Biomolecular NMR spectroscopy: Application to the study of the piRNA-pathway protein GTSF1, and backbone and side-chain spin relaxation methods development

Paul O'Brien

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University

2019

ABSTRACT

Biomolecular NMR spectroscopy: Application to the biophysical study of piRNA-pathway protein GTSF1, and backbone and side-chain spin relaxation methods development

Paul O’Brien

The structural dynamics of proteins and other macromolecules typically serve crucial roles for their respective biological function. While rigid protein structures are used in classic “lock and key” descriptions of enzymology and receptor-ligand interactions, more and more evidence suggest that the majority of molecular interactions occur on the spectrum between induced-fit binding and conformational selection binding. This model of biomolecular interaction requires, to differing degrees, conformation plasticity and dynamics of the protein itself. To characterize the determinants and implications of protein dynamics, there exists no more suited biophysical technique than nuclear magnetic resonance (NMR) spectroscopy. This method is capable of probing the individual atomic nuclei of proteins in a site-specific manner. Furthermore, NMR spectroscopy is unique in being able to access timescales from picoseconds to seconds, providing information on events from bond vibration and libration to protein folding and ligand binding. The breadth of biophysical information accessible by NMR spectroscopy has led to its widespread use in the study of protein dynamics. The work presented herein involves i) the use of NMR for investigation of structure and dynamics in two separate biological systems that demonstrate a high

degree of flexibility for folded proteins and ii) the improvement of pulse sequences and methodology for better characterizing picosecond to nanosecond backbone and side-chain dynamics. The organizing principle of this work, which is best exemplified in the structural studies of the piRNA-pathway protein Gametocyte-specific factor 1, is the unmatched capability of NMR spectroscopy to decipher molecular details within dynamic protein systems.

First, the molecular structure and RNA-binding properties of gametocyte-specific factor 1 (GTSF1) of the piRNA effector pathway were investigated. A partially disordered protein with two Zn finger domains, the work presented here describes the isolation of a GTSF1 protein construct amendable to study by NMR spectroscopy. Chemical shift assignment of GTSF1 allowed site-specific observation of amide correlations, which established the basis for NMR structure calculation of GTSF1 and the evaluation of binding to candidate RNA sequences, with goal of the identification of an *in vivo* RNA binding partner for GTSF1. The work presents compelling data that indicate GTSF1 Zn finger 1 specifically binds a motif *GGUUC(G/A)* RNA, which in this study was found in the T-arm loop of transfer RNA. Zn finger 2 is affected by the interaction with RNA, but the available structural and binding data indicate that the second Zn finger is a more dynamic, breathable entity, supported by cysteine chemical shift and structural differences between the two GTSF1 Zn fingers. Although it's currently speculative, the function of GTSF1 might first require binding of RNA to the more stable Zn finger 1, which then leaves Zn finger 2 poised for binding to another molecular species. tRNA-derived fragments that include the T-arm TΨC loop have been recently implicated in silencing of transposable elements in mammalian cells. GTSF1, which was identified in a genetic screen for piRNA-pathway proteins as vitally required for gene silencing, might plausibly act as a sensor of transcription of transposable elements and help initiate Piwi-piRISCs-mediated chromatin modification and heterochromatin formation.

Next, NMR spectroscopy is used to investigate protein thermostability in psychrophilic (cold-loving) cytochrome *c*$_{552}$. Isolated from the bacterium *Colwellia psychrerythraea (Cp)*, previous work has implicated two conserved *Cp*cyt *c*$_{552}$ methionine residues, which are both conserved across psychrophilic and psychrotolerant cytochromes, as acting in dynamical ligand substitution with a third methionine that is the axial heme ligand. It is proposed that elevated backbone dynamics in these methionine residues and the ability for them swap into the axial ligand position accounts for an uncharacteristically high melting temperature (T_m) compared to meso- and thermophile *c*-type cytochromes. Progress was made in NMR sample preparation and backbone chemical shift assignment of both redox states of *Cp*cyt *c*$_{552}$, and insight from 1D ^{1}H NMR experiments focused on the heme group bound to *Cp* cytochrome *c*$_{552}$ is discussed. Additionally, chemical shifts are used to predict protein dynamics as a first test of a multiple methionine axial ligand hypothesis. Initial data analysis predicts relatively large measures of Random Coil Index for residues surrounding the native axial heme ligand, and shows the hyperfine shifts localized to the residues surrounding the heme. Future experiments will selectively record methyl group dynamics of methionine residues for elucidation of rate constants of methionine substitution and to determine the structural properties of this minor conformation.

Finally, two NMR methodology studies are presented in this thesis: a novel simultaneous-acquisition TROSY pulse sequence for measurement of backbone spin relaxation rates (R_1 and $\{^1H\}$-^{15}N heteronuclear NOE) and a side-chain ^{2}H spin relaxation method for using multifield experimental datasets for better sampling of the spectral density function. Together, these pulse sequences represent significant advancements in NMR measurement of microscopic rate constants and more nuanced detail of protein dynamics.

Table of Contents

1 Background

2 Materials and Methods

3 The Study of Structural and RNA-binding Properties of GTSF1

3.4 GTSF1 M1-D80 Structure Shows 2 Zn fingers Connected by an α-helix linker

3.5 Hold on Tight: The Curious Case of RNA-binding to GTSF1

4 TROSY Pulse Sequence for Simultaneous Characterization of Backbone Dynamics in Deuterated Proteins

5 Enhanced Spectral Density Mapping through Combined Multi-field Deuterium ^{13}CH$_2$D Methyl Spin Relaxation NMR Spectroscopy

6 NMR Studies of Localized Dynamics in Psychrophilic Cytochrome c_{552} From *C. Psychrerythraea*

List of Figures

List of Tables

Acknowledgements

I would like to take the opportunity to recognize the many people who have been vitally important to my growth both as a scientist and as a person over the course of graduate school. Getting to this point would not have been possible without them.

First, I want to express my thanks and gratitude to Art Palmer for guidance, mentorship, and for being a vast resource of knowledge and advice all these years. Your support and confidence in me have kept me motivated and enthusiastic even after yet another NMR sample issue or the magnet suddenly shutting off. I also appreciate the responsibility and freedom he has afforded me to pursue a varying set of ideas and experiments.

Secondly, I thank my committee members Wayne Hendrickson and Ann McDermott. I am appreciative for both your interest and insightful comments in the varied work I have carried out during my graduate career. Additionally, I thank the members of my thesis committee, Kevin Gardner and Israel Sanchez Fernandez, for taking a Friday afternoon to hear about my passion for NMR.

Third, I want to recognize Palmer Lab members past and present. Palmer lab iteration one included Jae-Hyun Cho, Michelle Gill, Paul Harvilla, Paul Robustelli, Ying Li, Kate Stafford, and Tim Zeiske. Their contributions to both my intellectual growth and as members of productive and friendly workplace were much appreciated. As a bright-eyed and naïve NMR spectroscopist, I was especially lucky to learn from Ying and Michelle, whose patience and guidance I am thankful for. Palmer lab iteration two has included Hans Koss, Liz Kish, and James Martin. I enjoyed both the warm workplace, friendly banter, and collaborating on certain biological and NMR projects. I look forward to seeing the innovative, exciting developments in NMR methodology from the Palmer lab in the coming years. I also humbly bequeath liquid nitrogen fills of the spectrometer to all three of you. Use this power wisely.

Fourth, I thank my collaborator Jon Ipsaro from Leemor Joshua-Tor's laboratory Cold Spring Harbor. Your work in preparation of NMR protein sample, accompanying biochemical experiments, and bioinformatics work to clarify RNA binding was vital to the GTSF1 story (Chapter 3). I feel especially privileged to have been allowed the opportunity to explore this interesting biological story with you.

Finally, a special dedication to the many family and friends who have been there for me throughout graduate school and shared New York City during the past years: John, Christine, and William O'Brien, Stu and Fran Ehrlich, Seth Robey, Ryan Underwood, Liz Nagle, Mel Hullings, Jason Schifman, Sean Engles, John Goldston, and to a host of others. Your support, confidence in me, and assistance, big or small, was integral in getting me this far. Most importantly, I am thankful for the support, perseverance, and love of my partner Samantha Ehrlich. I am ecstatic and excited to have made you so proud with this accomplishment.

