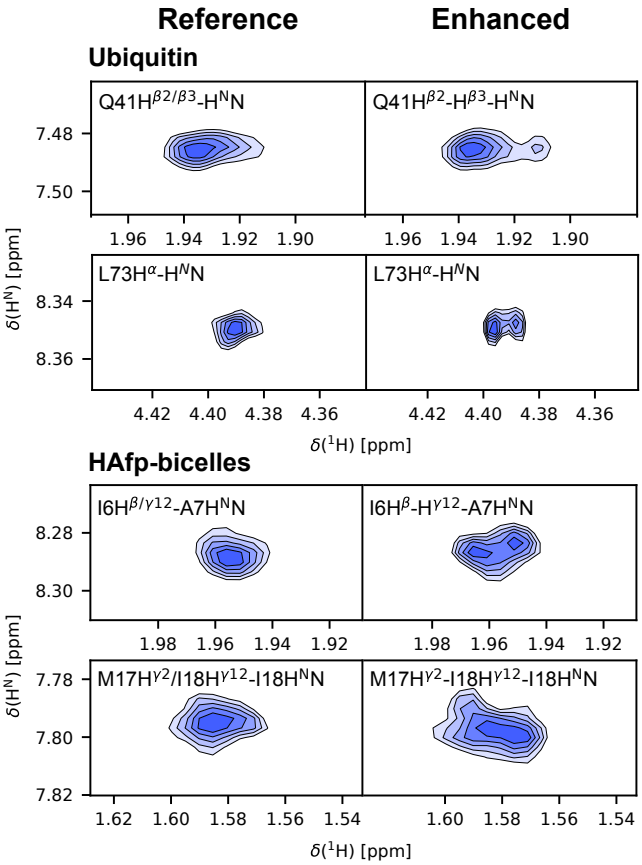


Figure 3.4. Contour plot comparison of reference NOESY-HSQC and srNOESY-HSQC. Selected contour plots of cross-peaks from the reference NOESY-HSQC (left) resolving into two peaks in the enhanced srNOESY-HSQC experiment (right). L73 H^{α} shows a resolved J_{HH}-coupling whereas other contour plots represent the deconvolution of two discrete spins. Contour plots were prepared by summing the ^{15}N dimension over its FWHH. The lowest contour represents the FWHH. The NOE peaks were identified from literature assignments [18], [19].

The distance restraints were calculated using calibration plots as described in the SI (see **Figure 3.S10**, Appendix B), and they were plotted to compare their accuracy with the reference experiment (**Figure 3.5A-B**). Interatomic ^{1}H-^{1}H distances from the reference NOESY-HSQC and enhanced srNOESY-HSQC are highly correlated ($R^2 > 0.99$), with linear regression slopes of 1.002 ± 0.003 for ubiquitin and 1.006 ± 0.008 for HAfp-bicelles. The conventional and enhanced experiments yield the same distance restraints.

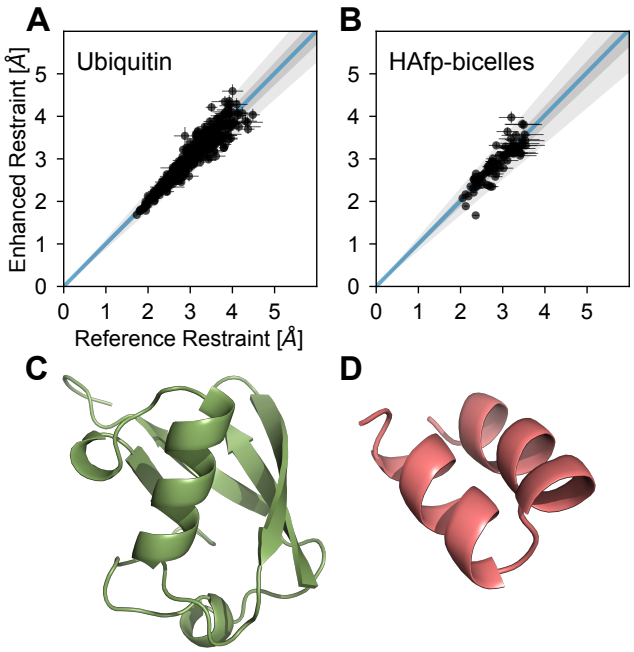


Figure 3.5. Distance restrain and structure comparison of the reference NOESY-HSQC and srNOESY-HSQC. Comparison of calculated distance restraints for the reference NOESY-HSQC and the enhanced srNOESY-HSQC for **(A)** ubiquitin and **(B)** HAfp-bicelles. Shaded regions represent the 1-σ (5.3% and 6.2%, ubiquitin and HAfp, dark gray) and 2-σ (12.2% and 16.1%, ubiquitin and HAfp, light gray) confidence intervals for the data. Error bars represent the error in the peak intensity propagated through the distance calibration. The solution NMR structures are shown for **(C)** ubiquitin and **(D)** HAfp generated with the inclusion of restraints from the enhanced srNOESY-HSQC. The original NMR restraints were obtained from the Protein Data Bank (PDB IDs: 1D3Z [18] and 2KXA [19]), and matched NOE restraints were replaced with those from the enhanced srNOESY-HSQC in an XPLOR-NIH refinement.

We refined the structures of ubiquitin and HAfp with the distance restraints from the srNOESY-HSQC (**Figure 3.5C-D**). The structures are superimposable to the published structures. The backbone heavy-atom root mean square deviations (RMSDs) between our refined structures and the previously published structures are 0.24 Å and 0.10 Å for ubiquitin and HAfp, respectively [18], [19]. The refinement statistics show that structures calculated from srNOESY-HSQC restraints have a comparable accuracy to published values (see Tables 3.S1 and 3.S2, Appendix B).

3.5 srNOESY Discussion

3.5.1 Optimal Parameters and Molecular Size

According to equation (16), the reduction in peak linewidth and the degree of resolution enhancement depends on the a_0 and a_1 parameters as well as the ^{1}H R_1 rate of spins (**Figure 3.6**). Generally, larger systems have longer tumbling times and larger R_1 rates, and they require shorter a_0 and smaller a_1 parameters to achieve maximum resolution enhancement.

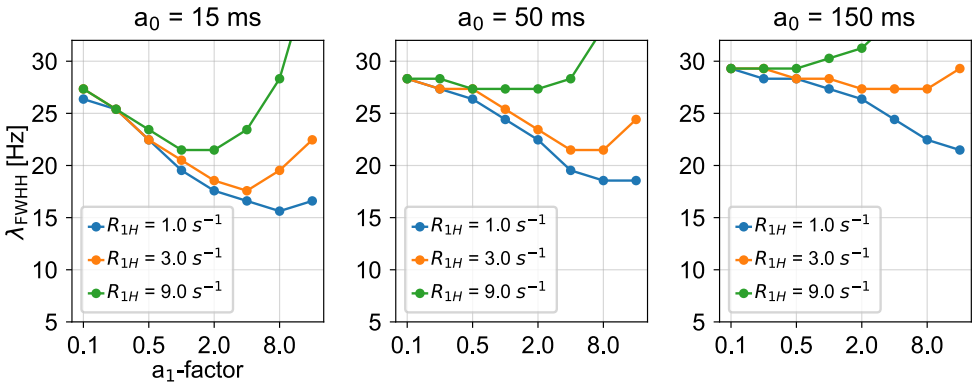


Figure 3.6. Simulation of the srNOESY cross-peak FWHH as a function of the a_1-factor, the effective ^{1}H R_1 and the NOESY mix time (a_0). The NOE rate was kept fixed at σ_{is} = 2.5 s^{-1} (r_{is} = 2.2 Å, τ_c = 4.9 ns) and the peaks were simulated with an R_2 of 30π s^{-1}.

Ideally, a resolution enhancement scheme would improve or remain unchanged as the size of the molecular system increases. Resolution enhancements were observed for the two protein systems in this study, but a less significant improvement was observed for the larger HAfp-bicelle system. The resolution enhancements are smaller for larger systems because the degree of enhancement does not depend on the tumbling time whereas the linewidths of cross-peaks are directly proportional to the tumbling time. For this reason, the degree of enhancement is reduced for larger systems where it would be most useful. In our case, the reduction in average linewidth is reduced from 31 ± 1% for ubiquitin to 18 ± 2% for HAfp-bicelles. This reduction in resolution enhancement can be attributed to the increase in the R_1 rate, which also increases proportionally with the size of the molecular system.

The simulations in **Figure 3.7** show the optimal a_0 and a_1 parameters for different rotational tumbling times, τ_c. For the best resolutions, the a_0 parameter should be as short as possible. However, shorter a_0 values also reduce the intensity of cross-peaks. Nevertheless, shorter a_0 values in the srNOESY experiment can be used in comparison to the t_{mix} times in a conventional NOESY because cross-peaks have greater intensity in the srNOESY experiment.

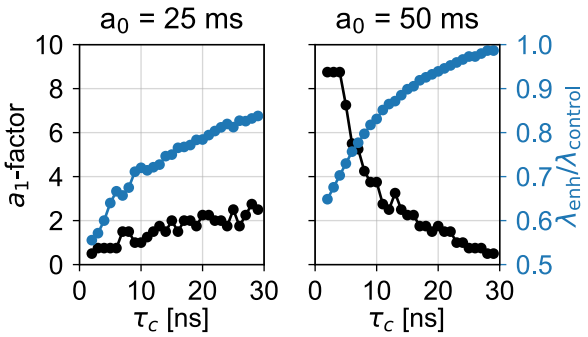


Figure 3.7. Simulated optimal a_1-factors and linewidth reductions at the FWHH as a function of the molecular tumbling time, τ_c. Simulations were conducted for an a_0 of 25 ms and 50 ms, which are optimal for large and mid-size molecules, respectively. Simulations were conducted using equations (4), (5), (9) and (12). The ^{1}H R_1 was made equal to the modeled NOE, σ_{is}, and the ^{1}H R_2 was made equal to 2.5 σ_{ij} π s^{-1}.

In selecting optimal a_0 and a_1 parameters, we suggest using an a_0 value that is short enough to give good cross-peak intensity (5-20% of the diagonal intensity) and still in the NOE linear build-up regime, then selecting the optimal a_1 parameter for a given τ_c with **Figure 3.7** as a guide. The selection of an a_0 parameter that is too long and near the NOE intensity maximum will produce cross-peaks with linewidths that are *larger* than the conventional NOESY experiment, according to equation (16).

For molecules with tumbling times up to 10 ns (*ca.* 24 kDa), an a_0 of 50 ms can be used with an a_1 parameter selected from **Figure 3.7**. For molecules with tumbling times of 10-20 ns (up to *ca.* 58 kDa), an a_0 of 25 ms will produce greater resolution enhancements. Shorter a_0 values are more complementary to larger molecular systems because the NOE transfer is much more efficient for larger τ_c values (equation (4)).

The reduced resolution enhancement for larger systems could also be circumvented, in part, with partial deuteration. Partial deuteration decreases the density of ^{1}H spins, thereby reducing the linewidths of ^{1}H peaks as well as spin diffusion pathways [38]. The principal drawback to partial deuteration is a reduction in ^{1}H signal intensity. However, the lower abundance of ^{1}H spins is partially compensated by the increase in intensity from sharper peaks [39]. The srNOESY experiments would achieve an additional resolution enhancement through a reduction in the ^{1}H R_1 rate. A smaller ^{1}H R_1 rate would enable larger a_0 and a_1 parameters and greater resolution enhancements.

3.5.2 Information Content of the srNOESY Lineshape.

The srNOESY cross-peak does not contain more information than the NOESY cross-peak at maximum intensity. **Figure 3.8A** shows the predicted cross-peak buildup using equation (6). A short mixing time (t_{short}, blue dashed lines) is typically selected in a conventional NOESY

since the cross-peak intensity more accurately encodes the distance between two ^{1}H spins [40]. At longer mixing times, the NOE cross-peak reaches a maximum intensity (t_{max}, orange dashed lines), and the peak intensity encodes the R_1 of the spin. The corresponding FIDs (**Figure 3.8B**) in t_1 are shown in panel B. Without noise, the information content of the FID at t_{short} (blue) and the FID at t_{max} (orange) would be the same. However, both FIDs are subject to experimental noise, and consequently, the FID at t_{max} contains more lineshape information. This can be seen in the larger amplitude oscillations later in the FID.

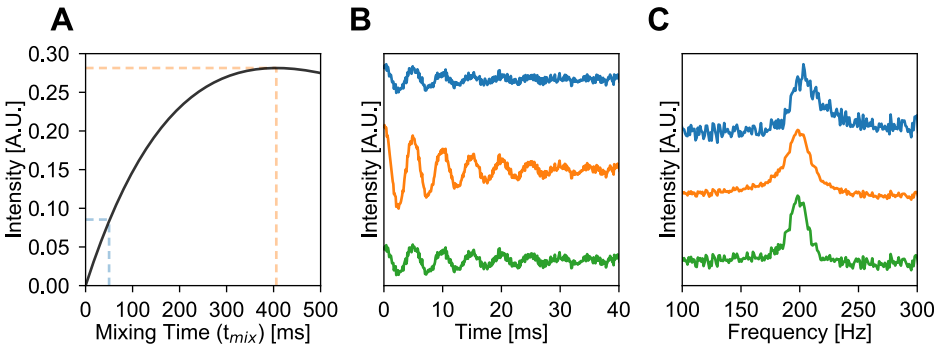


Figure 3.8. Comparison of the intensities and FIDs for NOESY cross-peaks at a short mixing time (blue), a cross-peak at a mixing time with maximum intensity (orange) and a srNOESY cross-peak (green). (**A**) Simulation of the cross-peak build-up as a function of the NOE mixing time for an H^N-$H^{\alpha 1}$ cross-peak modeled from a glycine in a helix for a protein with a τ_c = 4.9 ns. The NOE rate (σ_{ij}) is 1.9 s^{-1} and the ^1H R_1 (ρ_{ii}) is 2.5 s^{-1}. The NOE build-up was modeled with equation (**7**). (**B**) The simulated FIDs in the t_1 dimension for the cross-peak at a short mixing time (blue, t_{short} = 50 ms), the cross-peak with maximum intensity (orange, t_{max} = 406 ms) and the srNOESY crosspeak (green, a_0 = 50 ms and a1 = 4). A random Gaussian noise equal to 10% of the initial FID t_{max} intensity (1-σ) was added to each FID. (**C**) The corresponding Fourier transformed spectra of the simulated FIDs.

Likewise, the corresponding srNOESY cross-peak (green) encodes the distance between spins in its intensity, yet it also contains much of the resolution information from the FID at t_{max}, with the larger amplitude oscillations later in the FID. The srNOESY cross-peak contains more lineshape information than the FID at t_{short}, yet it still contains less information than the FID with t_{max}.

3.5.3 Comparison to apodization.

Apodization is the process of scaling an FID signal with a function to emphasize different regions of the time-domain data. The scaling function is convolved in the Fourier transformed spectrum to change peak shapes and emphasize either the signal-to-noise or the sharpness of peaks. Apodization functions generally fall under two classes: signal enhancing, using functions like an exponential decay or a Gaussian function, and "resolution enhancing," using Lorentz-to-Gauss window functions or the first lobe of an offset sine-bell function. The former sacrifices peak widths for an increased signal-to-noise ratio whereas the latter sacrifices signal-to-noise to improve peak linewidths. In either case, apodization only impacts the *appearance* of the Fourier Transformed spectrum.

As long as noise is non-deterministic in an existing dataset, mathematical operations, including apodization, cannot introduce new information in the dataset. "Resolution enhancing" apodization is a misnomer since the resolution of peaks, to potentially introduce new peaks, is not achieved. A more accurate term would be "peak sharpening" apodization. An experimental procedure only enhances the resolution if it can resolve new features in the dataset. For example, a mathematical procedure that replaces peaks with delta-functions may appear to have infinite resolution, yet no new peaks are resolved by this process.

Conceptually, the resolution enhancement of the srNOESY experiment may appear analogous to peak sharperning apodization. However, the srNOESY experiment will contain more lineshape information than the conventional NOESY experiment if both are collected in the linear, short mixing time regime. To get analogous spectral information from a conventional NOESY, the NOESY must be collected near the cross-peak intensity maximum with an aggressive peak sharpening apodization scheme. The drawback of this approach is that the cross-peak intensity does not accurately encode the distances between ^{1}H spins.

Figure 3.9 demonstrates contour plots for the lineshape of the conventional NOESY and the srNOESY with a short mixing time (t_{mix} and a_0 of 50 ms) and a conventional NOESY collected with t_{mix} closer to the NOE maximum (t_{mix} of 250 ms). The conventional NOESY with a longer mixing time and aggressive apodization was able to resolve 18 of the 45 newly resolved peaks from the srNOESY experiment (see **Figure 3.S6**, Appendix B).

For D21H$^{\beta 2}$, the peak appears broadened in the conventional NOESY with a short t_{mix}. The srNOESY resolves the splitting of the two peaks, which is also recovered in the conventional NOESY with a long mixing time and aggressive peak sharpening apodization. By contrast, L73H$^{\alpha}$ remains a singlet in both conventional NOESY experiments while the srNOESY is able to resolve two peaks for this assignment. This is likely due to a reduced $R_{1,H}$ for this spin, as it is in the dynamic C-terminal tail of ubiquitin. In this case, the t_{mix} would have to be increased substantially from 250 ms to resolve the doublet in the conventional NOESY. However, increasing the t_{mix} could decrease the intensity of other cross-peaks, and it would further reduce the accuracy of the NOEs. Moreover, different peaks with different $R_{1,H}$ rates will experience maxima at different t_{mix} values, and a single t_{mix} may not be used to achieve the best resolution of all peaks.

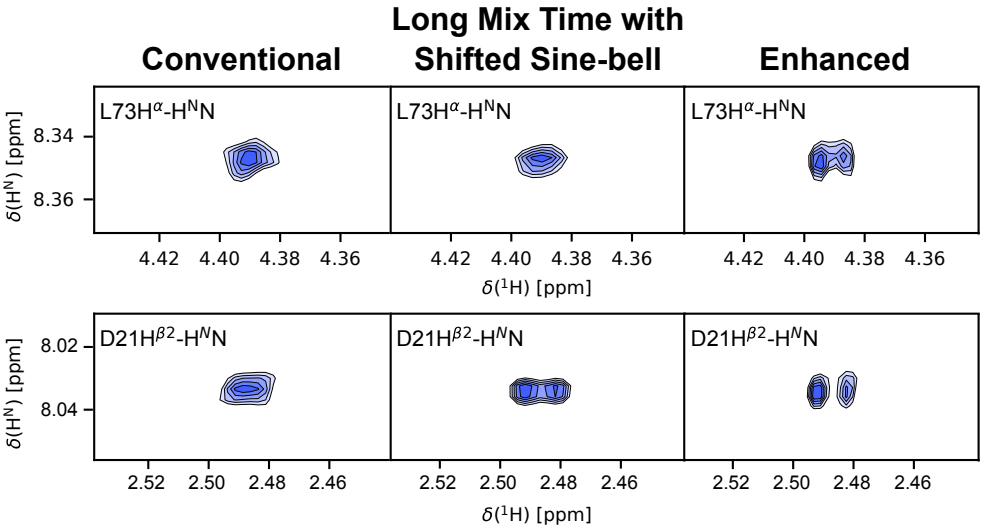


Figure 3.9. Contour plot comparison of reference NOESY-HSQC with aggressive apodization and srNOESY-HSQC. Contour plots from a conventional NOESY-HSQC (a_0 = 50 ms), a conventional NOESY-HSQC (a_0 = 250 ms) with an extremely shifted sine-bell window apodization function, and a srNOESY-HSQC (a_0 = 50 ms, a_1 = 4). Contour plots are presented from the ubiquitin datasets with otherwise matched experimental conditions. From left to right for L73 H^{α}, the FWHH of the peaks are 27.3, 24.1, 12.6 and 10.3 Hz. From left to right for D21 $H^{\beta 2}$, the FWHH of the peaks are 30.1, 20.8, 13.6, 19.9 and 11.2 Hz. The a_0 = 50 ms conventional and enhanced spectra were processed with a first-order sine-bell window with initial value of 0.45π, at 0.5 intensity, and a final value of 0.98π. The a_0 = 250 ms conventional spectrum was apodized using a first-order sine-bell window with initial value of 0.375π, at 0.25 intensity, and a final value of 1.0π. All apodization functions were applied using the NMRPipe software package [21].

3.5.4 Lineshape and the Effect of Spin Diffusion.

Spin diffusion adversely impacts the accuracy of NOE distances, and it is manifested at longer NOE mixing times. We characterized the extent of spin diffusion in our data by plotting the relationship between the cross-peak intensity and the internuclear distance (see **Figure 3.S10**, Appendix B) from reference structures [18], [19].

$$\log(\Delta S) = -m \log r_{ij} + C \tag{20}$$

In the absence of spin diffusion, the slope of the plot, m, should have a value of -6, following equation (2). Spin diffusion and relaxation increase the value of this slope. For cross-peaks with high intensity and short r_{ij} distances, the contribution from the direct 2-spin transfer tends dominate the cross-peak intensity. For more distant spins, the cross-peak intensity is much smaller, and spin diffusion pathways contribute more intensity to the cross-peak relative to the direct 2-spin transfer. Consequently, distant spins appear to have shorter internuclear distances and the experimental slope in equation (20) is more positive than -6. With ubiquitin as an example, which has a much larger number of cross-peaks, we calculated a slope 'm' of -3.6 ± 0.1 from the NOESY-HSQC and a slope of -3.4 ± 0.1 from the srNOESY-HSQC. The more positive slope of the srNOESY-HSQC indicates that the internuclear distances are subject to spin diffusion contamination to a slightly greater extent. This effect is unsurprising because the srNOESY-HSQC increases the mixing time throughout the experiment. The contribution of spin diffusion can be corrected using this procedure.

Additionally, the impact of spin diffusion was not directly observed in the peak line-shapes from the srNOESY-HSQC. In theory, cases with strong spin diffusion pathways may appear distorted if a short a_0 parameter is selected. A simulated example is presented in **Figure**

3.10. The blue FID and FT spectrum represents the srNOESY cross-peak for a linear NOESY buildup, characteristic of a 2-spin transfer, and the green FID and FT spectrum represents a cubic NOESY buildup, characteristic of spin diffusion. The spin-diffusion FID increases significantly in intensity before its decay, producing a lineshape with a distorted baseline. Peaks near this distorted peak will have a diminished intensity.

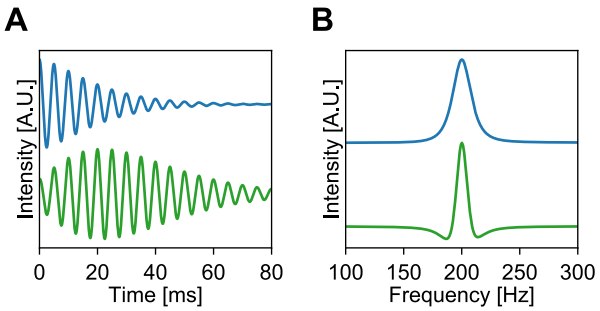


Figure 3.10. Effects of a_1 on srNOESY cross-peak lineshape. Comparison of the simulated (**A**) srNOESY cross-peak FID and (**B**) corresponding FT spectra for srNOESY cross-peaks with a linear NOE buildup (blue) and a cubic (green) NOE buildup.

The srNOESY-HSQC peaks do not appear to have this distortion (see **Figures 3.S8-3.S9**, Supplementary Information). Though we did not observe these distortions in our spectra, their appearance would be avoided by increasing the a_0 parameter in the experiment.

3.6 Conclusions of the srNOESY Study

We have shown that a simple modification to the NOESY pulse sequence can improve the resolution of spectra while maintaining the accuracy of distances from a conventional NOESY experiment. Resonances are resolved with reduced linewidths, and new peaks can be identified. The resulting spectra are greatly enhanced, without the use of costly isotopic labeling schemes. The enhancement is more modest for larger systems, yet the srNOESY experiment still presents a useful increase in resolution for both small and large systems to resolve new peaks. The srNOESY pulse program is simple to implement, effective for fully protonated molecules and readily applied with current technology.

3.7 Combination Approach: srNOESY and Random Fractional Deuteration

3.7.1 Study Aims

In order to improve the scope of the srNOESY experiment, we applied a random fractional deuteration strategy that was applied previously to 2D ^{1}H-NOESY experiments by LeMaster and Richards [7]. The scheme reduces the number spin diffusion pathways and disfavored ^{1}H-^{1}H dipolar relaxation effects that attenuate the FID. In this regard, fractional deuteration alone will reduce the ^{1}H spectral linewidths directly in both the traditional NOESY and srNOESY by reducing ^{1}H R_2 rates across the molecule. Similarly, R_1 (spin-lattice) relaxation

rates should also be reduced by effectively insulating ^{1}H nuclei from the bulk lattice with neighboring ^{2}H nuclei [41]. With the contribution of R_1 from the NOE build-up included in the signal decay of the srNOESY experiment, an additional reduction in ^{1}H linewidth should be achievable beyond what we were able to measure in uniformly protonated protein systems.

3.7.2 Random Fractional Deuteration Materials and Methods

Expression of random fractionally deuterated, uniformly ^{15}N-labeled ubiquitin ($^{2}H_x$,^{15}N-Ubq) was achieved by growing transformed *E. coli* BL21 (DE3) cells in 1 L of M9 minimal media [42] supplemented with 1.5 g $^{15}NH_4Cl$, 2.0 g ^{2}H-glucose, and 50% D_2O. Cells were grown at 37 °C and 200 rpm agitation until optical density at 600 nm reached 0.7. Induction of protein expression was achieved with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; GoldBio) for a period of 4 hours at 37 °C and 200 rpm. Cells were harvested by centrifugation at 6 000 x *g* at 4 °C for 30 minutes, flash frozen with liquid nitrogen, and stored at -80 °C until purification. Purification of $^{2}H_x$,^{15}N-Ubq was achieved as described in **section 3.2.1**.

A matched pair of NOESY-HSQC and srNOESY-HSQC experiments were collected using the same parameters as the uniformly ^{15}N-labeled ubiquitin experiments (**section 3.2.2**) on the same Bruker Avance-III 750 MHz wide-bore spectrometer equipped with a $^{1}H/^{13}C/^{15}N$ TXI 5 mm room temperature probe. However, the loss of a large number of ^{1}H neighboring spin pathways required the use of longer mixing times in the control experiment to achieve detectable NOEs. An a_0 parameter of 100 ms was sufficient to observe all of the $^{2}H_x$,^{15}N-Ubq cross-peaks and was selected for both the control and srNOESY experiments. Both spectra were processed using NMRPipe [21] and analyzed using NMRFAM-Sparky [22], [23].

3.7.3 Results

We first wanted to validate the effectiveness of our fractional deuteration labeling strategy. Uniformly ^{15}N- and $^{2}H_{x}$,^{15}N-Ubq were compared using electrospray ionization mass spectrometry (ESI-MS). Despite peak broadening from isotope distribution inequalities in the sample, a mass shift of approximately 315 Da was observed (**Figure 3.11**). There are 592 non-exchangeable ^{1}H atoms per ubiquitin molecule, indicating a 53% ^{2}H labeling efficiency on average with our protocol. Although this technique does not provide information on the distribution of ^{2}H throughout the molecule, the overall ^{2}H content is roughly what was expected.

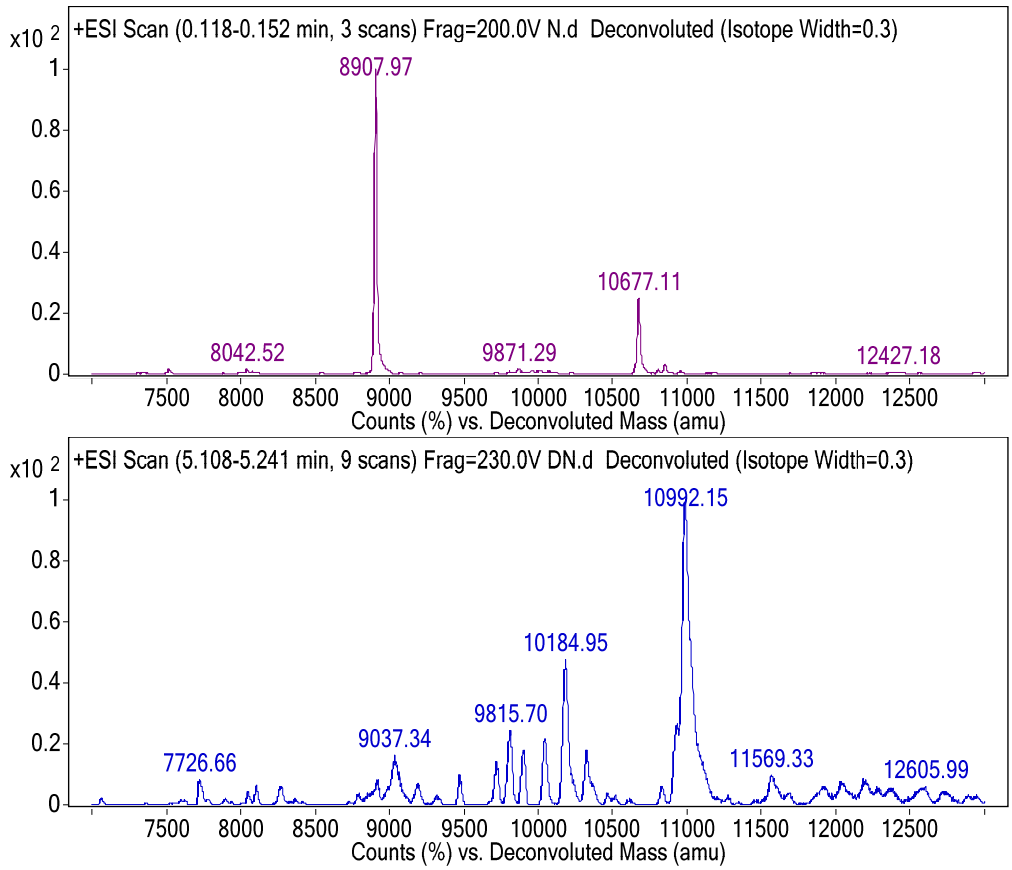


Figure 3.11. ESI mass spectrum of (top) ^{15}N-Ubq and (bottom) $^{2}H_x$,^{15}N-Ubq. M/Z peaks for the intact protein are 10677.11 Da for ^{15}N-Ubq and 10992.15 Da for $^{2}H_x$,^{15}N-Ubq. A mass difference of 315.04 Da indicates a 53% ^{2}H labeling efficiency by the 592 non-labile ^{1}H content in the uniformly protonated ^{15}N-Ubq sample. Spectra were recorded on a Bruker AmaZon-X ESI-MS attached to an Agilent 1200 series LC system.

In comparison to the control NOESY-HSQC, the srNOESY-HSQC produced spectra with an average reduction of 29 ± 1% in the linewidth of the indirect ^{1}H dimension for $^{2}H_{x}$,^{15}N-Ubq (**Figure 3.12A**). This corresponds to a resolution enhancement of 41 ± 2%. The peak intensities were enhanced by 53 ± 1% and maintained a very strong correlation coefficient with $R^2 = 0.972$ (**Figure 3.12B**). The resolved J_{HH}-couplings that had previously been observed in the ^{15}N-Ubq srNOESY-HSQC experiment were no longer apparent, which was expected of a 53% ^{2}H-labeled sample. The remaining newly resolved spins were already apparent in the control NOESY-HSQC experiment as an effect of the deuterium linewidth sharpening. Comparing the ^{15}N-Ubq and $^{2}H_{x}$,^{15}N-Ubq control NOESY-HSQC linewidths gave a 16% reduction in linewidths as a result of the fractional deuteration labeling alone (data not shown).

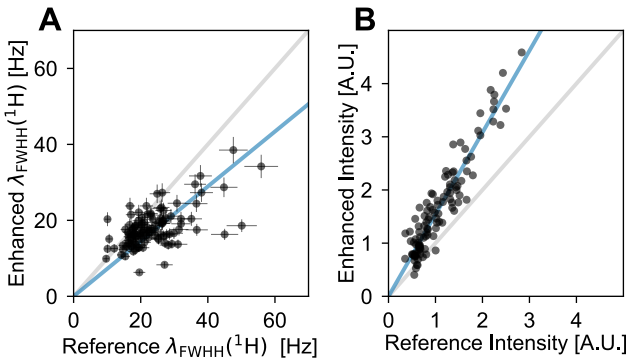


Figure 3.12. Linewidth and intensity comparison plots for a reference NOESY-HSQC and a srNOESY-HSQC of ^2H_x,^{15}N-Ubq. Comparison plots of the (**A**) ^{1}H indirect dimension FWHH linewidths and (**B**) signal intensities between a reference NOESY-HSQC and an enhanced srNOESY-HSQC of ^2H_x,^{15}N-Ubq. Visual guide 1:1 lines are shown in gray, and the linear regression lines (intercept=0) are shown in blue. The error bars in (**A**) were estimated from a 1-σ confidence interval, and the error bars in (**B**) are smaller than the markers. The intensities in (**B**) have been scaled down by $1 \cdot 10^7$. The NMR sample contained 0.8 mM ^2H_x,^{15}N-Ubq, 10 mM sodium phosphate, 7% D_2O, and 0.03% NaN_3. An a_0 of 100 ms and an a_1 of 8 was used for these experiments. All experimental parameters apart from the srNOESY a_1 factor were matched between experiments.

3.7.4 Conclusions and Future Directions

Using fractional deuteration in combination with the srNOESY experiment achieved a linewidth reduction of 29 ± 1% for $^{2}H_{x}$,^{15}N-Ubq. This was unfortunately lower than expected, and roughly the same, within error, to the enhancement observed on a fully protonated ^{15}N-Ubq sample. However this experiment was a first thorough analysis, and it is very likely that better resolution boosts can be achieved with optimized a_0 and a_1 parameters on fractionally deuterated systems. Similarly, I would expect that the combination of the deuteration content and the size of the protein would dictate specific requirements for the NOE mixing time parameters.

In contrast, the NOE intensity comparison showed a 53% increase for the fractionally deuterated srNOESY experiment compared to a 45% increase that was observed with the fully protonated srNOESY. This increase leads me to believe that larger a_0 or a_1 parameters could have achieved a more significant resolution and intensity boost. In our previous comparisons with fully protonated protein samples, the signal intensity and resolution would both begin to decrease if the mix time parameters were set at too large of a value. Additional comparisons will have to be performed to fully understand the effect of tandem srNOESY and fractional deuteration.

3.8 Bibliography

[1] J. E. Ollerenshaw, V. Tugarinov, and L. E. Kay, "Methyl TROSY: Explanation and experimental verification," *Magn. Reson. Chem.*, vol. 41, no. 10, pp. 843–852, 2003.

[2] V. Tugarinov and L. E. Kay, "Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins," *ChemBioChem*, vol. 6, no. 9, pp. 1567–1577, 2005.

[3] C. N. Chi, D. Strotz, R. Riek, and B. Vögeli, "NOE-Derived Methyl Distances from a 360 kDa Proteasome Complex," *Chem. - A Eur. J.*, vol. 24, no. 9, pp. 2270–2276, 2018.

[4] R. Otten, B. Chu, K. D. Krewulak, H. J. Vogel, and F. A. A. Mulder, "Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins," *J. Am. Chem. Soc.*, vol. 132, no. 9, pp. 2952–2960, 2010.

[5] K. Pervushin, R. Riek, G. Wider, and K. Wuthrich, "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.*, vol. 94, no. 23, pp. 12366–12371, Nov. 1997.

[6] J. Weigelt, "Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments," *Journal of the American Chemical Society*, vol. 120, no. 41. UTC, pp. 10778–10779, 1998.

[7] D. M. LeMaster and F. M. Richards, "NMR Sequential Assignment of Escherichia coli Thioredoxin Utilizing Random Fractional Deuteriation," *Biochemistry*, vol. 27, no. 1, pp. 142–150, 1988.

[8] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schulz, and D. E. Wemmer, "Site-specific isotopic labeling of proteins for NMR studies," *J. Am. Chem. Soc.*, vol. 114, no. 20, pp. 7959–7961, 1992.

[9] J. L. Battiste and G. Wagner, "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data," *Biochemistry*, vol. 39, no. 18, pp. 5355–5365, 2000.

[10] G. W. Vuister and A. Bax, "Resolution Enhancement and Spectral Editing of Uniformly 13C-Enriched Proteins by Homonuclear Broadband 13C Decoupling," *J. Magn. Reson.*, vol. 98, no. 2, pp. 428–435, Jun. 1992.

[11] B. Vögeli, M. Friedmann, D. Leitz, A. Sobol, and R. Riek, "Quantitative determination of NOE rates in perdeuterated and protonated proteins: Practical and theoretical aspects," *J. Magn. Reson.*, vol. 204, no. 2, pp. 290–302, 2010.

[12] W. Rieping, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges, "ARIA2: Automated NOE assignment and data integration in NMR structure calculation," *Bioinformatics*, vol. 23, no. 3, pp. 381–382, Feb. 2007.

[13] S. Macura, K. Wuthrich, and R. R. Ernst, "Separation and suppression of coherent transfer effects in two-dimensional NOE and chemical exchange spectroscopy," *J. Magn. Reson.*, vol. 46, no. 2, pp. 269–282, 1982.

[14] S. Macura, K. Wüthrich, and R. R. Ernst, "The relevance of J cross-peaks in two-dimensional NOE experiments of macromolecules," *J. Magn. Reson.*, vol. 47, no. 2, pp. 351–357, 1982.

[15] B. A. Borgias, M. Gochin, D. J. Kerwood, and T. L. James, "Relaxation matrix analysis of 2D NMR data," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 22, no. 1, pp. 83–100, Jan.

1990.

[16] J. D. Baleja, J. Moult, and B. D. Sykes, “Distance measurement and structure refinement with NOE data,” *J. Magn. Reson.*, vol. 87, no. 2, pp. 375–384, Apr. 1990.

[17] J. Marley, M. Lu, and C. Bracken, “A method for efficient isotopic labeling of recombinant proteins,” *J. Biomol. NMR*, vol. 20, no. 1, pp. 71–75, 2001.

[18] G. Cornilescu, J. L. Marquardt, M. Ottiger, and A. Bax, “Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase,” *J. Am. Chem. Soc.*, vol. 120, no. 27, pp. 6836–6837, Jul. 1998.

[19] J. L. Lorieau, J. M. Louis, and A. Bax, “The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 25, pp. 11341–11346, 2010.

[20] S. T. Smrt, A. W. Draney, and J. L. Lorieau, “The influenza hemagglutinin fusion domain is an amphipathic helical hairpin that functions by inducing membrane curvature,” *J. Biol. Chem.*, vol. 290, no. 1, pp. 228–238, 2015.

[21] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, “NMRPipe: A multidimensional spectral processing system based on UNIX pipes,” *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[22] W. Lee, M. Tonelli, and J. L. Markley, “NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy,” *Bioinformatics*, vol. 31, no. 8, pp. 1325–1327, 2015.

[23] T. D. Goddard and D. G. Kneller, “SPARKY 3,” *Univ. California, San Fransisco*, 2008.

[24] G. Lipari and A. Szabo, “Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results,” *J. Am. Chem. Soc.*, vol. 104, no. 17, pp. 4559–4570, 1982.

[25] R. Cole and J. P. Loria, “FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data,” *J. Biomol. NMR*, vol. 26, no. 3, pp. 203–213, 2003.

[26] M. W. Maciejewski *et al.*, “NMRbox: A Resource for Biomolecular NMR Computation,” *Biophys. J.*, vol. 112, no. 8, pp. 1529–1534, 2017.

[27] C. D. Schwieters, J. J. Kuszewski, and G. Marius Clore, “Using Xplor-NIH for NMR molecular structure determination,” *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 48, no. 1. pp. 47–62, Mar-2006.

[28] C. . Dobson, E. . Olejniczak, F. . Poulsen, and R. . Ratcliffe, “Time development of proton nuclear overhauser effects in proteins,” *J. Magn. Reson.*, vol. 48, no. 1, pp. 97–110, Jun. 1982.

[29] B. Vögeli *et al.*, “Exact distances and internal dynamics of perdeuterated ubiquitin from

NOE buildups," *J. Am. Chem. Soc.*, vol. 131, no. 47, pp. 17215–17225, 2009.

[30] N. A. Farrow *et al.*, "Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by15N NMR Relaxation," *Biochemistry*, vol. 33, no. 19, pp. 5984–6003, 1994.

[31] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," *J. Am. Chem. Soc.*, vol. 104, no. 17, pp. 4546–4559, Aug. 1982.

[32] C. R. C. and P.R.Schimmel, *Biophysical Chemistry, Part II: Techniques for the study of Biological Structure and Function*, 1st ed. New York: W.H. Freeman, 1980.

[33] A. J. Shaka, J. Keeler, T. Frenkiel, and R. Freeman, "An improved sequence for broadband decoupling: WALTZ-16," *J. Magn. Reson.*, vol. 52, no. 2, pp. 335–338, 1983.