Chapter 1

Background

The study of biology overwhelmingly begins with static descriptions of biomolecules. From classic undergraduate biochemistry lessons in glycolysis, gluconeogenesis, and metabolic biochemistry, to oxidative phosphorylation, receptor biochemistry and general enzymology, the page-bound image of a protein or enzyme blob reacting with a ligand is etched into the mind. Dig deeper and the two-dimensional description of an enzyme or receptor becomes the three-dimensional crystal structure of ribonuclease H or receptor tyrosine kinase. Deeper yet is a physiological understanding of biological macromolecules that more closely resembles a ship afloat at sea: as Å size biomolecules, tumbling in a solvent of ions and small molecules, with rigid structural elements replaced by a continuum of molecular flexibility that contribute to biological function. It is through this lens that biology is at its most deconstructed yet astonishing. The numerous studies in this thesis are moored by the use of nuclear magnetic resonance (NMR) spectroscopy as a tool to integrate molecular structure and dynamics for a more complete biophysical characterization of proteins and enzymes. Especially as the contribution of disordered or partially disordered proteins to the human proteome increase, the ability to specifically characterize the molecular dynamics becomes even more powerful. This thesis has two themes: i) methods development in NMR spectroscopy and ii) application of NMR spectroscopy to dynamic protein systems.

1.1 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful experimental technique that is governed by the interaction between atomic nuclei and electromagnetic radiation. Since the research projects detailed in this thesis take advantage of a cross-section of NMR spectroscopy experiments. To help guide readers, a certain fluency in NMR spectroscopy will be established. To begin, NMR active nuclei, that is atomic nuclei with non-zero spin (**I**), possess a magnetic dipole that interacts with an external magnetic field (B_0). Within this static field, magnetic dipoles populate a number of discrete, quantized orientations defined by $2\mathbf{I} + 1$, where **I** corresponds to the nuclei spin number. The potential energy of a particular orientation ("state") within an externally-applied field is given by:

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B}$$

where E is defined as the potential energy of the magnetic dipole, $\boldsymbol{\mu}$ is the magnetic dipole vector, and $\boldsymbol{B}$ is the external magnetic field vector. If the external field is oriented along the z-axis, the equation for potential energy can be reduced to the relevant z-component by:

$$E = -\mu_z \cdot B_0 = -\gamma \hbar m B_0$$

where μ_z is the z-component of the magnetic dipole interaction, B_0 is the external magnetic field aligned along the z-axis in the laboratory reference frame, γ is the nuclei-specific gyromagnetic ratio, which is a ratio of the magnetic moment to the angular momentum of the particle, $\hbar$ is the reduced Planck constant, and m corresponds to the quantum number of a particular state of the particle. The Boltzmann distribution describes the population of these different, quantized energy states by:

$$\frac{N_a}{N_b} = -e^{\left(-\frac{\Delta E}{kT}\right)}$$

where N_a and N_b are the populations of two different energy states, ΔE is the difference in energy between those two states, and kT, which is formally the product of the Boltzmann constant k and the system temperature T, serves the role of a discrete scale factor for energy.

To illustrate the preceding concept, typical solution-state NMR spectroscopy makes use of spin ½ (**I** = ½) nuclei. For this example, ^{1}H nuclei and a B0 field of 14.1 T will be used. Two states are predicted to be populated (2**I** + 1): $m = +½$ and $m = -½$. The energy difference between the two states is given by:

$$E = -\mu_z \cdot B_0 = -\gamma\hbar m B_0 = -\gamma\hbar(m_a - m_b)B_0$$

calculating a value for the energy difference between states and plugging this back into the above equation illustrates a fundamental property of NMR spectroscopy: when placed in an external magnetic field, there exists a small net magnet dipole aligned along the direction of B_0. Furthermore, choosing NMR active nuclei with larger gyromagnetic ratios or increasing the external magnetic field are two straightforward ways to increase sensitivity of NMR spectroscopy. On the other hand, the very small energy difference between populations makes NMR a relatively insensitive technique compared to other fluorescence and infrared (IR) spectroscopy. However, the experimental insensitivity of NMR spectroscopy is mitigated by the extremely high resolution and subsequent insight that is obtained from nuclei-specific probing of molecules and proteins.

In the archetype NMR experiment, the alignment of bulk magnetization along the *z*-axis (typically defined as the direction of the external magnetic field) serves as the starting point. This net +*z*-magnetization is perturbed by another magnetic field, which typically comes from a radiofrequency pulse passed through a coil in close proximity. This new field exerts a force on the bulk magnetization to tilt it away from the +*z*-axis to, say, the +*x*-axis. This newly aligned magnetization, which will be referred to as $\mathbf{M}_x$, now will precess around the z-axis at a frequency defined by:

$$\omega_0 = -\gamma B_0$$

where γ is the nuclei-specific gyromagnetic ratio, B_0 is the external magnetic field aligned along the *z*-axis in the laboratory reference frame, and ω_0 is referred to as the Larmor frequency. As inferred from the equation, changes in gyromagnetic ratio (differences between ^{1}H ^{13}C, ^{15}N, etc.) and magnetic field both influence the Larmor frequency. This will be revisited shortly. Going back to the consideration of $\mathbf{M}_x$ magnetization: with net magnetization precession in the transverse (*xy*) plane, the system is perturbed and exists in a non-Boltzmann thermodynamic state. Two processes act on the magnetization to restore magnetization to its equilibrium state. The first process is called spin-lattice relaxation (also known as non-adiabatic relaxation), by which equilibrium magnetization conditions are restored in the direction of the external magnetic field (in this case, the *z*-axis). This is an exponential process, and the recovery back to its equilibrium state is described by the rate constant R_1 ($1/T_1$). Spin-lattice relaxation occurs by exchange of energy between between the nuclei and its surroundings. The second process is called spin-spin relaxation (also known as adiabatic relaxation) and is described by the rate constant R_2 ($1/T_2$). As $\mathbf{M}_x$ precession occurs in the transverse plane and the electrical current it induces in the coil of a neighboring probe is detected, spin-spin relaxation leads to so called dephasing of magnetization in the *xy*-plane and damping of the recorded signal. This dephasing is caused by random, time-dependent fluctuations in the local magnetic field. These fluctuations affect the individual Larmor frequencies and lead to the loss of coherence between spins (nuclei). To properly contextualize these effects, a [I_z - 90_y°] sequence leads to I_x transverse magnetization that is initially affected by

R_2 relaxation (5-20 ms timescale for a 10-20 kDa protein) before slowly being affected by R_1 relaxation (400-600 ms timescale for a 10-20 kDa protein).

Another central concept of NMR spectroscopy is the chemical shift. The aforementioned high resolution that is possible in biomolecular NMR experiments comes from the rich diversity of protein topology that create unique chemical environments for nuclei. These different environments can drastically affect the chemical shift: unpaired electrons of a metal ion can cause large hyperfine shifts, while the aromatic ring current can likewise drastically affect a chemical shift. Chemical shielding is catch-all term for the differing local magnetic fields sensed by nuclei and it directly affects the Larmor frequency of the nucleus in question by:

$$\omega = -(1 - \sigma)\gamma B_0$$

where σ is termed the chemical shielding, and the overall chemical shift is affected by shielding or de-shielding chemical environments. From NMR structure calculations to secondary structure prediction and Random Coil Index prediction, a wealth of structural and dynamical information is encoded in the chemical shifts of proteins, and will be utilized extensively in this thesis.

Typically, NMR experiments are carried out at more than just a single magnetic field. The Larmor frequency, and thus the frequency that the nuclei will resonate at in the NMR spectrum, changes with increasing magnetic field strength. To normalize resonance frequencies across experiments, frequencies are converted into the chemical shifts:

$$\delta = 10^6 \times \frac{\omega - \omega_{ref}}{\omega_{ref}}$$

where the chemical shift, δ, is expressed in parts per million (ppm). An additional consideration for comparing NMR chemical shifts across field strengths and to chemical shift databases is referencing. 2,2-methyl-silapentane-5-sulfonate, or DSS, is commonly used as a reference for biomolecular NMR. The resonance frequency of DSS's methyl protons is set zero ppm, and used to linearly adjust the remaining proton spectrum or proton dimension. For heteronuclei referencing, the gyromagnetic ratio of proton and the heteronuclei are used to adjust the carrier frequency for the heteronuclei. This process is known as internal, indirect referencing.