[34] L. E. Kay, P. Keifer, and T. Saarinen, "Pure Absorption Gradient Enhanced Heteronucler Single Quantum Correlation Spectroscopy with Improved Sensitivity," *J. Am. Chem. Soc.*, vol. 114, pp. 10663–10665, 1992.

[35] R. Muhandiram and L. E. Kay, "Three-Dimensional HMQC-NOESY, NOESY-HMQC, and NOESY-HSQC," in *Encyclopedia of Magnetic Resonance*, no. i, 2007, pp. 1–11.

[36] J. Cavanagh, W. Fairbrother, A. Palmer III, M. Rance, and N. Skelton, *Protein NMR Principles And Practice*, 2nd ed. Burlington, MA: Academic Press, 2007.

[37] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 6th ed. New York: W. H. Freeman and Company, 2013.

[38] D. M. LeMaster, "Deuteration in protein proton magnetic resonance," in *Methods in Enzymology*, vol. 177, no. C, 1989, pp. 23–43.

[39] H. R. Kalbitzer, R. Leberman, and A. Wittinghofer, "1H-NMR spectroscopy on elongation factor Tu from Escherichia coli: Resolution enhancement by perdeuteration," *FEBS Lett.*, vol. 180, no. 1, pp. 40–42, 1985.

[40] D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed. New York: Wiley-VCH, 2000.

[41] A. Kalk and H. J. . Berendsen, "Proton magnetic relaxation and spin diffusion in proteins," *J. Magn. Reson.*, vol. 24, no. 3, pp. 343–366, Dec. 1976.

[42] F. C. Neidhardt, P. L. Bloch, and D. F. Smith, "Culture medium for enterobacteria.," *J. Bacteriol.*, vol. 119, no. 3, pp. 736–47, Sep. 1974.

Appendix A

Copyright and Permissions

Pro-Islet Amyloid Polypeptide in Micelles Contains a Helical Prohormone Segment
John Wiley & Sons (The FEBS Journal)

This work has been submitted to *The FEBS Journal* for publication on October 12th, 2019. The following excerpt from the *Author Guidelines* of the journal indicates permission for publication in a thesis prior to acceptance of the manuscript.

Submission Declaration

Submission of a manuscript implies: (1) that the work described has not been published before (except in the form of an abstract or as part of a lecture, review or thesis); (2) that it is not under consideration for publication elsewhere; (3) that its publication in the present form has been approved by all authors and by the responsible authorities in the institutions where the work was carried out; and (4) that, if accepted, it will not be published elsewhere in the same form, in any language, without the consent of FEBS, the licence holder. Prior or concurrent submission of the manuscript to an institutional repository or a not-for-profit subject-based preprint server does not constitute prior publication (see **Preprints**). Previously published abstracts, etc. should be referred to in the Introduction.

Super Resolution NOESY Spectra of Proteins
Springer Nature (Journal of Biomolecular NMR)

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Nov 15, 2019

This Agreement between Mr. Charles DeLisle ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4652630569025
License date	Aug 19, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Biomolecular NMR
Licensed Content Title	Super resolution NOESY spectra of proteins
Licensed Content Author	Charles F. DeLisle, H. Bhagya Mendis, Justin L. Lorieau
Licensed Content Date	Jan 1, 2019
Licensed Content Volume	73
Licensed Content Issue	3
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	yes
Title	Graduate Student
Institution name	University of Illinois at Chicago
Expected presentation date	Sep 2019

Mr. Charles DeLisle
466 W Elm St

Requestor Location

CHICAGO, IL 60610
United States
Attn: Mr. Charles DeLisle

Total 0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. **Grant of License**

    1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

    2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

    3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. **Scope of Licence**

    1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

    2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

    3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

    4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

    5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

- **Duration of Licence**

    1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

- **Acknowledgement**

1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

- **Restrictions on use**

1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

2. You must not use any Licensed Material as part of any design or trademark.

3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

- **Ownership of Rights**

1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

- **Warranty**

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER

CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- **Limitations**

1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

- **Termination and Cancellation**

1. Licences will expire after the period shown in Clause 3 (above).

2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:
Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix B

Super Resolution NOESY Spectra of Proteins

Chapter 3 Supporting Information

*Charles F. DeLisle, H. Bhagya Mendis, and Justin L. Lorieau**

Department of Chemistry, University of Illinois at Chicago, 845 W Taylor St, Chicago IL 60607

* Corresponding Author: justin@lorieau.com

Table 3.S1. Comparison of refinement statistics for ubiquitin with control and enhanced NOEs.

Ubquitin – 314 Restraints	**Control**[a]	**Enhanced**
Intra (i = j) – 161 Restraints		
Avg. NOE Viols. / Structure	0.6	2.8
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0.47	0.34
Avg. RMSD	0.026	0.059
Seq (i,j = 1) – 107 Restraints		
Avg. NOE Viols. / Structure	1.2	0.8
# Viols. > 0.5 A	2	1
Largest Viol. (A)	0.82	0.56
Avg. RMSD	0.047	0.041
Med (2 < i,j < 5) – 26 Restraints		
Avg. NOE Viols. / Structure	0	0
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0	0
Avg. RMSD	0.000	0.000
Long (2 < i,j < 5) – 20 Restraints		
Avg. NOE Viols. / Structure	0	0
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0	0
Avg. RMSD	0.021	0.008
Heavy Atom RMSD[b]	0.240 Å	

The control structure was 13DZ[1]
Backbone RMSD was calculated using residues 2-71

Table 3.S2. Comparison of refinement statistics for HAfp with control and enhanced NOEs.

HAfp23 – 82 Restraints	Control [a]	Enhanced
Intra (i = j) – 36 Restraints		
Avg. NOE Viols. / Structure	1	2
# Viols. > 0.5 A	2	3
Largest Viol. (A)	0.55	0.62
Avg. RMSD	0.075	0.111
Seq (i,j = 1) – 41 Restraints		
Avg. NOE Viols. / Structure	0	1.8
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0	0.43
Avg. RMSD	0.010	0.082
Med (2 < i,j < 5) – 3 Restraints		
Avg. NOE Viols. / Structure	0	0.2
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0	0.25
Avg. RMSD	0.000	0.095
Long (2 < i,j < 5) – 2 Restraints		
Avg. NOE Viols. / Structure	0	0
# Viols. > 0.5 A	0	0
Largest Viol. (A)	0	0
Avg. RMSD	0.019	0.028
Heavy Atom RMS	0.104 Å	

The control structure was 2KXA[2]

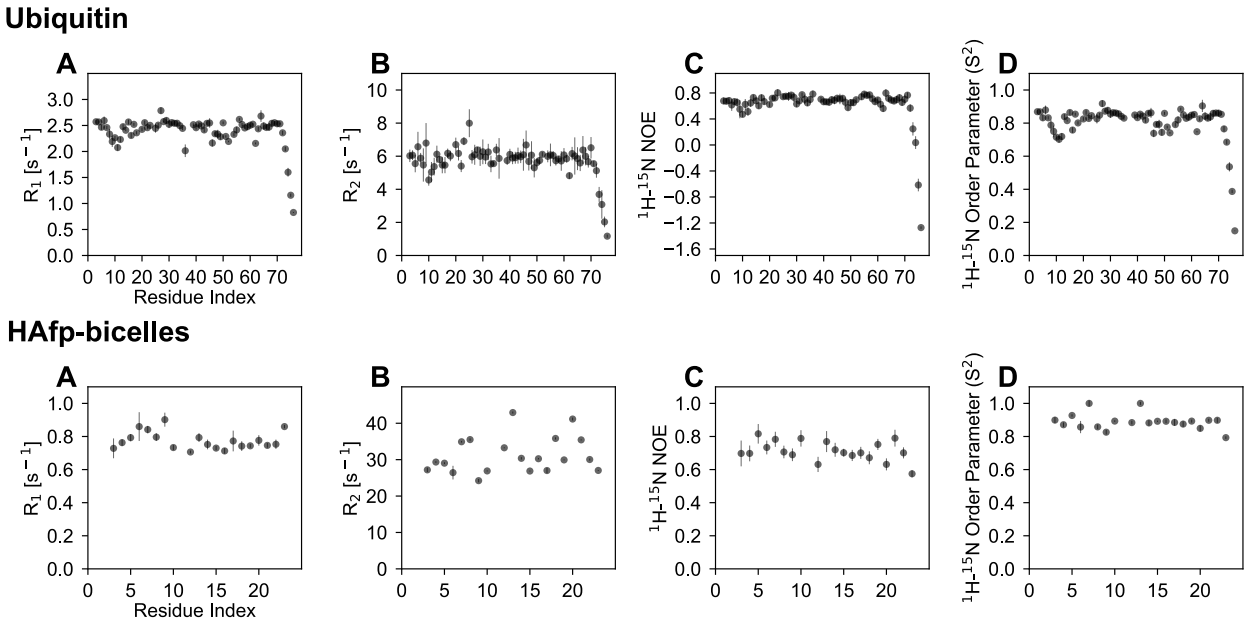


Figure 3.S1. ^{15}N relaxation rates for ubiquitin and HAfp-bicelles.

^{15}N (A) R_1, (B) R_2, (C) and ^{15}N-{^{1}H} NOE relaxation rates, and the resulting (D) order parameters from the Lipari-Szabo formalism for ubiquitin and HAfp-bicelles. The correlation times were determined to be 4.25 ± 0.03 ns and 18.95 ± 0.14 ns for ubiquitin and HAfp bound to ^{2}H-DMPC/DHPC bicelles, q = 0.44, respectfully .

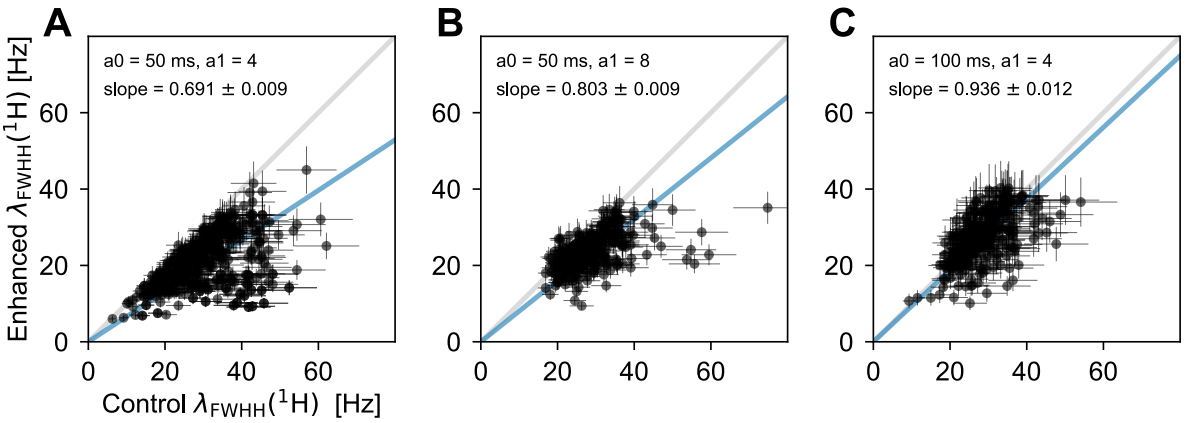


Figure 3.S2. Effects of base NOE mix time (a_0) and linear factor (a_1) on the ubiquitin indirect ^{1}H resolution.

An a_0 of 50 ms was employed in **(A)** and **(B)**, while 100 ms was used in **(C)**. An a_1 of 4 was employed in **(A)** and **(C)**, while 8 was used in **(B)**. A line of best-fit is shown in blue, and a 1:1 line is shown in gray. The average linewidth reduction was 31 ± 1%, 20 ± 1%, and 6 ± 1% for **(A)**, **(B)**, and **(C)**, respectively. Error bars represent the 1-σ confidence interval for the ^{1}H direct dimension for each corresponding experiment.

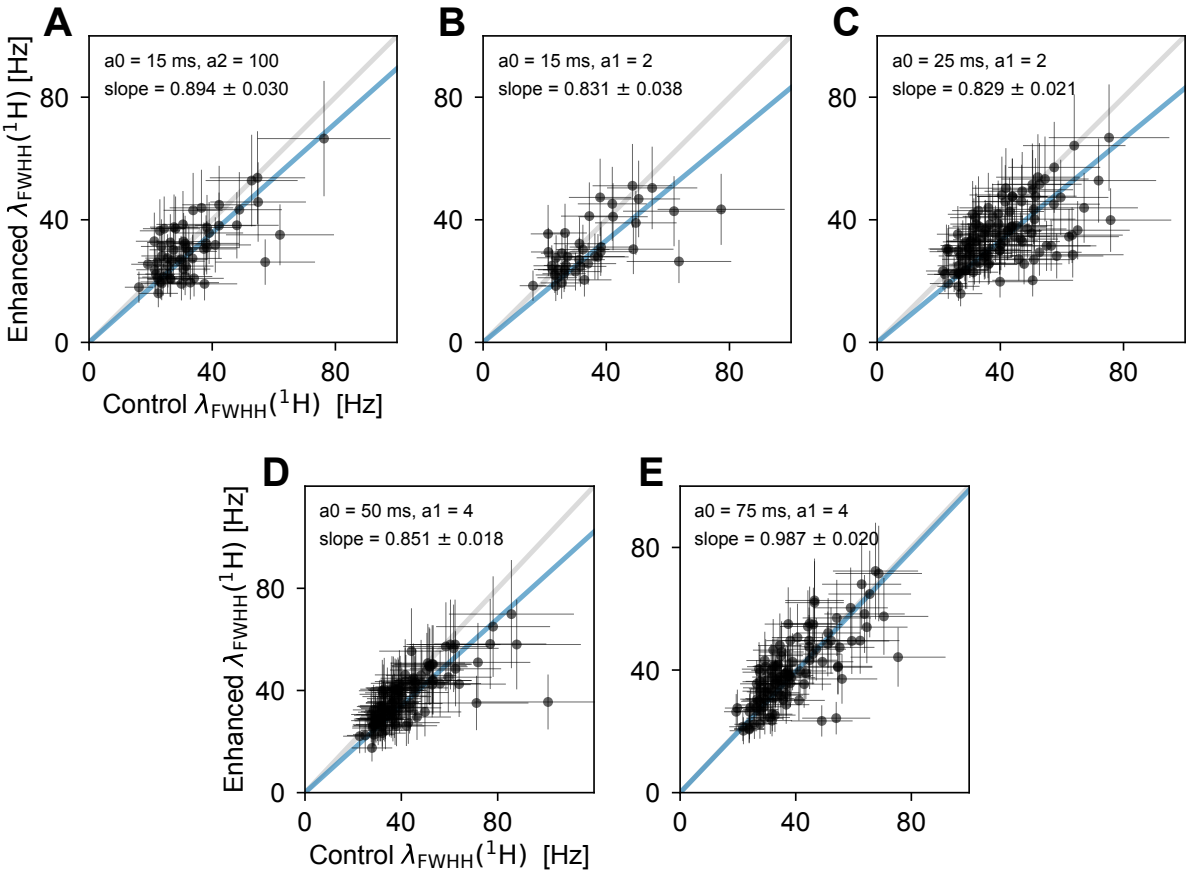


Figure 3.S3. Effects of base NOE mix time (a_0) and linear (a_1) or quadratic (a_2) factor on the HAfp-bicelle indirect ^1H resolution.

An **(A)** a_0 of 15 ms with an a_2 of 100 (t_1^2 * 100) and **(B)** a_0 of 15 ms with an a_1 of 2 (t_1 * 2) were compared to a 15 ms conventional NOESY-HSQC with matched experimental parameters to yield average linewidth reductions of 10.6 ± 3.0 and 16.9 ± 3.8%, respectfully. **(C)** An a_0 of 25 ms with an a_1 of 2 produced a similar resolution boost, within error, to the a_0 = 15 ms, a_1 = 2 experiment. An a_1 of 4 was also attempted with an a_0 of **(D)** 50 ms and **(E)** 75 ms, which gave average linewidth reductions of 14.9 ± 1.8 and 1.3 ± 2.0%, respectfully.

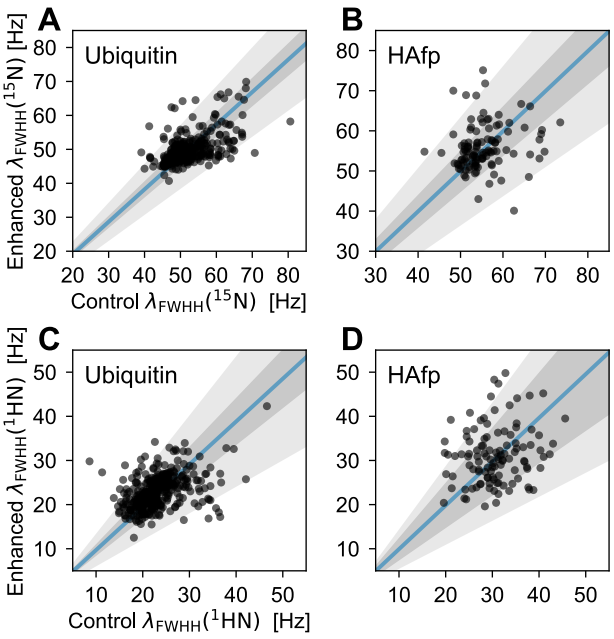


Figure 3.S4. Comparison of the fit linewidths for the indirect ^{15}N (A, B) and direct ^{1}H (C, D) dimensions for the conventional and enhanced srNOESY-HSQC for ubiquitin and HAfp-bicelles.

The gray shaded regions represent the 1-σ (68.5%) and 2-σ (95.5%) confidence intervals. For ubiquitin, the 1-σ and 2-σ confidence intervals are 6.3 and 19.7% for ^{15}N and 13.8 and 38.1% for ^{1}H. The slopes of the ubiquitin plots are 0.955 ± 0.004 and 0.998 ± 0.009 for ^{15}N and ^{1}H, respectively. For HAfp-bicelles, the 1-σ and 2-σ confidence intervals are 10.4 and 27.1% for ^{15}N and 26.0 and 45.1% for ^{1}H. The slopes of the HAfp-bicelle plots are 0.970 ± 0.012 and 0.991 ± 0.023 for ^{15}N and ^{1}H, respectively.