Before describing representative NMR experiments in more detail, the practical aspects of NMR spectroscopy will be briefly explained. First, a "well-behaved" sample, which is code for a protein sample that remains in solution, doesn't precipitate over time, and has low ionic conductivity (less than 500 mM salt), is matched and tuned, then shimmed. These steps tune the probe to the resonant frequency and match the impedance of the sample, and produce a homogenous magnetic field across the size of the NMR sample, respectively. Hard pulse lengths for all nuclei (^{1}H, ^{13}C, ^{15}N, etc.) are measured and shaped pulse lengths are determined. Pulses are applied at specific time intervals from a pre-compiled pulse sequence through use of a frequency synthesizer and RF transmitter. Equilibrium magnetization is perturbed, the pulse sequence executed, and transverse magnetization detected through a receiver coil. A time domain signal then passes through a preamplifier, phase modulator, and is converted to a digital signal. Several signal processing steps can then be applied: a window function, zero-filling of the data, zero-order and first-order phase corrections, and Fourier transform. This last step transforms a time domain signal to a frequency domain signal. The result is a spectrum, with peaks corresponding to the frequency (or energy levels) of different nuclei.

As a guide to the work in this thesis, three different NMR experiments will be considered here. The discussion is ordered by experimental dimensionality. First, one-dimensional proton experiments are described. Next is the ubiquitous two-dimensional ^{15}N heteronuclear single quantum coherence experiment (HSQC), which serves as the foundational pulse sequence for both backbone resonance assignment and spin relaxation experiments. Lastly, the multidimensional nuclear Overhauser effect spectroscopy (NOESY) experiment, used in this thesis for measurement of through-space proton-proton interactions, will be described. This introduction to experimental NMR spectroscopy should at once convey the expertise and ambition brought to fruition in this thesis as well as illustrate the level of qualitative and quantitative characterization possible for biological systems with this spectroscopic technique.

To begin, a basic one-dimensional experiment will be described. Assuming protons are the only NMR-active nuclei, an experiment starts with a hard 90° pulse that transforms equilibrium $\mathbf{M}_z$ magnetization to either $\mathbf{M}_x$ or $\mathbf{M}_y$ depending on the phase of the hard pulse. Residing in the transverse plane, magnetization precesses under the Zeeman Hamiltonian. Hypothetically, the current induced in a nearby coil from transverse magnetization could be immediately detected, and the time domain data Fourier-transformed into a spectrum. However, almost all protein and enzyme NMR samples are in aqueous solution, which means the concentration of solvent water protons (55.5 M) far exceeds solute protons (0.1-1 mM) of interest and overwhelms the detection of signal from solute as well as corrupting other experimental parameters. This is a recurring issue in almost all of solution-state NMR spectroscopy: suppression of solvent signal, which is almost exclusively from water-derived protons, is an absolute requirement for usable and intelligible NMR spectra. Several methods for water suppression exist. All schemes rely on strategic perturbation of the water proton resonance (~4.7 ppm), either by presaturation, gradient dephasing, or selective pulses that result in excitation of solute protons only while solvent protons are left undetected. Here, a version of the latter water suppression scheme will be discussed. Specifically, a jump-return pulse sequence uses a [$90_y°$ - τ - $90_{-y}°$] pulse block for detection [1]. With the carrier frequency placed on the solvent proton resonance and τ equal to:

$$\tau = \frac{1}{(4\Delta\nu_{max})}$$

where $\Delta\nu_{max}$ is the difference in frequency between the solvent resonance and the frequency of maximum excitation, this pulse sequence block results in no transverse solvent magnetization, complete water suppression, and detection of solute protons, with stronger peak signal intensities +($\Delta\nu_{max}$) from the carrier frequency. Jump-return read pulses are especially useful in situations where ^{1}H nuclei exchange rapidly with solvent. An additional pulse block consisting of a refocusing Hahn-echo sequence minimizes effects from magnetic field inhomogeneities. This thesis takes advantage of 1D ^{1}H jump-return Hahn echo experiments in Chapter 3 (Figure 1.1.1), where different RNA candidates are screened for their specific protein binding capabilities [2]. The imino protons of RNA (guanosine and uracil), which typically are only weakly observed with traditional one-dimensional proton NMR experiments, can be identified clearly with jump-return experiments as well as characteristic hydrogen-bonded chemical shifts [3]. Jump-return pulse sequence blocks, along with other water suppression methods, make one-dimensional NMR spectroscopy a powerful technique despite its seemingly low complexity. For heteronuclear experiments with increased dimensionality, NMR experiments are designed with combinations of pulse field gradients and selective pulses to accomplish water suppression. Examples of this in a 2D relaxation experiment are further described in Chapter 4.

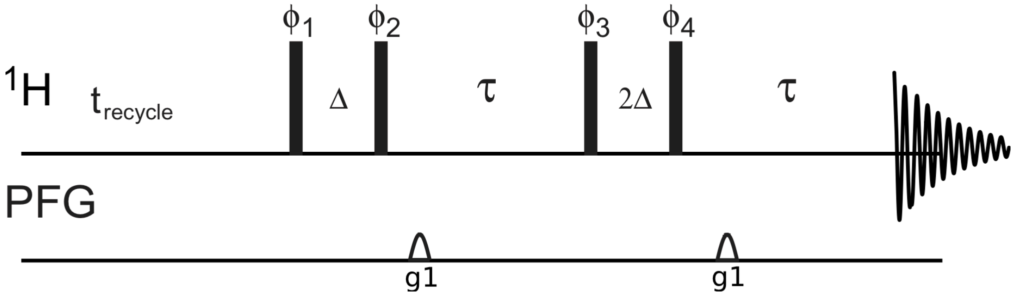


Figure 1.1.1: **1D ^{1}H Jump-Return Hahn Echo NMR Experiment.**
1D pulse sequence for ^{1}H jump return Hahn echo experiment. Relaxation delay $T = 2\tau$. Delay Δ is set to $1/(4^*\Delta\nu)$. The phase cycle is ϕ_1=(x), ϕ_2=(-x), ϕ_3=(x,y,-x,-y) and ϕ_4=(-x,-y,x,y); gradient g1 = 1 ms. Experiments recorded with a recycle delay of 1.5 seconds. For an example 1D 1H Jump-Return HE spectrum, see Figure 3.5.5. [2]

Next, two-dimensional ^{15}N HSQC spectroscopy is considered. This heteronuclear NMR experiment, which correlates amide proton magnetization with their attached nitrogen, is either used individually or serves as the starting point for multiples spectroscopic studies in this thesis: diagnostic ^{15}N HSQC spectra are used for identification of a suitable NMR sample and ^{2}J ^{15}N HMQC experiments are used to identify the side-chain histidine tautomers in Chapter 3 [4], backbone ^{15}N spin relaxation pulse sequences begin and end with HSQC-like magnetization transfer elements in Chapter 5, and two-dimension spectra and backbone chemical shift experiments with redox cytochrome in Chapter 6 utilize proton-nitrogen heteronuclear transfer. Compared to one-dimensional NMR spectroscopy, a 2D HSQC allows the alleviation of peak overlap through an increase in dimensionality of the experiment. Additionally, by way of polarization transfer from higher gyromagnetic ^{1}H to lower gyromagnetic ^{15}N (or ^{13}C) through large one-bond (^{1}J) or two-bond (^{2}J) scalar couplings, the experimental sensitivity is not substantially sacrificed. In a ^{15}N HSQC experiment, excited proton magnetization is transferred by ^{1}J scalar coupling interactions to an attached nitrogen, where the magnetization is t_1 labeled. Another pulse sequence element selects for conversion of labeled nitrogen magnetization back into proton magnetization, where it is directly detected in t_2. Varying the length of t_1 produces a 2D interferogram, which becomes a 2D spectrum after Fourier transform of both time dimensions. The result is a spectrum with proton-nitrogen correlations that allows confirmation of peak dispersion during NMR sample condition optimization, chemical shift perturbations for amide resonances upon addition of a ligand, as well as a host of other observations. For an example of a ^{15}N HSQC spectrum with assigned proton-nitrogen resonances, see Figure 3.3.1.