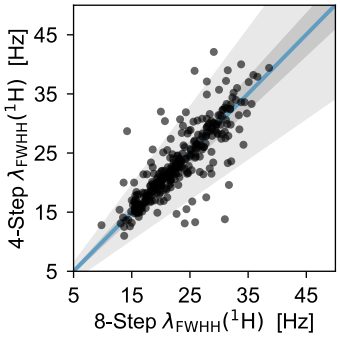
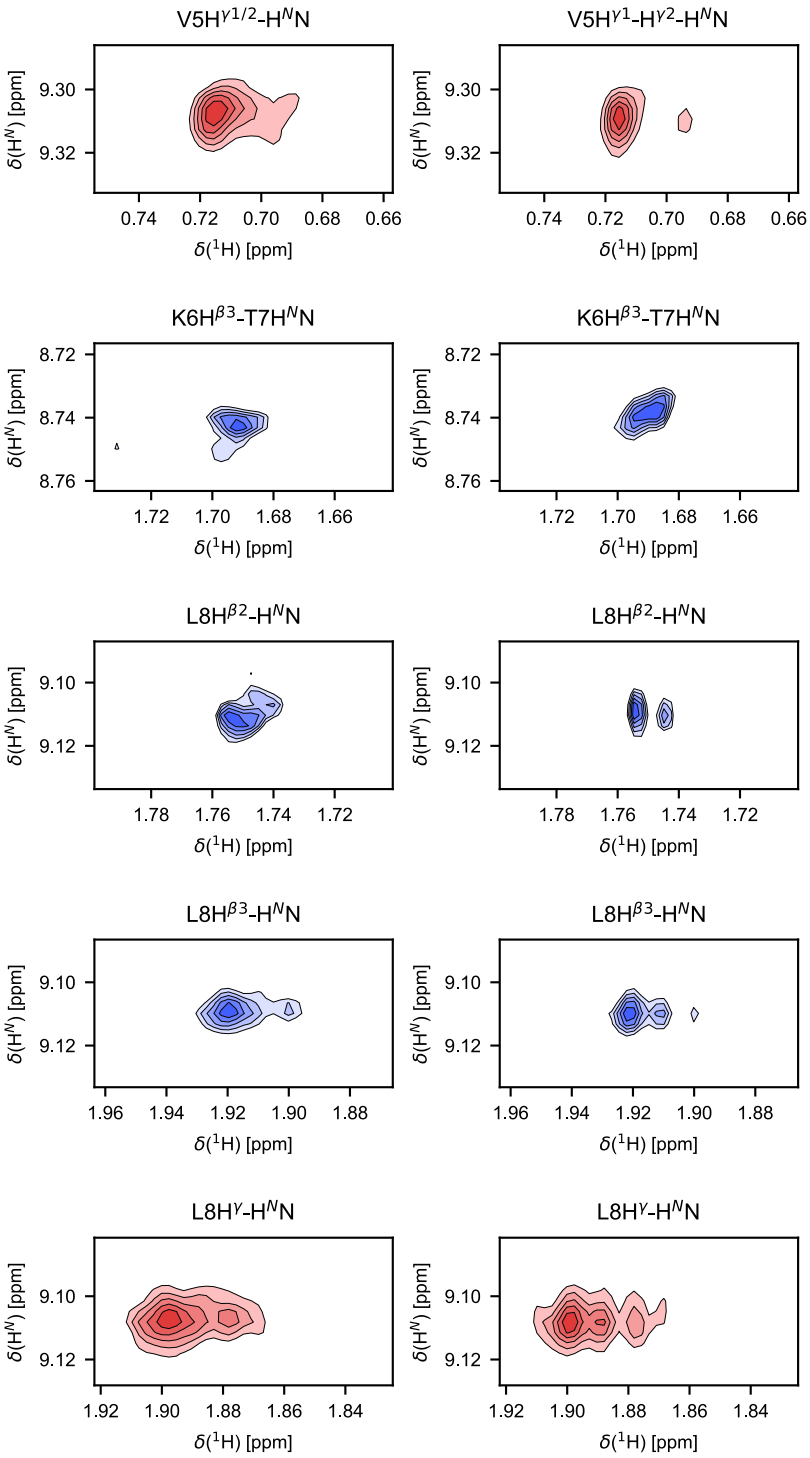
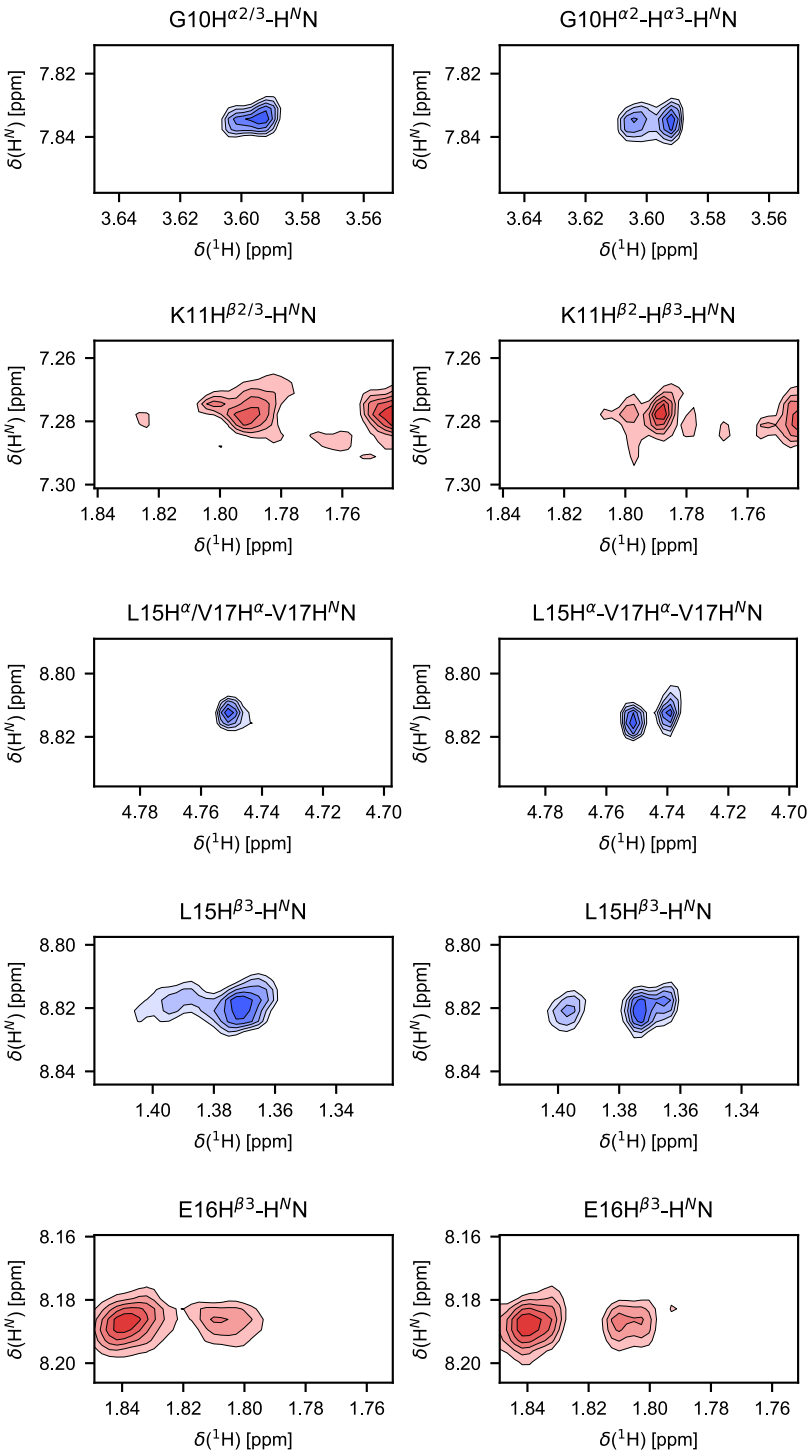
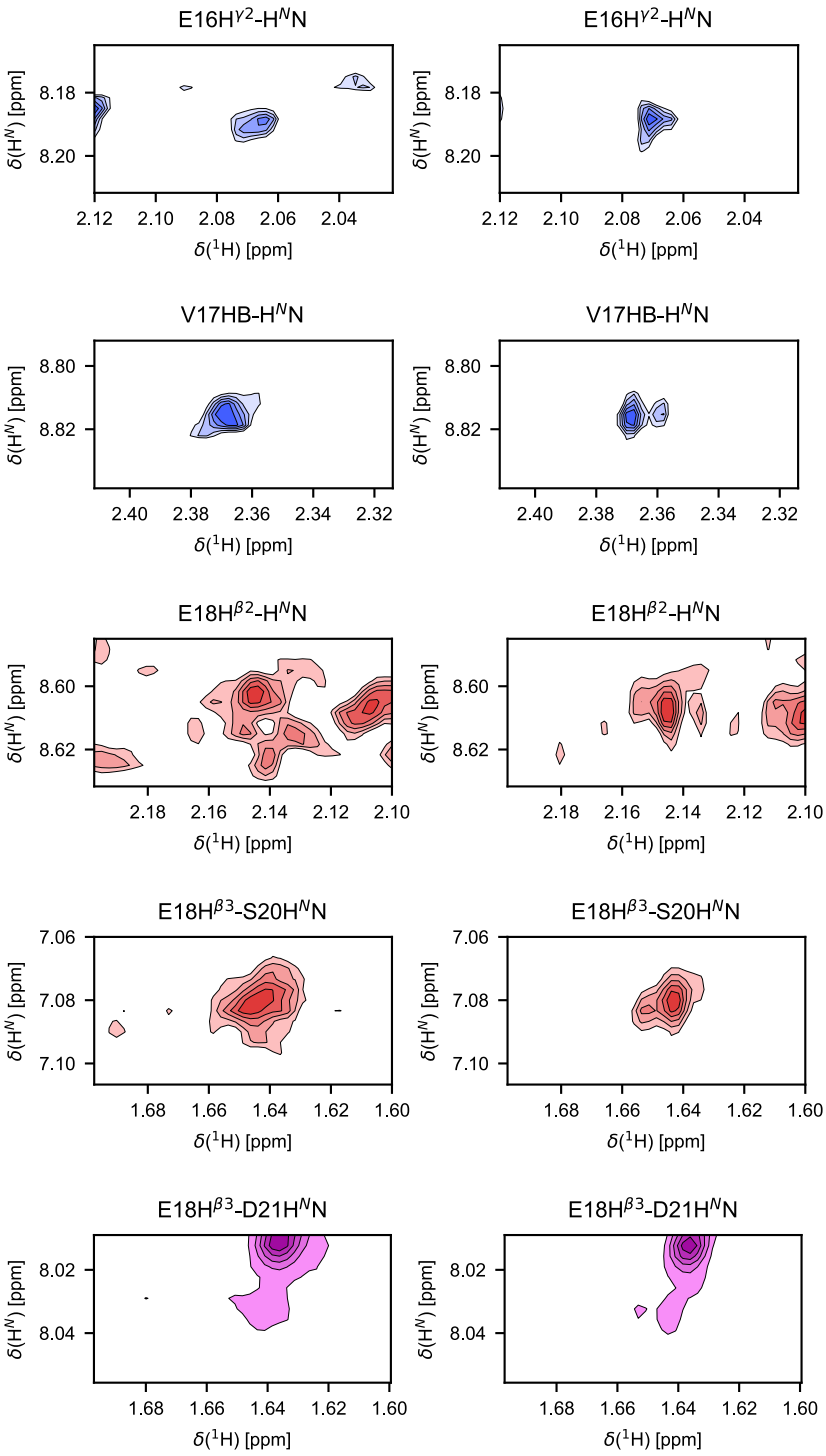


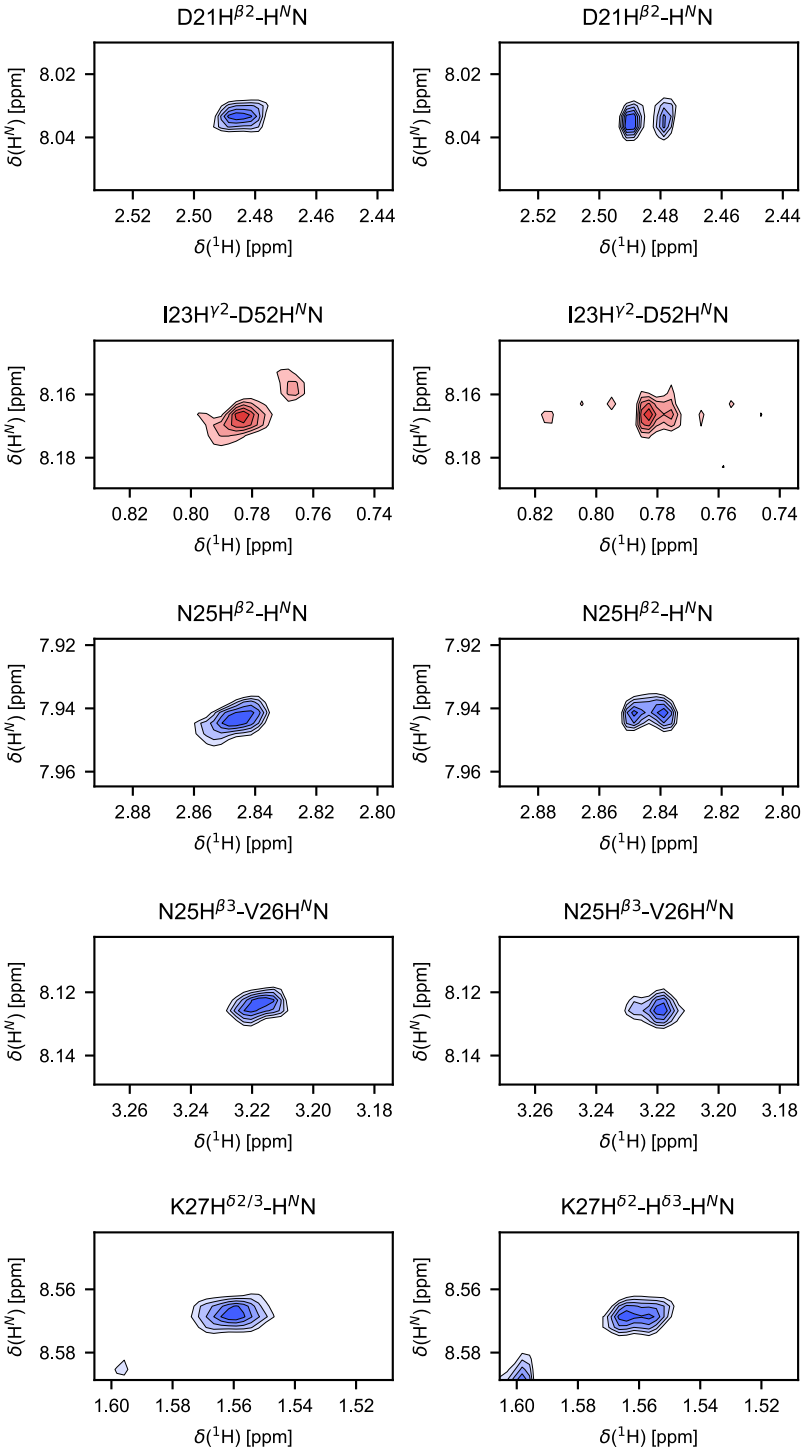
Figure 3.S5. Linewidth error estimation by confidence interval calculation.

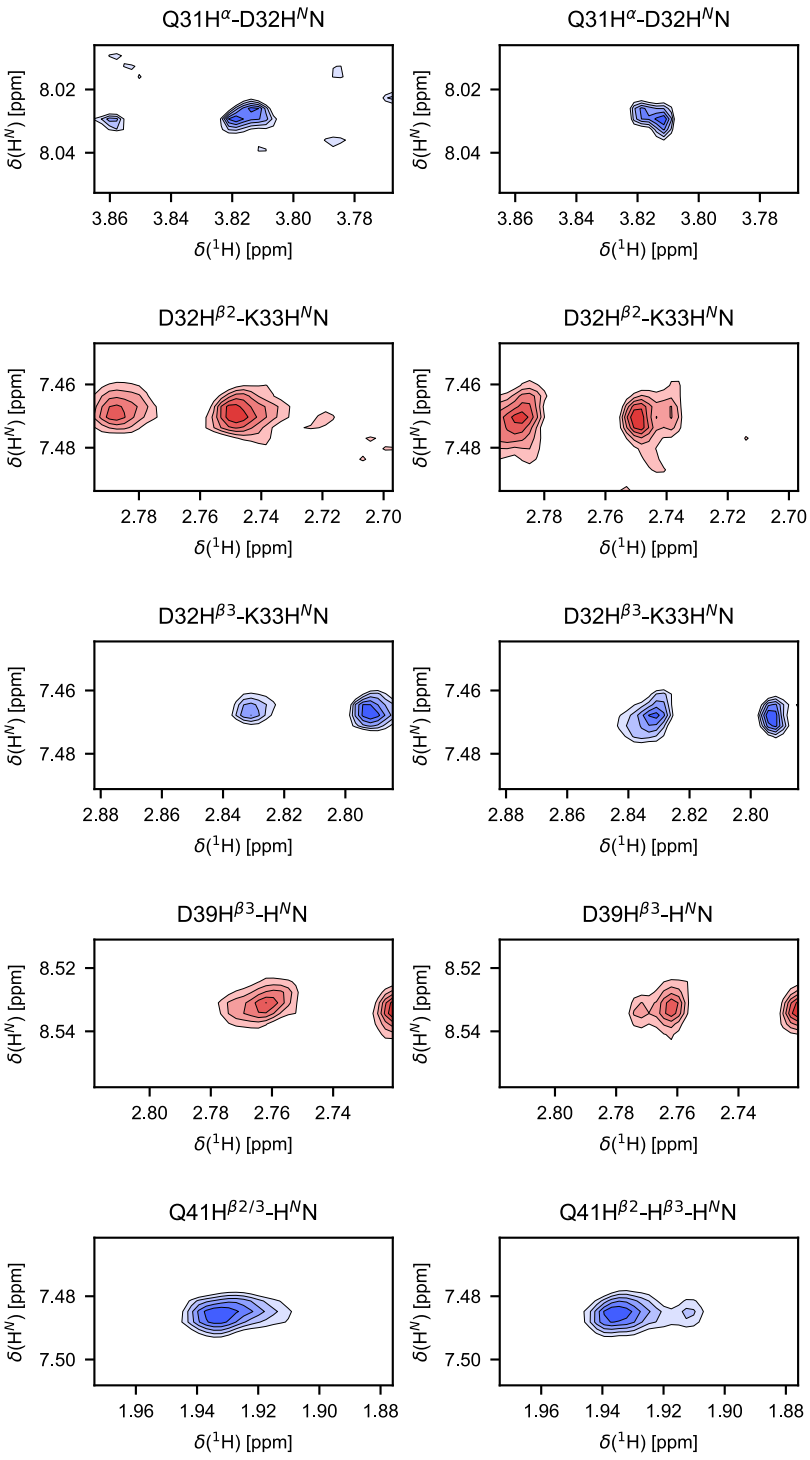
Two matched enhanced srNOESY-HSQC experiments were collected on ubiquitin with a conventional 8-step phase cycle or a truncated 4-step phase cycle. A line of best-fit with a slope of 1.001 ± 0.008 is shown in blue, and the experimental error was estimated by calculation of the 68.8 and 95.5% confidence intervals (gray shaded regions). The 1-σ and 2-σ confidence intervals were found to be 8.2 and 31.9%.

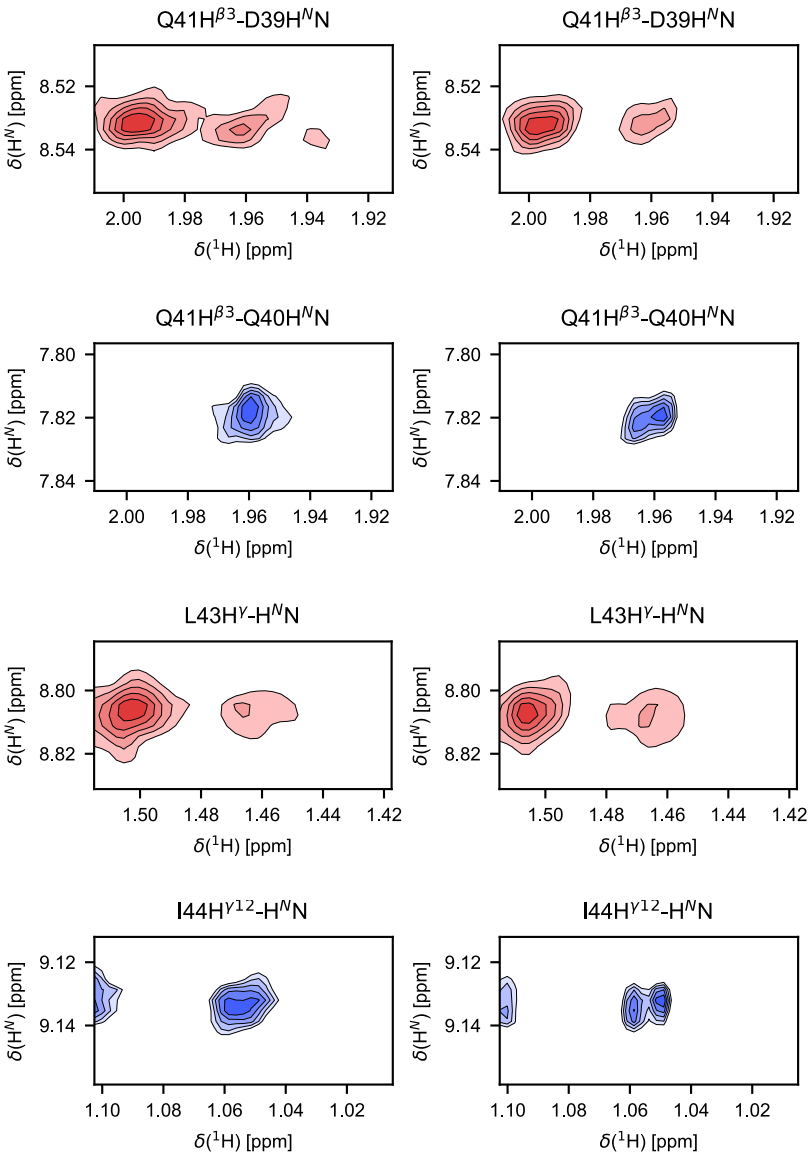

V5H$^{\gamma 1/2}$-H^{N}N
V5H$^{\gamma 1}$-H$^{\gamma 2}$-H^{N}N
K6H$^{\beta 3}$-T7H^{N}N
K6H$^{\beta 3}$-T7H^{N}N
L8H$^{\beta 2}$-H^{N}N
L8H$^{\beta 2}$-H^{N}N
L8H$^{\beta 3}$-H^{N}N
L8H$^{\beta 3}$-H^{N}N
L8H$^{\gamma}$-H^{N}N
L8H$^{\gamma}$-H^{N}N
δ(H^N) [ppm]
δ(^{1}H) [ppm]

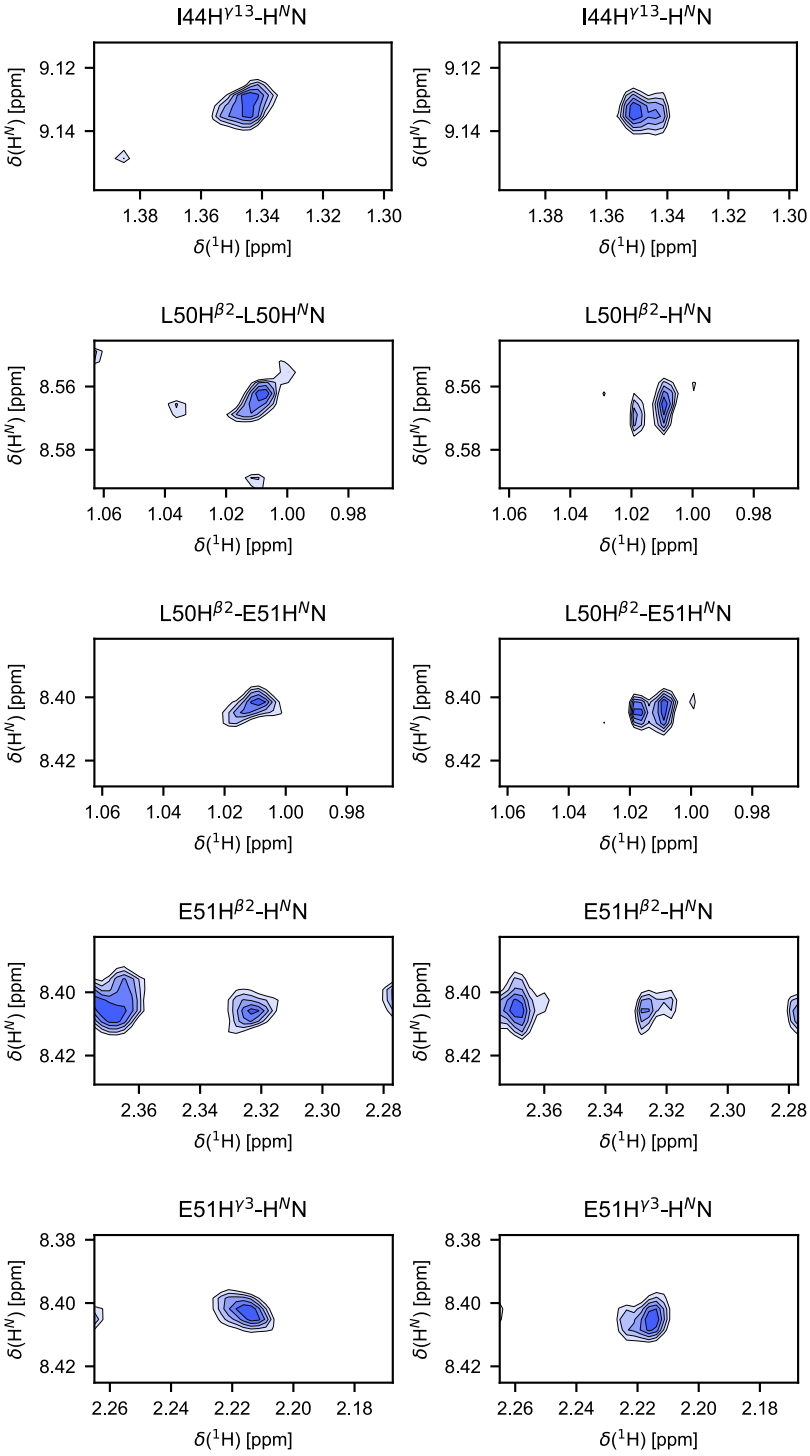

G10H$^{\alpha 2/3}$-H^{N}N
G10H$^{\alpha 2}$-H$^{\alpha 3}$-H^{N}N
K11H$^{\beta 2/3}$-H^{N}N
K11H$^{\beta 2}$-H$^{\beta 3}$-H^{N}N
L15H$^{\alpha}$/V17H$^{\alpha}$-V17H^{N}N
L15H$^{\alpha}$-V17H$^{\alpha}$-V17H^{N}N
L15H$^{\beta 3}$-H^{N}N
L15H$^{\beta 3}$-H^{N}N
E16H$^{\beta 3}$-H^{N}N
E16H$^{\beta 3}$-H^{N}N
δ(H^{N}) [ppm]
δ(^{1}H) [ppm]

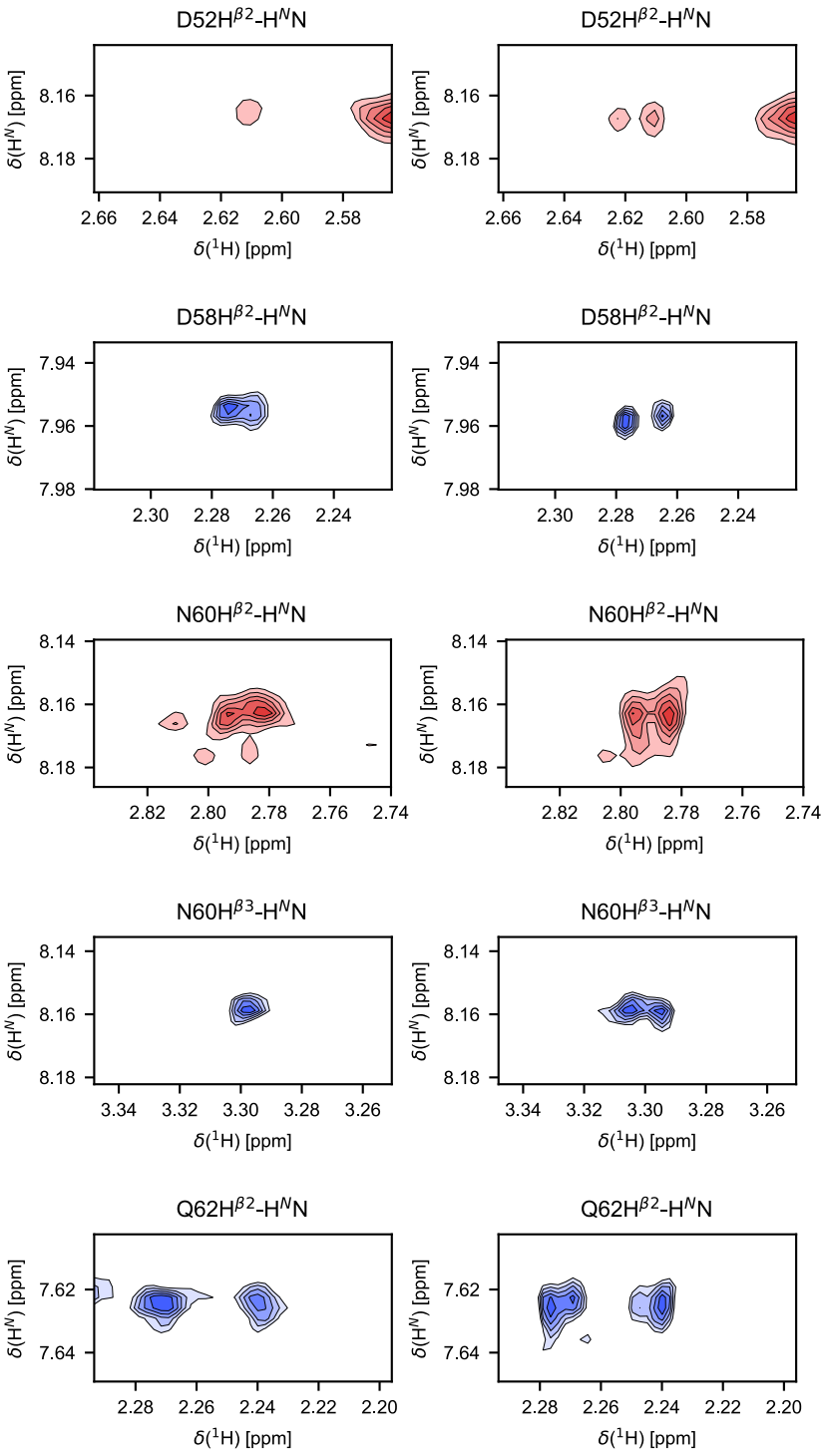

E16H$^{\gamma 2}$-H^{N}N
E16H$^{\gamma 2}$-H^{N}N
V17HB-H^{N}N
V17HB-H^{N}N
E18H$^{\beta 2}$-H^{N}N
E18H$^{\beta 2}$-H^{N}N
E18H$^{\beta 3}$-S20H^{N}N
E18H$^{\beta 3}$-S20H^{N}N
E18H$^{\beta 3}$-D21H^{N}N
E18H$^{\beta 3}$-D21H^{N}N
δ(H^N) [ppm]
δ(^{1}H) [ppm]

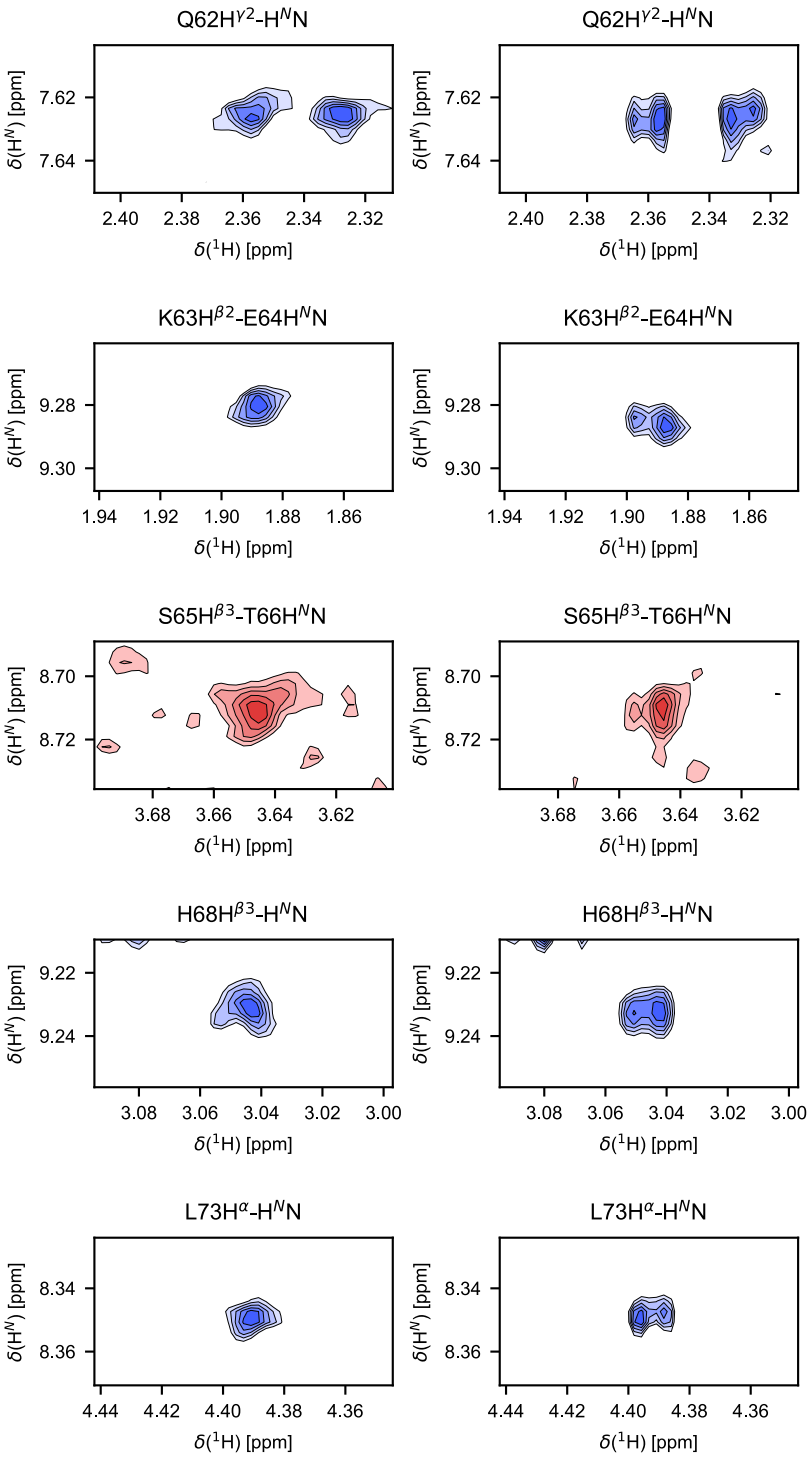

D21H$^{\beta2}$-H^{N}N
D21H$^{\beta2}$-H^{N}N
I23H$^{\gamma2}$-D52H^{N}N
I23H$^{\gamma2}$-D52H^{N}N
N25H$^{\beta2}$-H^{N}N
N25H$^{\beta2}$-H^{N}N
N25H$^{\beta3}$-V26H^{N}N
N25H$^{\beta3}$-V26H^{N}N
K27H$^{\delta2/3}$-H^{N}N
K27H$^{\delta2}$-H$^{\delta3}$-H^{N}N
δ(H^{N}) [ppm]
δ(^{1}H) [ppm]


Q31H$^{\alpha}$-D32H^{N}N
Q31H$^{\alpha}$-D32H^{N}N
D32H$^{\beta2}$-K33H^{N}N
D32H$^{\beta2}$-K33H^{N}N
D32H$^{\beta3}$-K33H^{N}N
D32H$^{\beta3}$-K33H^{N}N
D39H$^{\beta3}$-H^{N}N
D39H$^{\beta3}$-H^{N}N
Q41H$^{\beta2/3}$-H^{N}N
Q41H$^{\beta2}$-H$^{\beta3}$-H^{N}N
δ(H^{N}) [ppm]
δ(^{1}H) [ppm]


Q41Hβ3-D39HNN
Q41Hβ3-D39HNN
Q41Hβ3-Q40HNN
Q41Hβ3-Q40HNN
L43Hγ-HNN
L43Hγ-HNN
I44Hγ12-HNN
I44Hγ12-HNN
δ(HN) [ppm]
δ(1H) [ppm]


I44H$^{\gamma 13}$-H^{N}N
I44H$^{\gamma 13}$-H^{N}N
L50H$^{\beta 2}$-L50H^{N}N
L50H$^{\beta 2}$-H^{N}N
L50H$^{\beta 2}$-E51H^{N}N
L50H$^{\beta 2}$-E51H^{N}N
E51H$^{\beta 2}$-H^{N}N
E51H$^{\beta 2}$-H^{N}N
E51H$^{\gamma 3}$-H^{N}N
E51H$^{\gamma 3}$-H^{N}N
δ(H^N) [ppm]
δ(^{1}H) [ppm]


D52Hβ2-HNN
D52Hβ2-HNN
D58Hβ2-HNN
D58Hβ2-HNN
N60Hβ2-HNN
N60Hβ2-HNN
N60Hβ3-HNN
N60Hβ3-HNN
Q62Hβ2-HNN
Q62Hβ2-HNN
δ(HN) [ppm]
δ(1H) [ppm]


Q62H$^{\gamma 2}$-H^{N}N
Q62H$^{\gamma 2}$-H^{N}N
K63H$^{\beta 2}$-E64H^{N}N
K63H$^{\beta 2}$-E64H^{N}N
S65H$^{\beta 3}$-T66H^{N}N
S65H$^{\beta 3}$-T66H^{N}N
H68H$^{\beta 3}$-H^{N}N
H68H$^{\beta 3}$-H^{N}N
L73H$^{\alpha}$-H^{N}N
L73H$^{\alpha}$-H^{N}N
δ(H^N) [ppm]
δ(^{1}H) [ppm]

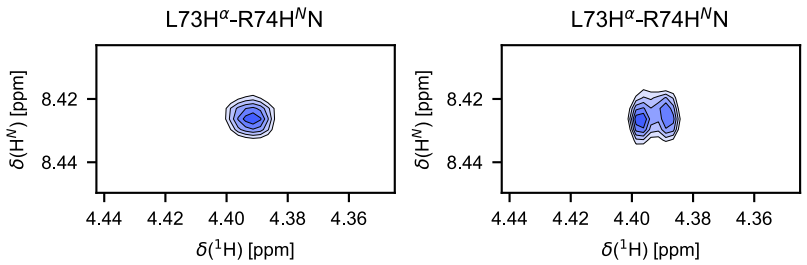


Figure 3.S6. Contour plots of conventional NOESY-HSQC (left column) and enhanced srNOESY-HSQC (right column) peak pairs for ubiquitin, where two peaks are resolved from one original peak.

Contours were summed over the FWHH in the ^{15}N dimension. Contours are drawn at the half-height (blue), quarter-height (red), or eighth-height (purple) of the peaks shown.