Different coherence transfer mechanisms, such as heteronuclear single quantum coherence (HSQC), heteronuclear multiple quantum coherence (HMQC), and transverse relaxation optimized (TROSY), can be used for polarization transfer between hetero-nuclei depending on the molecular size of the system, chemical moieties (deuterated background, methyl correlation, etc.), or relaxation properties of the biomolecular system. Not only do these heteronuclear correlation schemes lead to different two-dimensional experiments, they form the building blocks for multi-dimensional NMR experiments. For example, three-dimensional triple-resonance assignment experiments are used in Chapter 3 and Chapter 6 [5]. These pulse sequences start with coherence transfer from proton to attached nitrogen using the above description, but then add a coherence transfer step to carbon nuclei (typically C$^{\alpha}$, C$^{\beta}$, and CO). The third dimension in such experiments comes from a third labeling period (t_l) while the carbon magnetization of interest is on the transverse plane. Then, coherence transfer steps are run in reverse to take the ^{13}C$^{\alpha,\beta}$/^{13}CO – ^{15}N – ^{1}H^{N} pathway for detection of the NMR signal. As with the two-dimensional ^{15}N HSQC, the individual time domains are Fourier transformed and processed to yield a three-dimensional frequency spectrum. Why is a more complex pulse sequence so vital? Because, again, increases in the dimensionality of data resolves issues of peak overlap and degeneracy that occur in lower dimensional datasets. Triple resonance assignment experiments are extremely useful for elucidating the identify of peaks in a ^{15}N HSQC. Additionally, the information implicit in the chemical shifts of ^{1}H, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO contribute to qualitative understandings of molecular structure and dynamics. Finally, in addition to serving as the building block for multi-dimensional NMR experiments, heteronuclear correlation methods also serve as the initial step in spin relaxation experiments. Results in Chapter 4 and Chapter 5 demonstrate this: viewed as the preparation part of the spin relaxation pulse sequence, pure $^{15}N_z$ magnetization (Chapter 4) and proper deuterium magnetization (Chapter 5) are necessary for accurate measurement of various relaxation rates.

Third, the heteronuclear NOESY experiment is described [6]. This pulse sequence again incorporates a heteronuclear correlation. When used in Chapter 3 of this thesis, the heteronuclear correlation scheme used is a HSQC element (Figure 3.4.2 – 3.4.8 demonstrate the use of ^{15}N heteronuclear NOE data in NMR structure calculation). However, the pulse sequence contains a homonuclear NOESY block preceding the HSQC. This block consists of t_1 labeling scheme and a NOESY mixing period, τ_m. During the mixing time, the magnetization is transferred between protons according to the:

$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_H^4 \hbar^2}{10} \frac{\tau_c}{r_{IS}^6} \left[-1 + \frac{6}{1 + 4\omega_0^2 \tau_c^2}\right]$$

where γ_H is the nuclei-specific gyromagnetic ratio, ω_0 is the Larmor frequency, τ_c is the rotational correlation time of the molecule, μ_0 is the permeability of free space, $\hbar$ is the reduced

Planck constant, and r_{IS} is the distance between spin $\boldsymbol{I}$ and spin $\boldsymbol{S}$. To simplify it, the nuclear Overhauser effect is a through-space magnetization transfer between spins heavily dependent on i) the gyromagnetic ratio, ii) the tumbling time of the molecule, and iii) the physical distance between spins. Compared to the length of spin irradiation in the steady-state NOE experiment, the NOESY mixing time is typically set much shorter to avoid spin diffusion (100 ms). After the mixing period, the NOESY-HSQC consists of a heteronuclear transfer from proton to nitrogen by ^{1}J scalar coupling, followed by t_2 frequency labeling, reverse coherence transfer back to proton followed by detection (t_3). In Chapter 3, three separate NOESY-HSQC experiments were recorded to measure NOE crosspeaks between amide protons, aliphatic, and aromatic protons to surrounding protons. These crosspeaks serve as restraints for derivation of the protein structure [6]. More detail and explanations regarding structure calculations are included in Chapter 3.

As a last point, it's prudent to highlight the versatility of NMR spectroscopy. From characterization of molecular motion on timescales faster than the rotational correlation time (picoseconds to nanoseconds timescale) by measurement of the R_1, R_2, and $\{^{1}H\}$-^{15}N NOE and characterization of the spectral density function [7], to measurement of ligand binding, catalysis, and folding (microseconds to seconds timescale) by $R_{1\rho}$, CPMG, or ZZ-exchange[8]. Figure 1.1.2 illustrates this experimental utility.

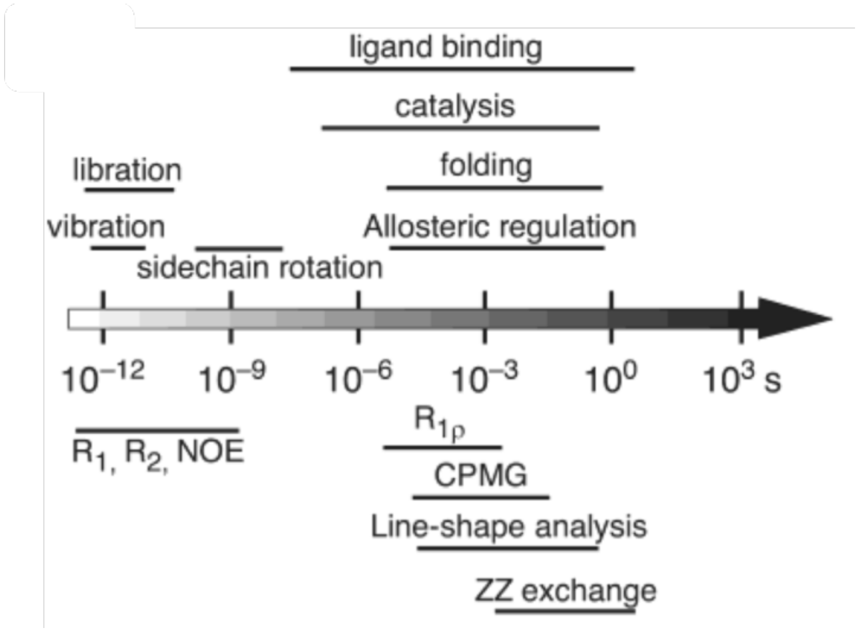


Figure 1.1.2: **Dynamics Regimes Accessible by NMR Spectroscopy.** Experimental NMR spectroscopy allows observation of molecular dynamics spanning the timescale from seconds to picoseconds. Relaxation rates R_1, R_2, and heteronuclear NOE permit characterization of dynamics faster than the rotational correlation (τ_c) of the molecule (10^{-9} seconds). Molecular dynamics that affect the chemical shift or the linewidth are observable by ZZ exchange, line-shape analysis, Carr-Purcell-Meiboom-Gill (CPMG), or $R_{1\rho}$ NMR experiments. These experiments measure dynamics parameters capable of characterizing protein-ligand interactions, folding, and enzyme catalysis. Reprinted with permission from Chemical Reviews 104: 3623 3640. Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules [9]. Copyright 2004 American Chemical Society.

1.2 RNA Interference and piRNA

1.2.1 Discovery and Mechanistic Description of RNAi

RNA interference (RNAi) is a process by which non-coding RNA acts through different mechanisms to regulate gene expression. Published work, beginning with the studies of Fire and Mello in 1998 [10], built on previous observations of RNA-mediated gene regulation to explicitly define RNA interference in *C. elegans*. Specifically, double-stranded RNA exhibited potent, specific, and amplified (non-stoichiometric) gene knockdown effects [11]. The study pointed to a mechanism more complex more than just an antisense model. In the avalanche of research that followed, the mechanisms of RNAi were further delineated and its presence identified in multiple organisms beyond *C. elegans*.

Three main types of non-coding RNA have been identified: small-interfering RNA (siRNA), microRNA (miRNA) [12], and piRNA [13]. These non-coding RNA types differ in their trigger transcript length, nucleotide modification, source of trigger, and function [14]. Table 1.2.1 highlights these differences. What these three types of non-coding RNA do share is a similar mechanistic engine for RNAi: the RNA-induced silencing complex (RISC) is formed between the single-stranded guide RNA (mature siRNA, miRNA, or piRNA) and an Argonaute (Ago) protein, which is an RNA-binding protein combining sequence complementarity between guide and target RNA with RNaseH-like cleavage activity [15]. Ago proteins have evolved for each individual lineage of non-coding RNA. Furthermore, guide RNA-loaded Ago proteins serve as input for the different methods of gene silencing.

ncRNA Type	Transcript Length	Chemical Modification	Source of Trigger	Function
miRNA	21 nucleotides	No significant enrichment	Intronic- and exonic-sequences; 70 nt stem-loop structure	Sequence complementarity and translational inhibition
siRNA	20-25 bp dsRNA	5' phosphorylation; 3' hydroxylation; 2 nt overhang on each strand	Heterochromatin-derived sequences	Post-transcriptional gene silencing
piRNA	26-31 nucleotides	5' bias for uridine; 2'-O-methylation at the 3' end	piRNA clusters	Transcriptional gene silencing, ping-pong cycle

Table 1.2.1: **Types of non-coding RNA, their Chemistry, and Function.**
The three major types of ncRNA are miRNA, siRNA, and piRNA. Their characteristics and chemistries differ, but all RNAi pathways utilize a biogenesis step to produce mature transcript (Dicer family proteins) and an effector step that uses the RNAi signal to affect gene expression (Argonaute family proteins) [16].

1.2.2 RNAi: Biogenesis

RNAi is best understood by splitting the pathway in half and categorizing by each's mechanistic goal: i) biogenesis and ii) effector steps of RNAi. The first step of biogenesis is distinct between miRNA, siRNA, and piRNA. For miRNA, a source of transcripts is from intronic and exonic regions of the genome [17]. RNA polymerase II typically transcribes a region of RNA that folds back on itself and makes a short hairpin. Primary miRNA is also produced, which consists of several stem-loop precursors assembled into one long, linear nucleotide sequence. In the case of siRNA, long double-stranded RNA (dsRNA) is the source of transcripts [18]. These dsRNAs can come from viral nucleic acid [19], transgenic introduction of RNA [20], or transcription by RNA polymerase [21]. Finally, piRNA biogenesis starts with production of long, single-stranded RNA sequences derived from piRNA clusters that function as the template for individual guide piRNA transcripts [22]. Second, the mechanism for conversion to mature mi/si/piRNA transcript differs between the three RNAi pathways. At a general level, both miRNA and siRNA use a similar type of processing mechanism for production of mature transcripts. Two proteins, Drosha and Pasha, separate primary RNA into precursor miRNA (pre-miRNA), which are approximately 70-nucleotide long stem-loop fragments [23]. Pre-miRNA is then exported to the cytoplasm, where it binds to the endonuclease protein Dicer and double-stranded RNA is cleaved, separated, and the miRNA guide strand is loaded onto Ago protein. Conversely, double-stranded siRNA transcripts are exported from the nucleus, processed by the endonuclease protein Dicer to mature dsRNA, and the guide strand shunted to Ago protein [24]. Compared to miRNA and siRNA, biogenesis of piRNA is conceptually distinct. To begin with, the process does not involve activity of the Dicer protein. Instead, the transcriptional products of piRNA clusters, which are typically derived from repeat and transposon-like genetic elements, are exported from the nucleus and localize to perinuclear or perimitochondrial regions of the cytoplasm. Termed the nuage, these areas surrounding the nucleus and mitochondria are locations with molecular machinery that function in additional processing of piRNA cluster transcripts [25]. Specifically, the endonuclease protein Zucchini is an HKD family phosphodiesterase that has been shown to possess single-strand RNA binding and cleavage functionality [26]. The cleaved product from Zucchini is a processed piRNA transcript that is loaded onto Piwi-clade Ago3 protein. Here, the maturation of piRNA transcripts is completed: trimming of the piRNA [27] occurs through the action of a 3'-to-5' exonuclease (Trimmer in *B. mori*; PNLDC1 in mammals), while the *S*-adenosyl methionine (SAM)-dependent methyltransferase HEN1 adds 2'-O methylation to the 3' end of the piRNA transcript [28]. Of note, there is a difference between germline and secondary piRNA pathways. Germline-specific biogenesis of mature piRNA transcripts occurs as detailed above. A secondary piRNA biogenesis pathway adds so-called "ping-pong amplification" on top of the primary piRNA processing [29]. The exact importance of the ping-pong cycle and its contribution to overall piRNA-mediated gene silencing is not fully understood [30]. A rough outline depicts the mature piRNA-Ago complex interacting with a counterpart piRNA-Aubergine (piRNA-Aub) complex. Overlap in a section of at the 5' ends of these piRNAs allows for refined slicing of the nascent piRNA to give a mature piRNA that is antisense to the genomic target sequence. Consequently, the presence of both mature sense and antisense piRNA guides adds to the gene silencing potential of piRNA[31]. The implications of this will be developed in further detail when the effector steps of piRNA-driven RNAi are discussed.

1.2.3 RNAi: Effector Steps

Similar to the diversity in RNAi biogenesis mechanisms, the effector steps, which are defined here as the biochemical pathways that use mature RNAi transcripts to accomplish gene silencing, differ among the three main noncoding-RNA classes. Two will be described briefly (miRNA and siRNA), while more time is devoted to piRNA effector steps, which set the stage for studies in Chapter 3 of this thesis. Figure 1.2.1 shows detailed schematics for each of these pathways.

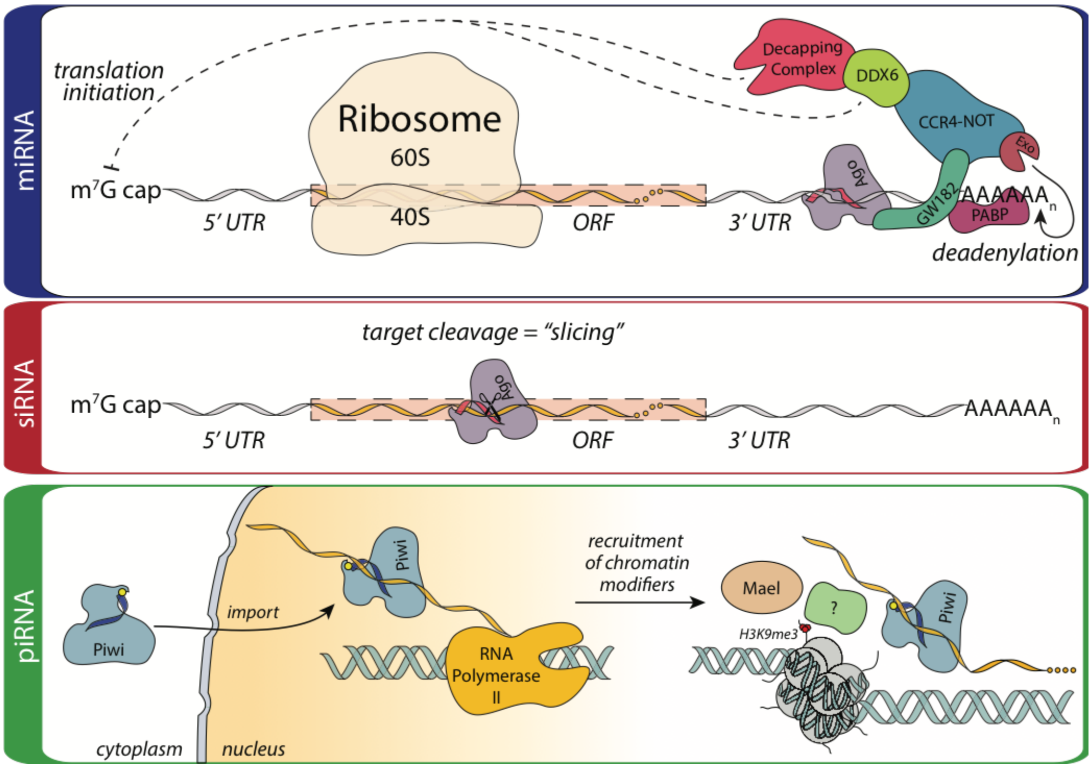


Figure 1.2.1: **Effector Steps of Different RNAi Pathways.**
The three RNAi effector steps vary in the mode by which they accomplish gene silencing. miRNA-bound Ago associates with mRNA to promote deadenylation, which is mediated by GW182 and PABP. Alternatively, RISC binding to the m^{7}G cap of mRNA promotes translational inhibition. Post-transcriptional gene silencing by siRNA pathways relies on guide strand complementarity with target mRNA, followed by Ago-mediated cleavage of target mRNA. Finally, nuclear import of piRNA-bound Piwi complexes precedes association with nascent mRNA transcripts, localization of chromatin modifying enzymes to target genomic locations, and formation of heterochromatin. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature Structural and Molecular Biology, 22, Ipsaro, J.J., Joshua-Tor, L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. 20-28, Copyright 2015. [32]

With mature RNA transcripts in tow, both miRNA- and siRNA-mediated RISCs achieve gene silencing by exclusively targeting mRNA corresponding to the target gene. In the miRNA pathway, this occurs through imperfect base-pairing between the guide sequence and mRNA target. Silencing of mRNA by miRISCs proceeds by repression of translation (interference with initiation of translation machinery) or deadenylation of mRNA's poly(A) tail sequence and subsequent destabilization [33, 34]. On the other hand, siRISCs match their mRNA target exactly, with perfect base-pairing between the guide sequence and mRNA target [35, 36]. Once bound, the catalytic functionality of Ago slices the mRNA into two smaller nucleotides along a phosphodiester bond in the phosphate backbone. The remnants of siRISC action are further degraded by cytoplasmic nucleases. For both mi- and siRISCs, the Ago protein portion of the complex serves as conduit for the gene silencing signal [37].