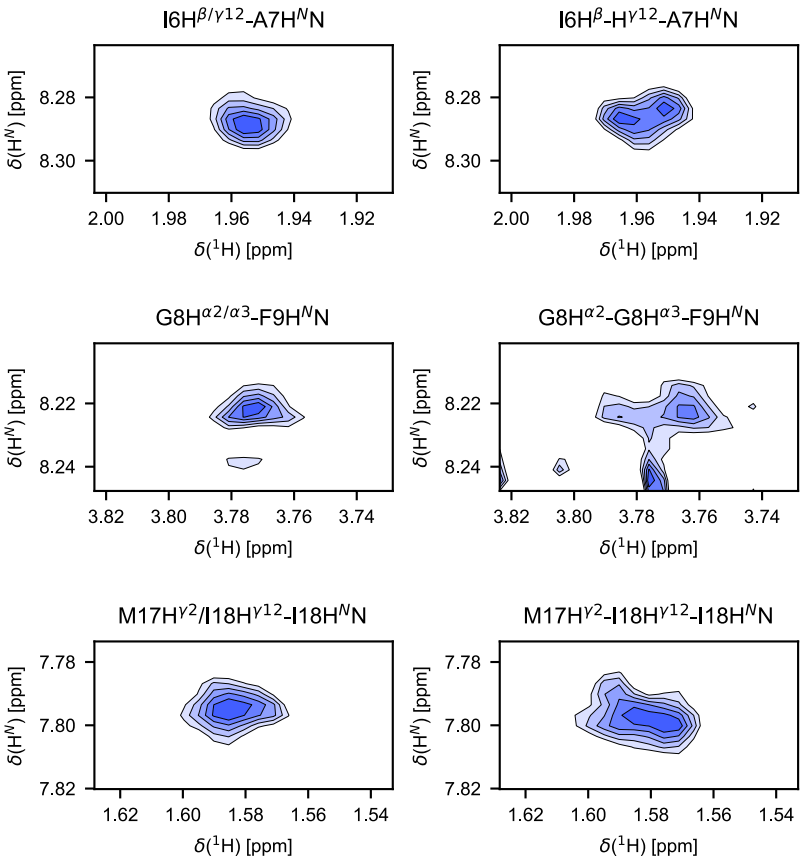
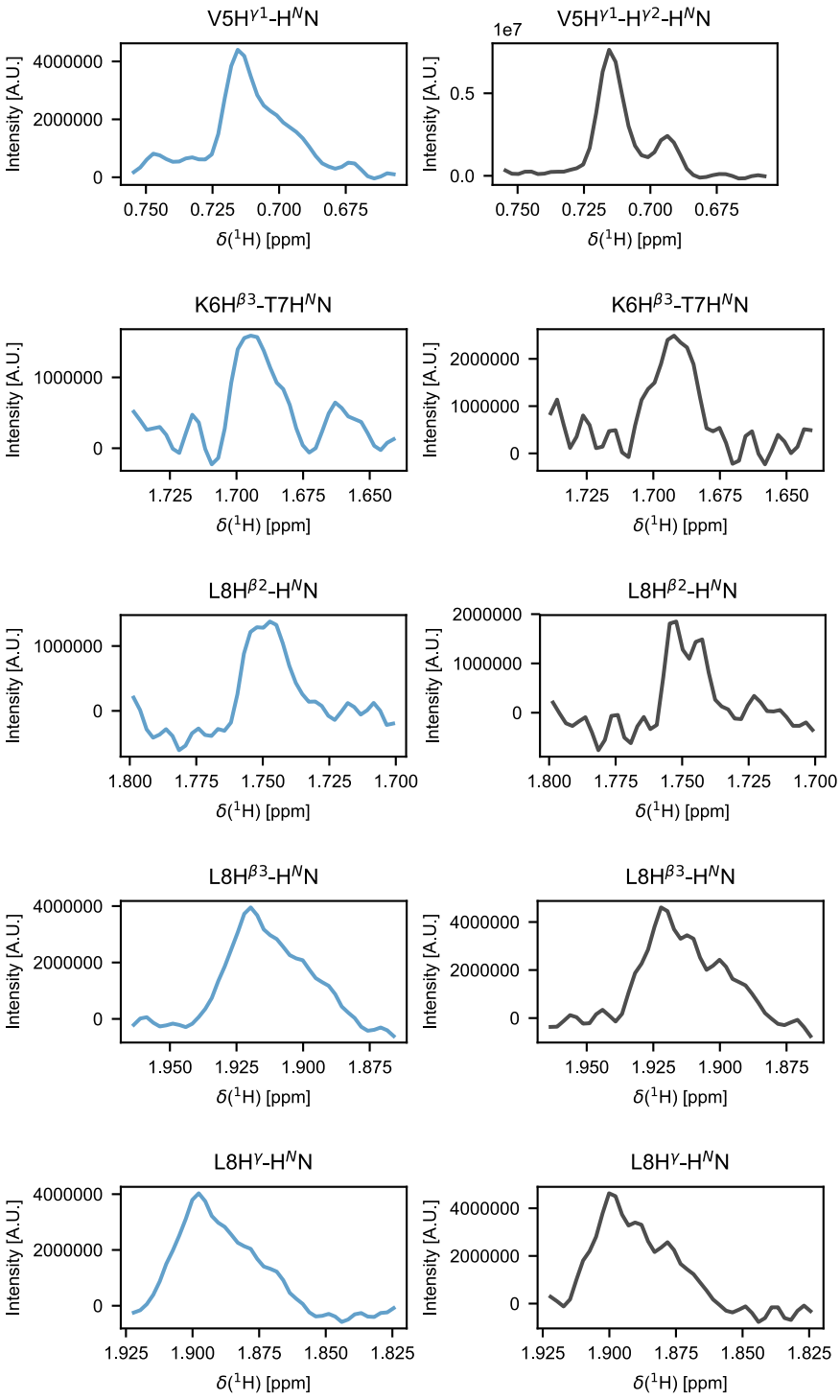
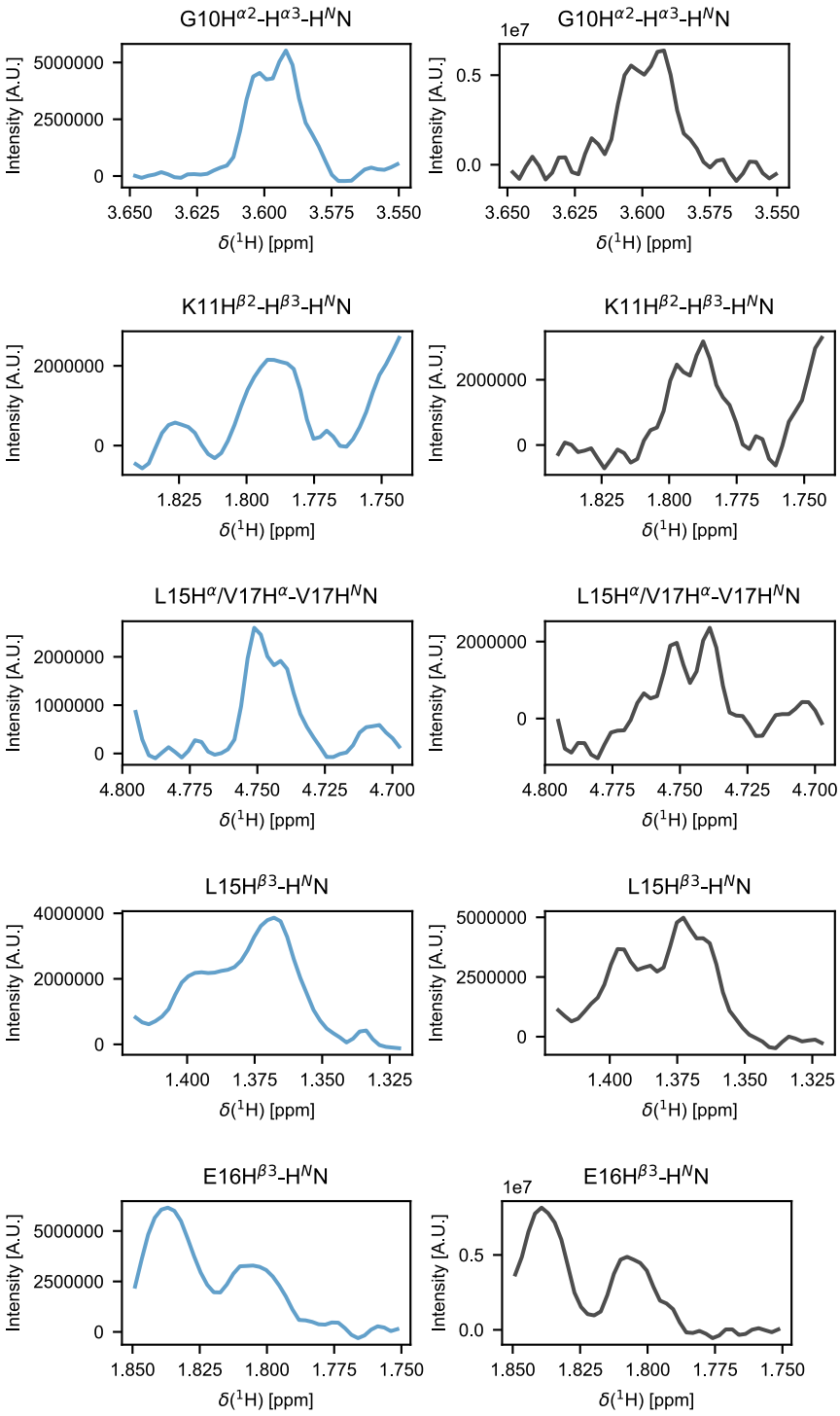
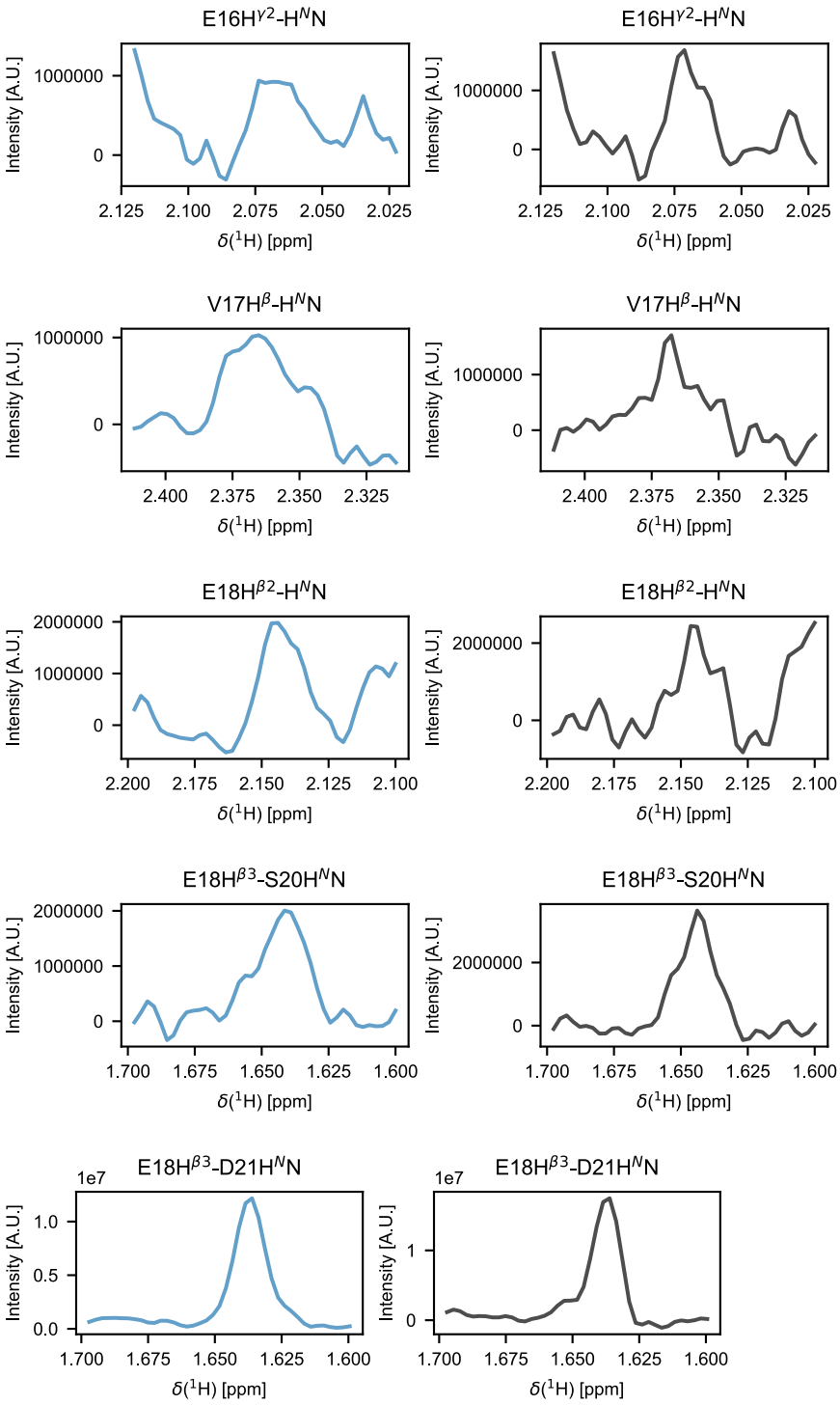


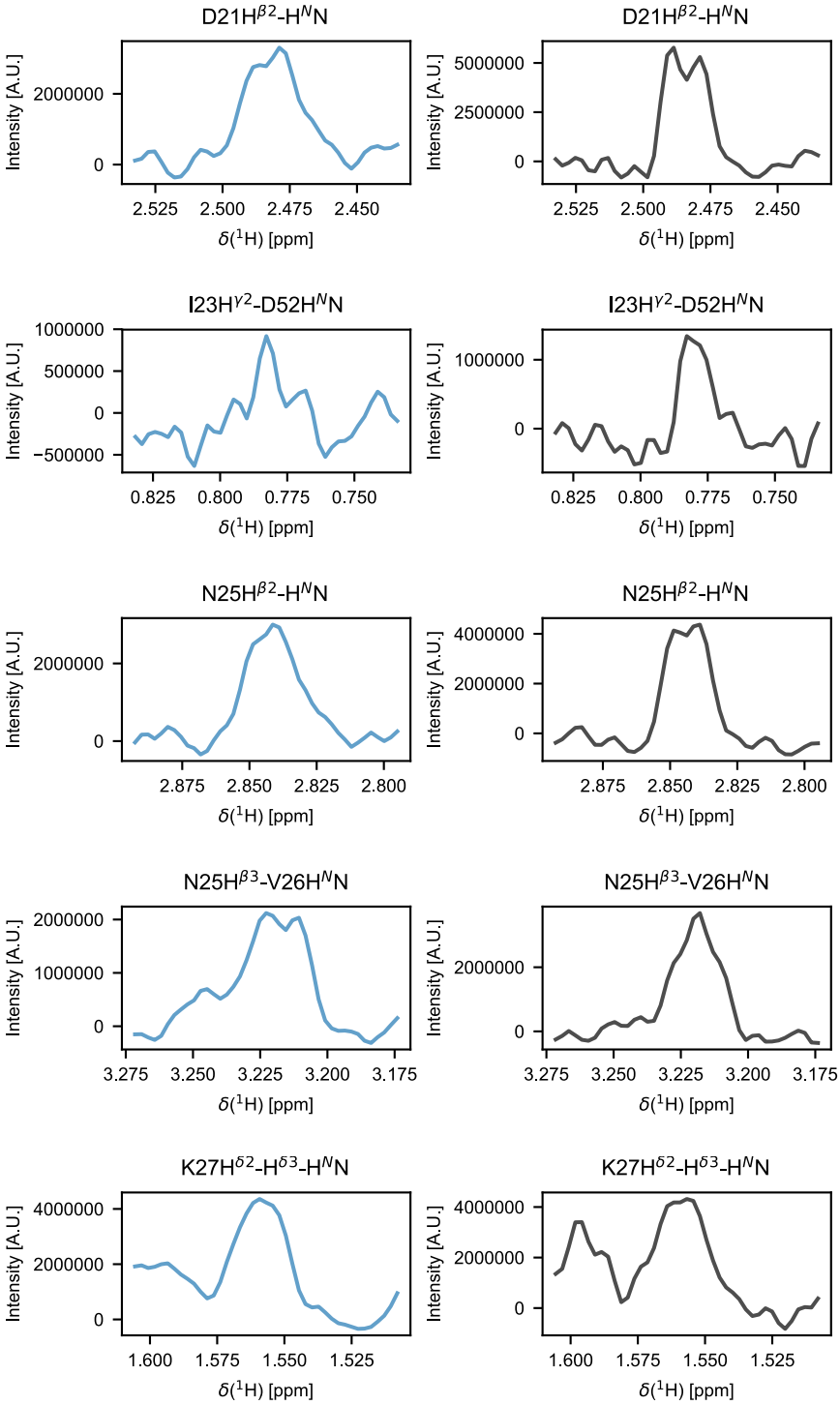
Figure 3.S7. Contour plots of conventional NOESY-HSQC (left column) and enhanced srNOESY-HSQC (right column) peak pairs for HAfp-bicelles, where two peaks are resolved from one original peak.

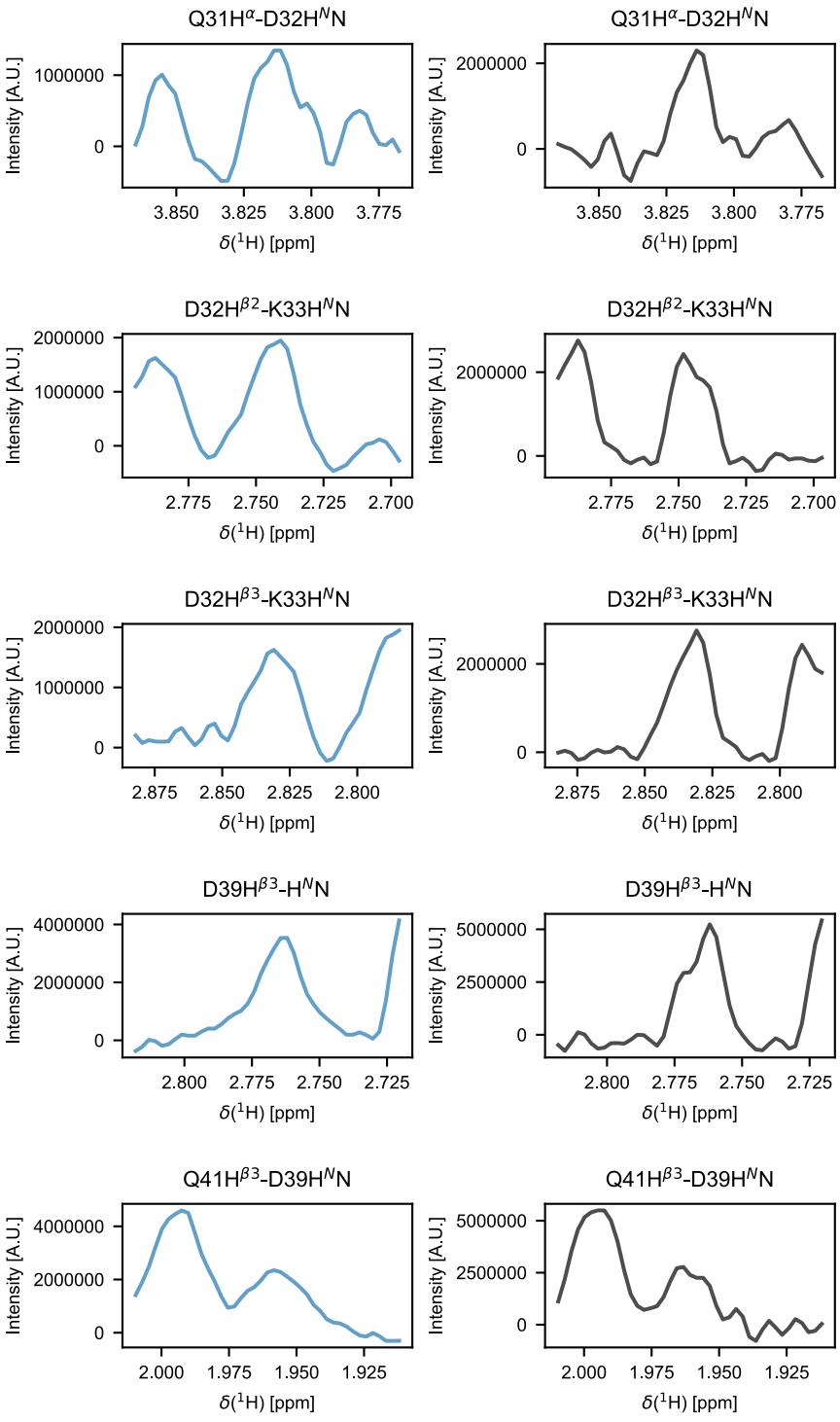
Contours were summed over the FWHH in the ^{15}N dimension. Contours are drawn at the half-height (blue) of the peaks shown.

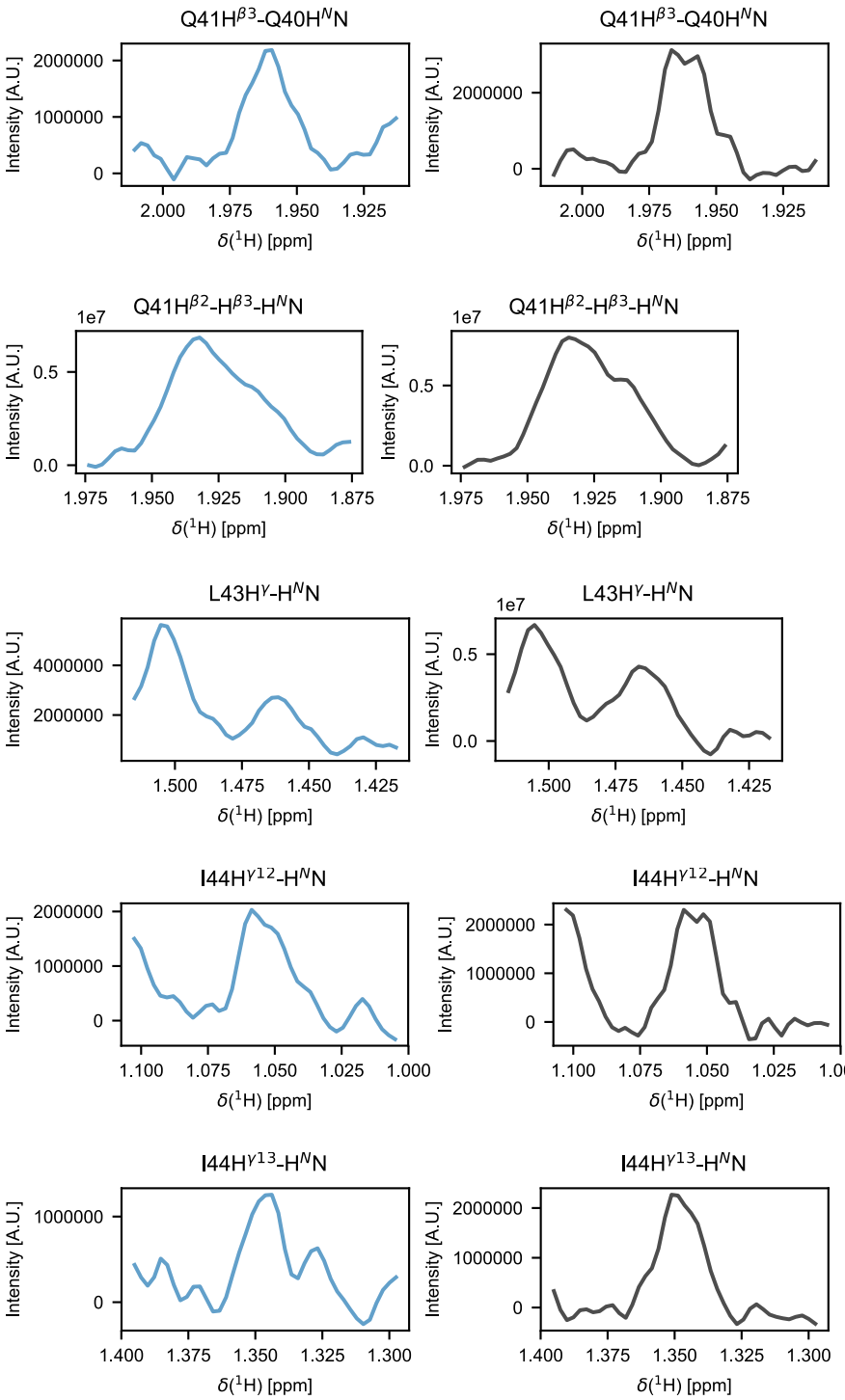

V5H$^{\gamma1}$-H^{N}N
V5H$^{\gamma1}$-H$^{\gamma2}$-H^{N}N
1e7
K6H$^{\beta3}$-T7H^{N}N
K6H$^{\beta3}$-T7H^{N}N
L8H$^{\beta2}$-H^{N}N
L8H$^{\beta2}$-H^{N}N
L8H$^{\beta3}$-H^{N}N
L8H$^{\beta3}$-H^{N}N
L8H$^{\gamma}$-H^{N}N
L8H$^{\gamma}$-H^{N}N
Intensity [A.U.]
δ(^{1}H) [ppm]