The effector steps of piRNA RISCs are distinct from those of miRISCs and siRISCs. Though the contribution of it does differ between organisms, the most notable distinction of the piRNA RNAi response is the predominance of transcriptional gene silencing (TGS) over physical cleavage of the mRNA target driven by sequence complementarity [38]. As noted earlier, one implication of the ping-pong cycle is that antisense transcripts are available for use as templates for binding and slicer activity towards the mRNA target. However, several studies have demonstrated that the principal contributor to piRNA-mediated gene silencing comes from TGS [39]. A rudimentary mechanism depicts piRISCs (Piwi-clade Ago bound to mature piRNA) imported to the nucleus, where it localizes to sites of piRNA target sequence active transcription [32]. Observations show DNA, heterochromatin, and chromatin-modifying factors interacting with piRISCs to methylation marks and gene silencing of piRNA clusters. Bringing clarity to the rather opaque piRNA-induced gene silencing step is the focus of Chapter 3 of this thesis.

1.2.4 Delineation of piRNA-mediated TGS

Before adding more detail to the piRNA effector step, the driving force behind piRNA-mediated RNAi should be elaborated on. Compared to miRNA and siRNA, the primary function of the piRNA class of non-coding RNA is to silence transposable elements in germline cells, which include retrotransposons and DNA transposons [40]. The potential deleterious effects from these mobile genetic elements "jumping out" of the genome and reinserting back in random locations makes their regulation highly important. This is particularly important in germline cells, where transcriptional markers, as measured by DNA methylation, exist in a derepressed state, which can potentially invite aberrant transposon activity.

To recapitulate, piRNA-bound RISCs are imported into the nucleus and localize to transcriptionally-active piRNA clusters. The complex is then able to locate the DNA tract for silencing. Concurrently, several protein factors are recruited to this bridged site. Maelstrom and chromatin-modifying proteins function to transmit silencing marks to the genomic DNA through methylation of heterochromatin [41]. Conversely, gametocyte-specific factor 1 (GTSF1) [42] serves an important but as of yet unclear role in silencing of transposable elements. The role of nuclear piRNA in suppression of transposable elements will be explored in three different animal systems, which serve as a representation of the diversity of piRNA effector steps.

D Melanogaster The piRNA pathways in germline cells are well characterized in *Drosophila*. To better understand the robustness of piRNA, it's helpful to compare piRNA-directed silencing in somatic follicle cells and in oocytes [31, 43]. In the context of somatic follicle cells, transcription of the piRNA *flamenco* cluster primes the PIWI-mediated response to *gypsy* family long terminal repeat (LTR) retroviruses. This predominantly accomplished by a TGS mechanism, which includes H3K9 methylation of heterochromatin and association with HP1A. Alternatively in oocytes, both ping-pong amplification and TGS mechanisms combine to eliminate transposon mRNA and transfer silencing marks to heterochromatin, respectively, in response to assorted transposon stimuli.

M. Musculus The piRNA-mediated transposon response in mouse germline cells is markedly different than *Drosophila*. Mouse germline cells exists in a derepressed state (hypomethylation). Therefore, global regulation of transposons by establishment of nuclear H3K9 methylation marks is the main piRNA effector step in mice [44, 45]. MIWI is a mouse version of PIWI protein and is implicated as the main effector protein in this instance as well. piRISCs primed with guide piRNAs then act to silence the multitude of active transposons.

C. Elegans The 21U small RNA pathway is thought to be the *C. elegans* version of the piRNA pathway [46, 47]. Again, no evidence of ping-pong amplification has been observed. Biogenesis steps involve production of 22G small RNAs from 21U small RNAs, while effector steps involve 22G small RNA import into the nucleus along with WAGO-9 and WAGO-10, which are both Ago/Piwi homologues. The RNA-WAGO-9/10 complex localizes to active transcription sites, where it associates with nascent transcripts and results in H3K9 methylation of heterochromatin. Two histone methyltransferases, SET-25 and SET-32, catalyze establishment of silencing marks [48]. Finally, the *C. elegans* protein HPL-2 plays the role of HP1A. It recognizes H3K9 methylation marks and stabilizes heterochromatin [49].

The TGS step of piRNA-mediated RNAi is notably shared among these three examples. Yet, the proteins involved and mechanisms underlying the propagation of the transposon silencing signal remain not well characterized.

1.2.5 Gametocyte-Specific Factor 1: A Dual CHHC Zinc Finger Protein

One effector protein appears common to all piRNA effector steps and is conserved across invertebrates and vertebrates. Gametocyte-specific factor 1, or GTSF1, is a 167 amino acid protein containing two zinc finger domains of the Cys-His-His-Cys class (Figure 1.2.2) [50]. These two zinc binding domains account for approximately 25% of the entire protein. The remaining portion of GTSF1 has no gene annotation nor is it predicted to have any secondary structure (assigned as random coil from primary sequence prediction algorithms). Furthermore, these ~130 remaining residues are enriched in charged amino acid types (aspartate/glutamate and lysine/arginine) as well as proline residues. Both of these observations provide evidence for characterization of the non-zinc finger portion of GTSF1 as an intrinsically disordered region.

^{1}MEDTYIDSLDPEKLLQ**C**PYDKN**H**QIRACRFPY**H**LIK**C**RKNHPDVANKLAT**C**PFNAR**H**QVPRAEISH**H**ISS**C**DDKSCIEQDVVNQTRNLGQETL^{93}
^{94}AESTWQCPPCDEDWDKDLWEQTSTPFVWGTASFCGNNSPANNIVMEHKSNLASGMRVPKSLPYVLPWKNNGNAQ^{167}

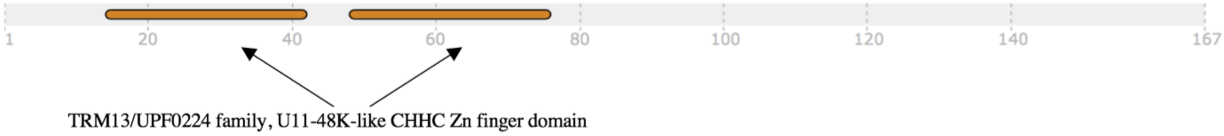


Figure 1.2.2: **Gametocyte-specific Factor 1: Primary Sequence and Homology.** From its primary sequence, GTSF1 is predicted to contain two Zn finger domains: one comprising residues 14-41, and the other comprising residues 48-75. The suspected Zn ligands are bolded and underlined. The conserved **C**-($X_{3\text{-}5}$)-**H**-(X_{9})-**H**-(X_{3})-**C** Zn finger domain shows homology to the TRM13 methyltransferase family and the U11-48K minor spliceosomal protein. (InterPro software for protein sequence analysis and classification; IPR007917)

The gene coding for GTSF1, also known as Asterix, was discovered in RNAi screens in *Drosophila* for previously uncharacterized piRNA-pathway factors [42, 51]. These screens were carried out in cultured ovarian somatic sheet (OSS) cells, which express all three classes of non-coding RNA. One notable observation was that specific knockdown of GTSF1 led to upregulation of transposable element markers, indicating that the cell environment was derepressed. A second important observation was that gene knockdowns of Piwi protein resembled knockdown of GTSF1 in both levels and type of transposon derepression. This was shown in both somatic and germline cells. Third, experimental results confirmed GTSF1's exclusive role in the transposon silencing step: expression levels nor subcellular localization of piRNA biogenesis factors (Piwi, Aub, Ago3) were affected, and these results strongly mimicked the effects seen from Maelstrom knockdown, which is protein similarly implicated in the transposon silencing step. Finally, experiments in mice showed knockout of mouse GTSF1 causes sterility in male mice and increased transcription of LINE-1 and IAP retrotransposons through loss of methylation in their respective promoter regions. These genetic experiments confirm the importance of the GTSF1 protein to piRNA-mediated TGS *in vivo*.