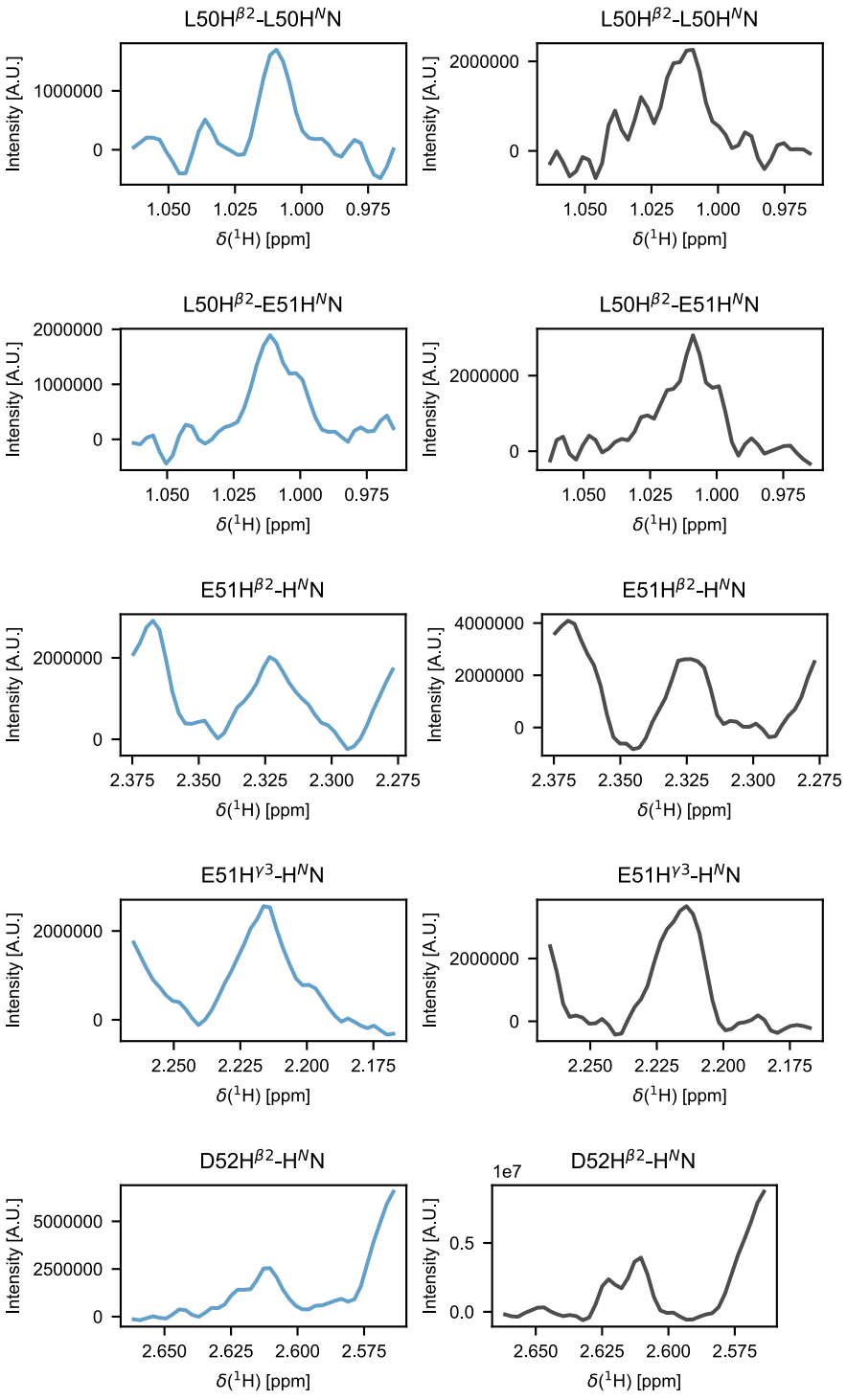

G10H$^{\alpha2}$-H$^{\alpha3}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
3.650
3.625
3.600
3.575
3.550
$\delta(^1H)$ [ppm]
G10H$^{\alpha2}$-H$^{\alpha3}$-H^{N}N
1e7
0.5
0.0
Intensity [A.U.]
3.650
3.625
3.600
3.575
3.550
$\delta(^1H)$ [ppm]
K11H$^{\beta2}$-H$^{\beta3}$-H^{N}N
Intensity [A.U.]
2000000
0
1.825
1.800
1.775
1.750
$\delta(^1H)$ [ppm]
K11H$^{\beta2}$-H$^{\beta3}$-H^{N}N
Intensity [A.U.]
2000000
0
1.825
1.800
1.775
1.750
$\delta(^1H)$ [ppm]
L15H$^{\alpha}$/V17H$^{\alpha}$-V17H^{N}N
Intensity [A.U.]
2000000
1000000
0
4.800
4.775
4.750
4.725
4.700
$\delta(^1H)$ [ppm]
L15H$^{\alpha}$/V17H$^{\alpha}$-V17H^{N}N
Intensity [A.U.]
2000000
0
4.800
4.775
4.750
4.725
4.700
$\delta(^1H)$ [ppm]
L15H$^{\beta3}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
1.400
1.375
1.350
1.325
$\delta(^1H)$ [ppm]
L15H$^{\beta3}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
1.400
1.375
1.350
1.325
$\delta(^1H)$ [ppm]
E16H$^{\beta3}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
1.850
1.825
1.800
1.775
1.750
$\delta(^1H)$ [ppm]
E16H$^{\beta3}$-H^{N}N
1e7
Intensity [A.U.]
0.5
0.0
1.850
1.825
1.800
1.775
1.750
$\delta(^1H)$ [ppm]

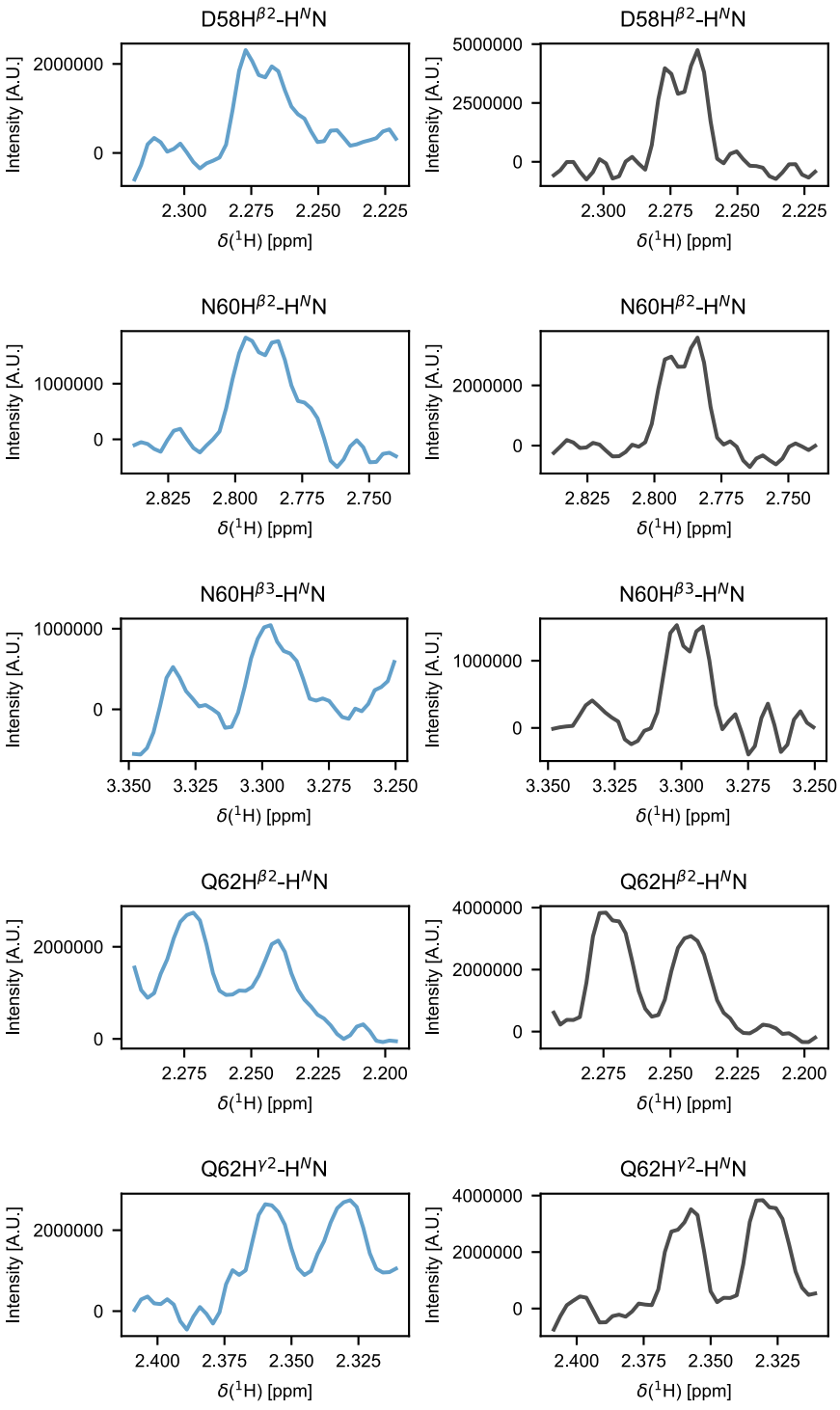

E16H$^{\gamma 2}$-H^{N}N
Intensity [A.U.]
1000000
0
2.125 2.100 2.075 2.050 2.025
$\delta(^1H)$ [ppm]
E16H$^{\gamma 2}$-H^{N}N
Intensity [A.U.]
1000000
0
2.125 2.100 2.075 2.050 2.025
$\delta(^1H)$ [ppm]
V17H$^{\beta}$-H^{N}N
Intensity [A.U.]
1000000
0
2.400 2.375 2.350 2.325
$\delta(^1H)$ [ppm]
V17H$^{\beta}$-H^{N}N
Intensity [A.U.]
1000000
0
2.400 2.375 2.350 2.325
$\delta(^1H)$ [ppm]
E18H$^{\beta 2}$-H^{N}N
Intensity [A.U.]
2000000
1000000
0
2.200 2.175 2.150 2.125 2.100
$\delta(^1H)$ [ppm]
E18H$^{\beta 2}$-H^{N}N
Intensity [A.U.]
2000000
0
2.200 2.175 2.150 2.125 2.100
$\delta(^1H)$ [ppm]
E18H$^{\beta 3}$-S20H^{N}N
Intensity [A.U.]
2000000
1000000
0
1.700 1.675 1.650 1.625 1.600
$\delta(^1H)$ [ppm]
E18H$^{\beta 3}$-S20H^{N}N
Intensity [A.U.]
2000000
0
1.700 1.675 1.650 1.625 1.600
$\delta(^1H)$ [ppm]
E18H$^{\beta 3}$-D21H^{N}N
1e7
Intensity [A.U.]
1.0
0.5
0.0
1.700 1.675 1.650 1.625 1.600
$\delta(^1H)$ [ppm]
E18H$^{\beta 3}$-D21H^{N}N
1e7
Intensity [A.U.]
1
0
1.700 1.675 1.650 1.625 1.600
$\delta(^1H)$ [ppm]


D21H$^{\beta2}$-H^{N}N
Intensity [A.U.]
2000000
0
2.525
2.500
2.475
2.450
$\delta(^1H)$ [ppm]
D21H$^{\beta2}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
2.525
2.500
2.475
2.450
$\delta(^1H)$ [ppm]
I23H$^{\gamma2}$-D52H^{N}N
Intensity [A.U.]
1000000
500000
0
−500000
0.825
0.800
0.775
0.750
$\delta(^1H)$ [ppm]
I23H$^{\gamma2}$-D52H^{N}N
Intensity [A.U.]
1000000
0
0.825
0.800
0.775
0.750
$\delta(^1H)$ [ppm]
N25H$^{\beta2}$-H^{N}N
Intensity [A.U.]
2000000
0
2.875
2.850
2.825
2.800
$\delta(^1H)$ [ppm]
N25H$^{\beta2}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
2.875
2.850
2.825
2.800
$\delta(^1H)$ [ppm]
N25H$^{\beta3}$-V26H^{N}N
Intensity [A.U.]
2000000
1000000
0
3.275
3.250
3.225
3.200
3.175
$\delta(^1H)$ [ppm]
N25H$^{\beta3}$-V26H^{N}N
Intensity [A.U.]
2000000
0
3.275
3.250
3.225
3.200
3.175
$\delta(^1H)$ [ppm]
K27H$^{\delta2}$-H$^{\delta3}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
1.600
1.575
1.550
1.525
$\delta(^1H)$ [ppm]
K27H$^{\delta2}$-H$^{\delta3}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
1.600
1.575
1.550
1.525
$\delta(^1H)$ [ppm]


Q31H$^{\alpha}$-D32H^{N}N
Intensity [A.U.]
1000000
0
3.850 3.825 3.800 3.775
$\delta(^1H)$ [ppm]
Q31H$^{\alpha}$-D32H^{N}N
Intensity [A.U.]
2000000
0
3.850 3.825 3.800 3.775
$\delta(^1H)$ [ppm]
D32H$^{\beta 2}$-K33H^{N}N
Intensity [A.U.]
2000000
1000000
0
2.775 2.750 2.725 2.700
$\delta(^1H)$ [ppm]
D32H$^{\beta 2}$-K33H^{N}N
Intensity [A.U.]
2000000
0
2.775 2.750 2.725 2.700
$\delta(^1H)$ [ppm]
D32H$^{\beta 3}$-K33H^{N}N
Intensity [A.U.]
2000000
1000000
0
2.875 2.850 2.825 2.800
$\delta(^1H)$ [ppm]
D32H$^{\beta 3}$-K33H^{N}N
Intensity [A.U.]
2000000
0
2.875 2.850 2.825 2.800
$\delta(^1H)$ [ppm]
D39H$^{\beta 3}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
2.800 2.775 2.750 2.725
$\delta(^1H)$ [ppm]
D39H$^{\beta 3}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
2.800 2.775 2.750 2.725
$\delta(^1H)$ [ppm]
Q41H$^{\beta 3}$-D39H^{N}N
Intensity [A.U.]
4000000
2000000
0
2.000 1.975 1.950 1.925
$\delta(^1H)$ [ppm]
Q41H$^{\beta 3}$-D39H^{N}N
Intensity [A.U.]
5000000
2500000
0
2.000 1.975 1.950 1.925
$\delta(^1H)$ [ppm]


Q41H$^{\beta3}$-Q40H^{N}N
Intensity [A.U.]
2000000
1000000
0
2.000
1.975
1.950
1.925
$\delta(^1H)$ [ppm]
Q41H$^{\beta3}$-Q40H^{N}N
Intensity [A.U.]
2000000
0
2.000
1.975
1.950
1.925
$\delta(^1H)$ [ppm]
Q41H$^{\beta2}$-H$^{\beta3}$-H^{N}N
1e7
Intensity [A.U.]
0.5
0.0
1.975
1.950
1.925
1.900
1.875
$\delta(^1H)$ [ppm]
Q41H$^{\beta2}$-H$^{\beta3}$-H^{N}N
1e7
Intensity [A.U.]
0.5
0.0
1.975
1.950
1.925
1.900
1.875
$\delta(^1H)$ [ppm]
L43H$^{\gamma}$-H^{N}N
Intensity [A.U.]
4000000
2000000
1.500
1.475
1.450
1.425
$\delta(^1H)$ [ppm]
L43H$^{\gamma}$-H^{N}N
1e7
Intensity [A.U.]
0.5
0.0
1.500
1.475
1.450
1.425
$\delta(^1H)$ [ppm]
I44H$^{\gamma12}$-H^{N}N
Intensity [A.U.]
2000000
1000000
0
1.100
1.075
1.050
1.025
1.000
$\delta(^1H)$ [ppm]
I44H$^{\gamma12}$-H^{N}N
Intensity [A.U.]
2000000
1000000
0
1.100
1.075
1.050
1.025
1.000
$\delta(^1H)$ [ppm]
I44H$^{\gamma13}$-H^{N}N
Intensity [A.U.]
1000000
0
1.400
1.375
1.350
1.325
1.300
$\delta(^1H)$ [ppm]
I44H$^{\gamma13}$-H^{N}N
Intensity [A.U.]
2000000
1000000
0
1.400
1.375
1.350
1.325
1.300
$\delta(^1H)$ [ppm]


L50H$^{\beta2}$-L50H^{N}N
Intensity [A.U.]
1000000
0
1.050 1.025 1.000 0.975
$\delta(^1H)$ [ppm]
L50H$^{\beta2}$-L50H^{N}N
Intensity [A.U.]
2000000
0
1.050 1.025 1.000 0.975
$\delta(^1H)$ [ppm]
L50H$^{\beta2}$-E51H^{N}N
Intensity [A.U.]
2000000
1000000
0
1.050 1.025 1.000 0.975
$\delta(^1H)$ [ppm]
L50H$^{\beta2}$-E51H^{N}N
Intensity [A.U.]
2000000
0
1.050 1.025 1.000 0.975
$\delta(^1H)$ [ppm]
E51H$^{\beta2}$-H^{N}N
Intensity [A.U.]
2000000
0
2.375 2.350 2.325 2.300 2.275
$\delta(^1H)$ [ppm]
E51H$^{\beta2}$-H^{N}N
Intensity [A.U.]
4000000
2000000
0
2.375 2.350 2.325 2.300 2.275
$\delta(^1H)$ [ppm]
E51H$^{\gamma3}$-H^{N}N
Intensity [A.U.]
2000000
0
2.250 2.225 2.200 2.175
$\delta(^1H)$ [ppm]
E51H$^{\gamma3}$-H^{N}N
Intensity [A.U.]
2000000
0
2.250 2.225 2.200 2.175
$\delta(^1H)$ [ppm]
D52H$^{\beta2}$-H^{N}N
Intensity [A.U.]
5000000
2500000
0
2.650 2.625 2.600 2.575
$\delta(^1H)$ [ppm]
D52H$^{\beta2}$-H^{N}N
1e7
Intensity [A.U.]
0.5
0.0
2.650 2.625 2.600 2.575
$\delta(^1H)$ [ppm]


D58Hβ2-HNN
Intensity [A.U.]
2000000
0
2.300 2.275 2.250 2.225
δ(1H) [ppm]
D58Hβ2-HNN
Intensity [A.U.]
5000000
2500000
0
2.300 2.275 2.250 2.225
δ(1H) [ppm]
N60Hβ2-HNN
Intensity [A.U.]
1000000
0
2.825 2.800 2.775 2.750
δ(1H) [ppm]
N60Hβ2-HNN
Intensity [A.U.]
2000000
0
2.825 2.800 2.775 2.750
δ(1H) [ppm]
N60Hβ3-HNN
Intensity [A.U.]
1000000
0
3.350 3.325 3.300 3.275 3.250
δ(1H) [ppm]
N60Hβ3-HNN
Intensity [A.U.]
1000000
0
3.350 3.325 3.300 3.275 3.250
δ(1H) [ppm]
Q62Hβ2-HNN
Intensity [A.U.]
2000000
0
2.275 2.250 2.225 2.200
δ(1H) [ppm]
Q62Hβ2-HNN
Intensity [A.U.]
4000000
2000000
0
2.275 2.250 2.225 2.200
δ(1H) [ppm]
Q62Hγ2-HNN
Intensity [A.U.]
2000000
0
2.400 2.375 2.350 2.325
δ(1H) [ppm]
Q62Hγ2-HNN
Intensity [A.U.]
4000000
2000000
0
2.400 2.375 2.350 2.325
δ(1H) [ppm]