Two papers contain the bulk of biochemical characterization for GTSF1. In one, three takeaways add clarity to the C-terminal functionality of GTSF1 (Figure 1.2.3) [52]. First, compared to both a C-terminal deletion construct and a zinc-finger mutation construct, only transfection of full-length GTSF1 plasmid is able to rescue transposon silencing (*mdg1* and *gypsy*) as assessed by transcript level in a GTSF1 knockdown background. Second, immunoprecipitation (IP) experiments using GTSF1, Piwi, and a C-terminal peptidomimetic for pull-down of GTSF1 demonstrate that a GTSF1 associates with Piwi protein through its C-terminal region. Specifically, two aromatic amino acids, W89 and Y98, are implicated as governing the interaction with Piwi protein. Lastly, IP experiments using increased concentrations of peptidomimetic (1x, 2.5x, 10x) exhibited increasing pull-down of Piwi protein. This last observation speaks to the large pool of Piwi capable of binding GTSF1. The second paper again finds evidence for the association between Piwi and GTSF1(Figure 1.2.4) [53]. Additionally, the authors carry out measurements of *mdg1* expression in GTSF1-depleted somatic cells transfected with GTSF1 Zn finger mutations. Cysteines of Zn finger 1 and Zn finger 2 are mutated to alanine. Both intra- and inter-Zn finger mutations show increased levels of expression, with even a single point mutation of one cysteine failing to rescue wild-type GTSF1 transposon silencing.

In summary, the results of genetic and biochemistry experiments implicate the dual Zn finger protein GTSF1 as occupying an essential role in the effector step of the piRNA pathway. The Zn finger motifs and the C-terminal disordered segment clearly play a role in depositing methylation marks and formation of heterochromatin in genomic areas corresponding to the piRNA in piRISC; investigating the molecular underpinnings of this process is the focus of Chapter 3 of this thesis.

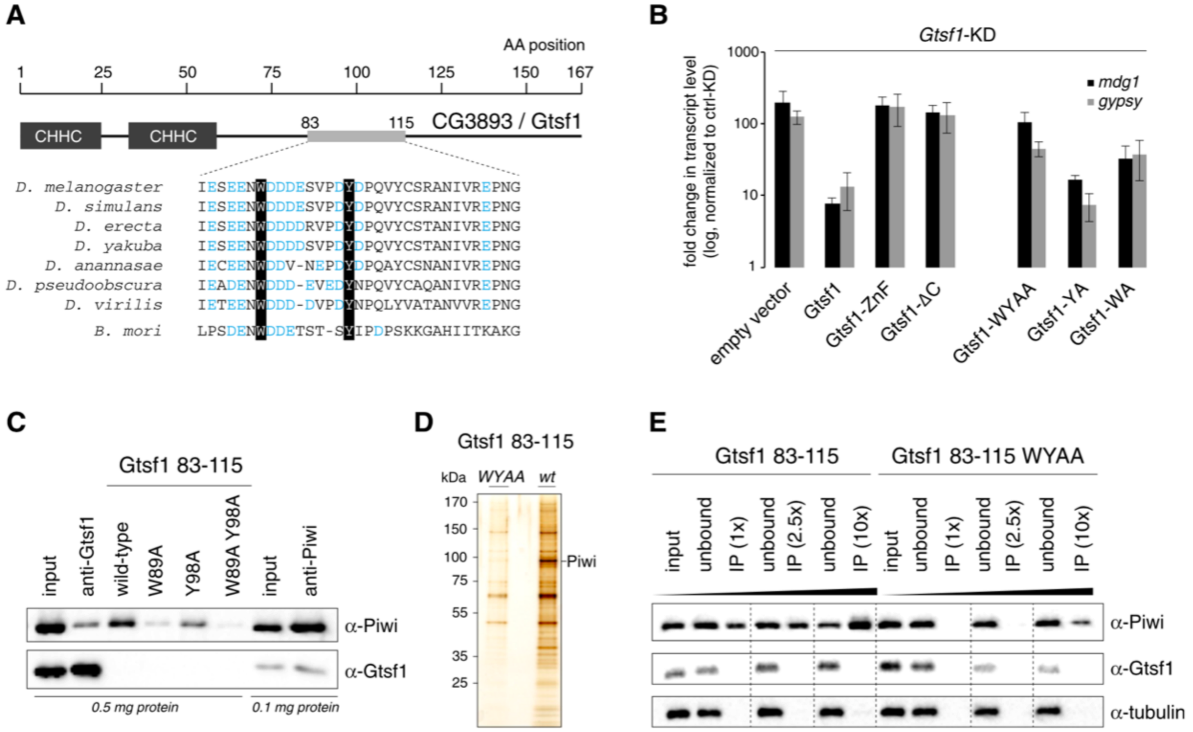


Figure 1.2.3: **The C-terminal Residues of *Drosophila* GTSF1 Associate with Piwi.** Two aromatic residues, W89 and Y98 in *Dm*GTSF1 (A), prove essential for Piwi association by immunoprecipitation with a peptide mimetic of GTSF1 C-terminal residues (C). Point mutations W89A and Y98A both reduce the pulldown of Piwi protein, with W89A qualitatively showing the most effect. Pulldown with the double mutant W89A/Y98A peptide entirely abrogates Piwi binding (D). Finally, *Dm*GTSF1-Y98A is the only mutation among W89A, W89A/Y98A, DC, and ZnF mutation experiments that rescues transposon repression compared to wild-type in a GTSF1 knockdown background (B). Reprinted by permission from Cold Spring Harbor Laboratory Press: Genes and Development, 27, Donertas, D., et al., *Drosophila* GTSF1 is an essential component of the Piwi-mediated transcriptional silencing complex. 1693-1705, Copyright 2013. [52]

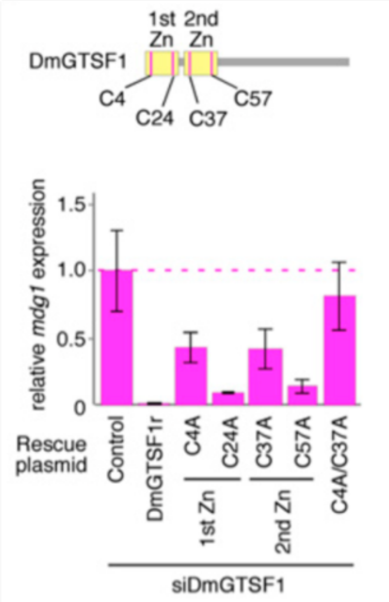


Figure 1.2.4: **Effect of *Dm*GTSF1 Zn finger Mutation on Transposon Silencing.** The *Drosophila* GTSF1 construct differs slightly from that of *M. musculus*. Alanine mutations of individual Zn finger cysteine residues resulted in loss of between 20% and 40% transposon control, depending on the mutated cysteine, as judged by RT-qPCR of *mdg1* transcripts. Furthermore, cysteine-to-alanine mutations of both the first cysteine residues of each Zn finger show an additive effect on loss of transposon silencing. Reprinted by permission from Cold Spring Harbor Laboratory Press: Genes and Development, 27, Ohtani, H., et al., DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. 1656-1661, Copyright 2013. [53]

Chapter 2

Materials and Methods

2.1 Nuclear Magnetic Resonance Spectroscopy

2.1.1 Protein Expression and Purification

Samples of gametocyte-specific factor 1 (GTSF1) for NMR experiments, including U-[^{15}N] and U-[^{13}C, ^{15}N] GTSF1, were prepared using standard protein expression and purification protocols. A plasmid coding for the *Mus musculus* mmGTSF1 gene was used to transform BL21 (DE3) expression-line *E. coli*. Transformed colonies on LB-Agar plus antibiotic were used to inoculate LB liquid cultures. After overnight growth, the cultures were spun down, the LB media decanted, and 50 mL M9 minimal media cultures were inoculated and grown at 37°C until OD_{600} reached 0.5. These 50 mL cultures were then spun down, the minimal media decanted off, and the bacterial pellet was resuspended in 1 L of M9 minimal media prepared with 5 g per L of ^{13}C-glucose and 2 g per L of ^{15}N ammonium chloride and supplemented with 10 mg of both biotin and thiamine. The 1 L M9 minimal media expression culture was grown at 37°C until OD_{600} read 0.7. Addition of IPTG to a final concentration of 500 μM induced GTSF1 expression. Cells were harvested after overnight protein expression.

Purification of GTSF1 commenced with cell lysis and sonication. Cells were lysed into a (contents) buffer. After sonication, crude lysate was cleared by centrifugation at 14K rpm for one hour. The supernatant was applied to freshly charged HisTrap-HP columns (GE Life Sciences) and GTSF1 with His-tag bound to the column. After wash with buffer plus 5 mM imidazole, GTSF1 was eluted with a linear gradient of 5 mM to 500 mM imidazole. Elution fractions containing GTSF1 protein were pooled, and A_{280} was measured and an SDS-PAGE gel was run to judge recombinant protein concentration and purity, respectively. Pooled fractions were dialyzed into a buffer containing 50 mM TrisHCl (pH 7.5) and 0.5 mM EDTA, then incubated with commercially purchased TEV protease in the presence of 1 mM DTT. Finally, size exclusion chromatography was used for removing the TEV protease, cleaved His-tag, and any remaining impurities. Gel-filtration chromatography in 50 mM MES (pH 6.5) buffer gave a single, monodisperse peak. Elution fractions corresponding to the peak were pooled and final protein concentration was determined by measurement of A_{280}. All purification steps were done at 4°C.