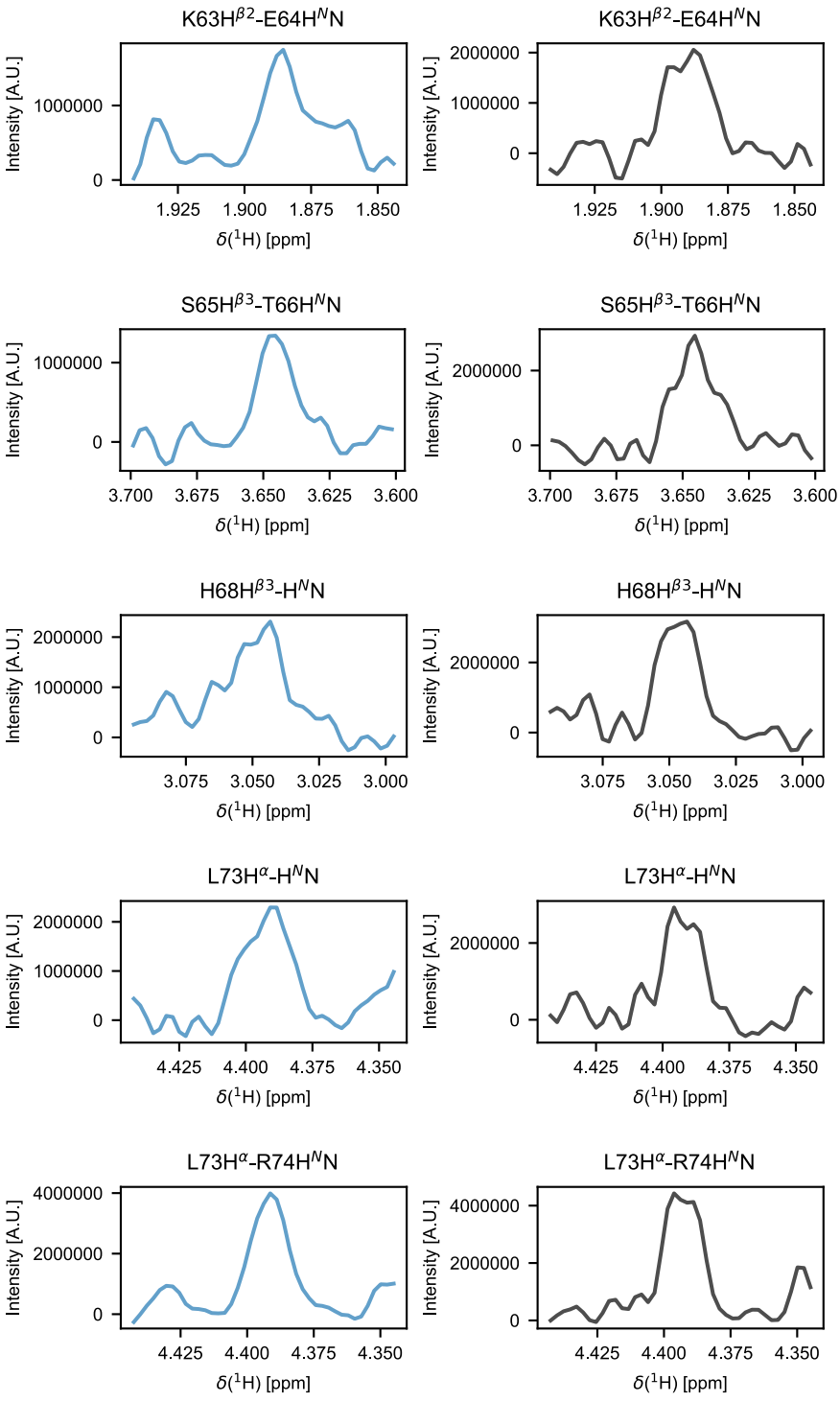


Figure 3.S8. Ubiquitin ^{1}H indirect dimension cross sections of the contour plots in Figure 3.S6.

The conventional NOESY spectra are shown in blue (left), and the srNOESY spectra are shown in black (right).

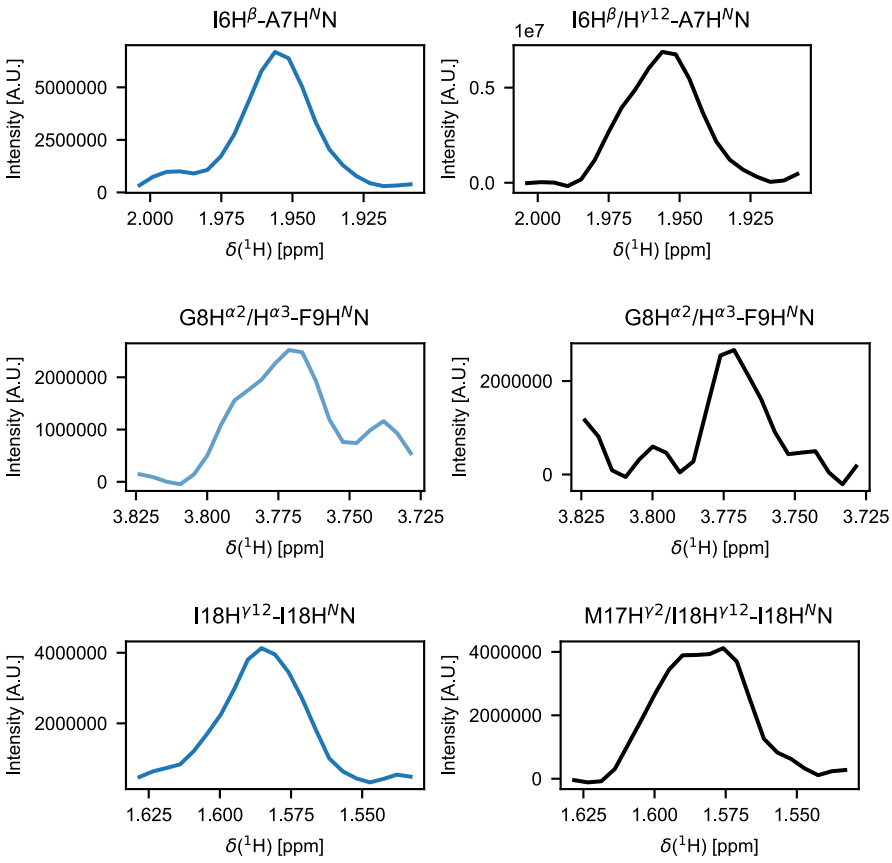


Figure 3.S9. HAfp-bicelles ^{1}H indirect dimension cross sections of the contour plots in Figure 3.S7.

The conventional NOESY spectra are shown in blue (left), and the srNOESY spectra are shown in black (right).

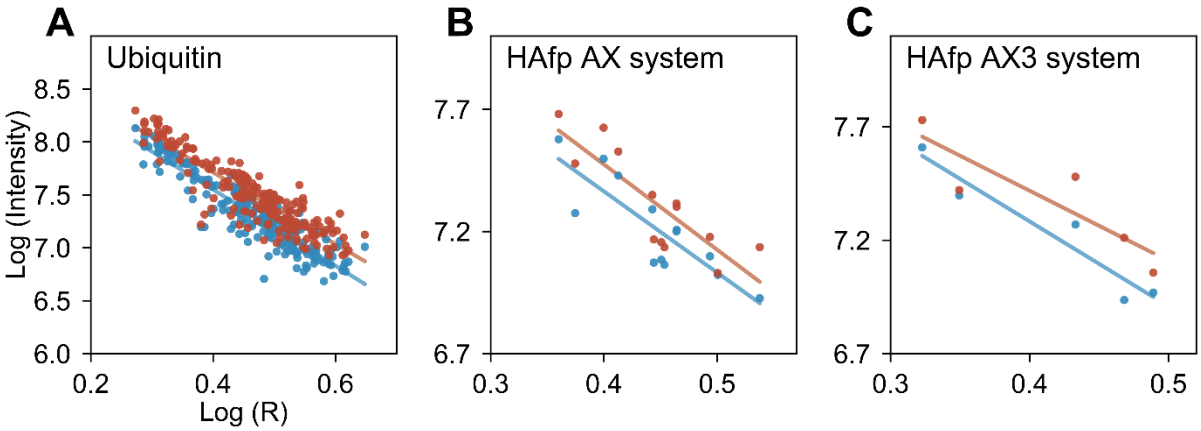


Figure 3.S10. NOE distance restraint calibration plots for (A) all ubiquitin restraints, (B) HAfp-bicelles non-methyl (AX/AX2) restraints, and (C) HAfp-bicelles methyl (AX3) restraints.

Published distance restraints from 13DZ and 2KXA structures in the PDB were used as the accepted distance restraint value (x-axis). The peak intensities from the conventional experiment are shown in blue, while peak intensities from the enhanced srNOESY-HSQC are shown in red. A total of 205 known distances were used in the calibration of ubiquitin, generating best-fit lines with slopes of -3.595 ± 0.104 (blue) and -3.381 ± 0.112 (red). A total of 13 and 5 known distances were used in the calibration of non-methyl and methyl groups, respectively, for HAfp-bicelles. The non-methyl best-fit lines had slopes of -3.324 ± 0.619 (blue) and -3.491± 0.667 (red), while the methyl best-fit lines had slopes of -3.741 ± 0.658 (blue) and -3.087 ± 0.990 (red). Vertical error bars are smaller than the figure symbols in all panels.

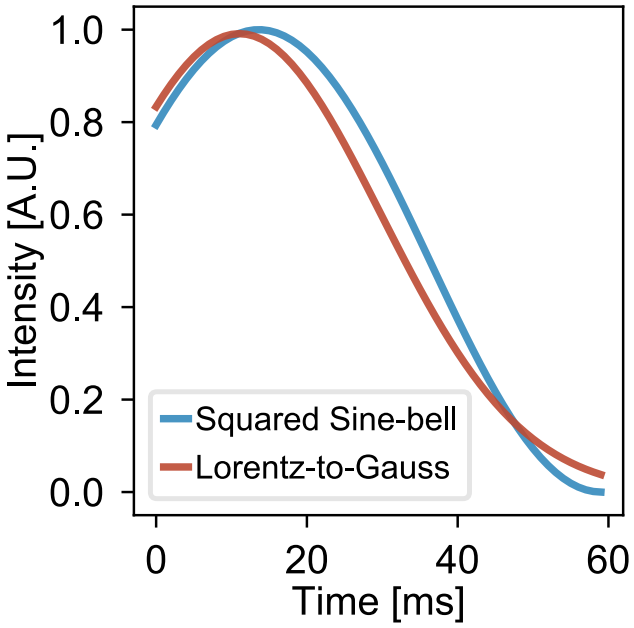


Figure 3.S11. Comparison plots of a squared sine-bell apodization window function and a Lorentzian-to-Gaussian conversion.

The blue line depicts a squared sine-bell window as a function of time starting at 0.30π and ending at 1.00π. The orange line represents a Lorentz-to-Gauss apodization function with an inverse exponential Lorentzian window of 10 Hz and a Gaussian broadening window of 20 Hz, with the intensity normalized to the squared sine-bell function

References

[1] G. Cornilescu, J. L. Marquardt, M. Ottiger, and A. Bax, "Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase," *J. Am. Chem. Soc.*, vol. 120, no. 27, pp. 6836–6837, Jul. 1998.

[2] J. L. Lorieau, J. M. Louis, and A. Bax, "The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface," *Proc. Natl. Acad. Sci.*, vol. 107, no. 25, pp. 11341–11346, 2010.

Pulse program executed on a Bruker AVIII-750 wide-bore spectrometer with TopSpin 3.1. The changes to the Bruker standard noesyhsqcf3gpsi3d pulse program have been highlighted below.

Parameters used for the ubiquitin srNOESY-HSQC dataset were: $p_{90}(^{1}H)$ = 9.2 μs, $p_{180}(^{1}H)$ = 18.4 μs, $p_{90}(^{15}N)$ = 40 μs, $p_{180}(^{15}N)$ = 80 μs, a_0 = 50 ms, a_1 = 4.

```
;noesyhsqcf3gpsi3d
;avance-version (12/01/11)
;NOESY-HSQC
;3D sequence with
;   homonuclear correlation via dipolar coupling
;   dipolar coupling may be due to noe or chemical exchange.
;   H-1/X correlation via double inept transfer
;      using sensitivity improvement
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho-TPPI gradient selection (t2)
;with decoupling during acquisition
;using trim pulses in inept transfer
;using f3 - channel
;(use parameterset NOESYHSQCF3GPSI3D)
;
;A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
;   Reson. 93, 151-170 (1991)
;L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
;   10663-5 (1992)
;J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky,
;   S.J. Glaser, O.W. Sorensen & C. Griesinger, J. Biomol. NMR 4,
;   301-306 (1994)
;
;$CLASS=HighRes
;$DIM=3D
```

```
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
"p22=p21*2"
"d11=30m"
"d12=20u"
"d13=4u"
"d26=1s/(cnst4*4)"


"d0=3u"
"d10=3u"

"in0=inf1/2"
"in10=inf2/2"


"DELTA1=d13+p16+d16+4u"

#   ifdef LABEL_CN
"DELTA=p16+d16+larger(p2,p14)+d10*2"
```

```
"DELTA2=larger(p14,p22)+d0*2"
#   else
"DELTA=p16+d16+p2+d10*2"
"DELTA2=p22+d0*2"
#   endif /*LABEL_CN*/

"TAU=d8-p16-d16+d0*2*d9"


aqseq 321


1 ze
  d11 pl16:f3
2 d11 do:f3
3 d12
  d1
  (p1 ph8)
  DELTA2 pl3:f3
  (p2 ph9)
  d0

#   ifdef LABEL_CN
  (center (p14:sp3 ph1):f2 (p22 ph1):f3 )
#   else
  (p22 ph1):f3
#   endif /*LABEL_CN*/

  d0
```

```
  (p1 ph10)
  TAU UNBLKGRAD
  p16:gp1
  d16
  (p1 ph11)
  d26
  (center (p2 ph1) (p22 ph6):f3 )
  d26
  p28 ph1
  d13
  (p1 ph2)
  3u
  p16:gp2
  d16
  (p21 ph3):f3
  d10

#   ifdef LABEL_CN
  (center (p2 ph7) (p14:sp3 ph1):f2 )
#   else
  (p2 ph7)
#   endif /*LABEL_CN*/

  d10
  p16:gp3*EA
  d16
  (p22 ph4):f3
  DELTA
  (center (p1 ph1) (p21 ph4):f3 )
```

```
  d24
  (center (p2 ph1) (p22 ph1):f3 )
  d24
  (center (p1 ph2) (p21 ph5):f3 )
  d26
  (center (p2 ph1) (p22 ph1):f3 )
  d26
  (p1 ph1)
  DELTA1
  (p2 ph1)
  d13
  p16:gp4
  d16 pl16:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2
     F1PH(calph(ph8, +90) & calph(ph9, +90), caldel(d0, +in0))
     F2EA(calgrad(EA) & calph(ph5, +180), caldel(d10, +in10) & calph(ph3,
+180) & calph(ph6, +180) & calph(ph31, +180))
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
```

```
ph8=0  0  0  0  2  2  2  2

ph9=1  1  1  1  3  3  3  3

ph10=0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
     2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2

ph11=0  0  0  0  0  0  0  0  2  2  2  2  2  2  2  2

ph31=0  2  2  0  2  0  0  2  2  0  0  2  0  2  2  0
     2  0  0  2  0  2  2  0  0  2  2  0  2  0  0  2



;pl1 : f1 channel - power level for pulse (default)

;pl3 : f3 channel - power level for pulse (default)

;pl16: f3 channel - power level for CPD/BB decoupling

;sp3: f2 channel - shaped pulse 180 degree (adiabatic)

;p1 : f1 channel -  90 degree high power pulse

;p2 : f1 channel - 180 degree high power pulse

;p14: f2 channel - 180 degree shaped pulse for inversion (adiabatic)

;p16: homospoil/gradient pulse                          [1 msec]

;p21: f3 channel -  90 degree high power pulse

;p22: f3 channel - 180 degree high power pulse

;p28: f1 channel - trim pulse                           [1 msec]

;d0 : incremented delay (F1 in 3D)                     [3 usec]

;d1 : relaxation delay; 1-5 * T1

;d8 : = a0 (srNOESY base mixing time)

;d9 : = a1 (srNOESY linear t1 factor)

;d10: incremented delay (F2 in 3D)                     [3 usec]

;d11: delay for disk I/O                               [30 msec]

;d12: delay for power switching                        [20 usec]

;d13: short delay                                      [4 usec]

;d16: delay for homospoil/gradient recovery
```

```
;d24: 1/(4J)YH for YH
;     1/(8J)YH for all multiplicities
;d26: 1/(4J(YH))
;cnst4: = J(YH)
;inf1: 1/SW(H) = 2 * DW(H)
;inf2: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(H)) = DW(H)
;nd0: 2
;in10: 1/(2 * SW(X)) = DW(X)
;nd10: 2
;ns: 8 * n
;ds: >= 16
;td1: number of experiments in F1
;td2: number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence


;use gradient ratio:    gp 1 : gp 2 : gp 3 : gp 4
;                         30 :   50 :   80 :  8.1      for N-15
;                         30 :   50 :   80 : 20.1      for C-13

;for z-only gradients:
;gpz1: 30%
;gpz2: 50%
;gpz3: 80%
;gpz4: 8.1% for N-15, 20.1% for C-13
```

```
;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
;gpnam4: SMSQ10.100


                                          ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;             option -DLABEL_CN (eda: ZGOPTNS)
                                          ;preprocessor-flags-end


;$Id: noesyhsqcf3gpsi3d,v 1.5.8.1 2012/01/31 17:56:34 ber Exp $
```

Vita

Name	Charles Frederick DeLisle	
Education	B.S., Biomedical Science, Grand Valley State University	2012
	B.S., Chemistry, Grand Valley State University	2012
	M.S., Chemistry, University of Illinois at Chicago	2014
	Ph.D., Biochemistry, University of Illinois at Chicago	2019
Awards	Merit Award for Excellence, Grand Valley State University	2008-2011
	Hensel & Jeanette McKee Undergraduate Scholar, Delta Sigma Phi National Fraternity	2011
	Thomas M. Seykora Award for Outstanding Campus Contributions, Grand Valley State University	2012
	Hensel & Jeanette McKee Graduate Scholar, Delta Sigma Phi National Fraternity	2012, 2016
	Chancellor's Student Service & Leadership Award, University of Illinois at Chicago	2016
	Crowd Favorite Presentation, UIC Chemistry Research Symposium	2018
	Graduate College Research Travel Award (Paris, FR), University of Illinois at Chicago	2019
	Graduate Student Council Research Travel Award (Paris, FR), UIC Graduate Student Council	2019
Publications	**DeLisle, C. F.**, Mendis, H. B., Lorieau, J. L. (2019) Super Resolution NOESY Spectra of Proteins. *J. Biomol. NMR* 73(3): 105-116.	

DeLisle, C.F., Malooley, A.L., Banerjee, I., Lorieau, J.L. (2019) Pro-islet Amyloid Polypeptide in Micelles Contains a Helical Prohormone Segment. *FEBS J.* Under Revision.

Presentations

GVSU Student Scholar's Day — Allendale, MI "Synthesis of Multidentate CMPO Ligands for Heavy Metal Ion Chelation"	2011
Argonne Undergraduate Symposium — Lemont, IL "Synthesis of Multidentate CMPO Ligands for Heavy Metal Ion Chelation"	2011
West Michigan Regional Undergraduate Science Research Conference — Grand Rapids, MI "Synthesis of Water Soluble Multidentate CMPO Ligands for Lanthanide Chelation"	2011
GVSU Chemistry: Senior Seminar — Allendale, MI "Synthesis of Water Soluble Multidentate CMPO Ligands for Lanthanide Chelation"	2012
Chicago Biomedical Consortium — Chicago, IL "Super Resolution NMR"	2018
UIC Chemistry Research Symposium — Chicago, IL "Making NOESY Spectra Super Resolution"	2018
NMR: A Tool for Biology XI — Paris, FR "Super Resolution NOESY Spectra of Proteins"	2019
Chicago Biophysics Networking Symposium — Chicago, IL "Micelle-bound pro-Islet Amyloid Polypeptide Contains a Carboxy-terminal Helical Segment"	2019
Gateway NMR — Ann Arbor, MI "Micelle-bound pro-Islet Amyloid Polypeptide Contains a Carboxy-terminal Helical Segment"	2019
Chicago Area NMR Discussion Group — Madison, WI "Structure of pro-Islet Amyloid Polypeptide in Micelles Contains a Helical Prohormone Segment"	2019