Final yields for expressed and purified GTSF1 constructs were: GTSF1.FL (4 mg/mL), GTSF1.125 (15 mg/mL), and GTSF1.115 (14 mg/mL). To prepare samples for NMR experiments (^{15}N and ^{13}C, ^{15}N), purified GSTF1 samples were either: i) buffer exchanged into 50 mM MES (pH 6.5), 200 mM NaCl, 2:1 stochiometric addition of $ZnCl_2$, and 10 mM TCEP with 10% D_2O; or ii) lyophilized and resuspended in a 99% D_2O solution with 50 mM MES (pD 6.5), 200 mM NaCl, 2:1 stochiometric addition of $ZnCl_2$, and 10 mM TCEP. Final concentrations for GTSF1 samples used in NMR experiments were 0.5 mM for U-[^{15}N] GTSF1 and 0.8 mM for U-[^{13}C, ^{15}N] GTSF1.

2.1.2 ^{15}N HSQC Test of Multiple Protein Constructs and Different Buffer Conditions

Proper solvent conditions are a requirement for successful NMR spectroscopy experiments [54]. A first attempt at recording NMR experiments for U-[^{15}N] GTSF1 in 20 mM Tris (pH 8.0) and 200 mM NaCl resulted in a 2D ^{15}N HSQC spectrum with little chemical shift dispersion, sharp peaks totaling less than the number expected for a 115-residue protein construct, and lacking other indications of a folded, well-behaved protein sample. To fix this, several solvent conditions were optimized: buffer, pH, TCEP, and metal ligand. Following parts of the solvent screening method by Lepre and Moore [55], we first optimized across buffers. These included TRIS (pH 8.0), HEPES (pH 7.5), Na^+ phosphate (pH 7.0), MES (pH 6.5 and pH 6.0), and Na^+ acetate (pH 5.5). All buffering agents were used at 100 mM, while ionic strength was maintained at 200 mM NaCl. The reducing agent TCEP was added for final sample concentrations of 1 mM, 5 mM, and 10 mM from a stock solution that was previously adjusted to the NMR sample pH. Finally, $ZnCl_2$ was added to give final stoichiometric ratios of 1:1 and 2:1 Zn ion to protein from a stock solution buffered to the corresponding pH of the sample. Peak dispersion and intensity were used as parameters to rank spectra from solvent optimization.

2.1.3 Backbone Resonance Assignment of GTSF1

A Bruker DRX600, AVANCE800 and AVANCE900 (5 mm TXI, 5 mm TCI, and 5 mm TCI cryoprobes, respectively) were used to acquire a suite of three-dimensional backbone resonance assignment experiments for GTSF1. Sample temperatures across all experiments was calibrated to 298 K using 98% $^2H_4$-methanol [56]. 100 μM DSS was included in samples for internal referencing of ^1H chemical shifts, followed by indirect referencing for ^{13}C and ^{15}N chemical shifts. In total, eight different NMR backbone assignment experiments were collected. Aggregate peak information from all spectra were necessary for near-complete backbone resonance assignment of GTSF1. Table 2.1 lists the relevant NMR experiments for backbone chemical shift assignment as well as experimental details.

Backbone Assignments	**Spectral Width (kHz)**	**TD ($t_3 \times t_2 \times t_1$)**	**Scans (NS)**
HNCA	$10.8 \times 2.4 \times 7.2$	$1024 \times 64 \times 80$	8
HN(CO)CA	$10 \times 2.1 \times 6.4$	$1024 \times 64 \times 100$	8
HNCO	$10 \times 2.7 \times 3.4$	$1024 \times 64 \times 128$	8
HN(CA)CO	$12 \times 2.8 \times 3.0$	$2048 \times 40 \times 60$	16
CBCANH	$12.6 \times 3.3 \times 17$	$2048 \times 64 \times 100$	32
CBCA(CO)NH	$12.6 \times 3.3 \times 17$	$2048 \times 64 \times 80$	48
HNCACB	$11.1 \times 3.2 \times 15$	$2048 \times 64 \times 180$	32
HN(CO)CACB	$12.6 \times 3.3 \times 17$	$2048 \times 40 \times 128$	56
^{15}N HSQC	11.1×2.4	2048×300	8
^{13}C HSQC	8.8×15.1	1024×256	16

Table 2.1: **Backbone Triple-Resonance Experiments Recorded for Assignment of GTSF1 H^N, N, C^α, C^β, and CO Chemical Shifts.**

2.1.4 Side-chain Resonance Assignment of GTSF1

Side-chain chemical shift assignment of GTSF1 was less straightforward than backbone assignment. Initially, HCCH-TOCSY experiments were recorded, and C^α & C^β chemical shift assignments were used to correlate side-chain carbon nuclei and then side-chain proton nuclei. However, the large peak intensity disparity as well as minimal resolution for certain amino acid residue types made for spectra that were difficult to decipher. Instead, three experiments were used for side-chain ^{1}H and ^{13}C assignment: HBHA(CO)NH, H(CCCO)NH, and (H)C(CCCO)NH. These experiments take advantage of the high spectral resolution in the nitrogen dimension (with only a few exceptions) for GSTF1 spectra. Additionally, by correlating all *i*th-1 side-chain protons (or carbons) to the *i*th amide nitrogen and proton, these three experiments serve as a check on the backbone chemical shift assignments since many of the amino acids show a diagnostic side-chain proton (or carbon) footprint. Finally, constant time ^{13}C HSQC and ^{13}C SOFAST-HMQC spectra were recorded. Table 2.2 lists the relevant NMR experiments for side-chain chemical shift assignment as well as experimental details.

Side-chain Assignments	**Spectral Width (kHz)**	**TD ($t_3 \times t_2 \times t_1$)**	**Scans (NS)**
HBHA(CO)NH	$8.4 \times 1.9 \times 8.4$	$1024 \times 64 \times 220$	8
H(CCCO)NH	$8.4 \times 1.9 \times 7.7$	$1024 \times 64 \times 256$	16
(H)CC(CO)NH	$8.4 \times 1.9 \times 12.7$	$1024 \times 64 \times 128$	16

Table 2.2: **Triple-Resonance Experiments Recorded for Assignment of GTSF1 Side-Chain ^{1}H and ^{13}C Chemical Shifts.**

2.1.5 Aromatic Resonance Assignment of GTSF1

Aromatic protons and carbons were assigned through the combined use of aromatic constant time ^{13}C HSQC and non-constant time ^{13}C HSQC spectra, with the carrier frequency set to 125 ppm in the indirect dimension for both experiments, aromatic and aliphatic ^{13}C NOESY-HSQC spectra, and a modified ^{15}N HMQC experiment that relies on ^{2}J couplings between aromatic protons and imidazole ring nitrogen. HBCBCGCDHD/HBCBCGCDHD experiments were recorded as well, but the signals observed originated only from the C-terminal residues of GTSF1.

2.1.6 NMR Restraints for Structure Calculations

NOESY-HSQC experiments were recorded with U-[^{15}N] GTSF1 and 0.8 mM for U-[^{13}C, ^{15}N] GTSF1. To guarantee the highest digital resolution possible in t_1, all three NOESY-HSQC experiments were run with 8 or 16 scans. All three experiments were recorded at 18.8 T (800 MHz) at the New York Structural Biology Center. Finally, the sample buffer for the ^{15}N NOESY experiment was 90% H_2O/10% D_2O, while the two 13C NOESY-HSQC experiments were recorded with GTSF1 in 99% D_2O. Table 2.2 lists the relevant NMR experiments for side-chain chemical shift assignment as well as experimental details.

NOESY Crosspeak	**Spectral Width (kHz)**	**TD ($t_3 \times t_2 \times t_1$)**	**Scans (NS)**
^{15}N NOESY	9.6 × 2.3 × 9.6	1024 × 64 × 512	8
^{13}C NOESY aliphatic	11.1 × 14 × 11.2	1024 × 128 × 400	8
^{13}C NOESY aromatic	8.8 × 8.0 × 9.6	1024 × 64 × 128	16

Table 2.3: **3D NOESY-HSQC Experiments for Assignment of GTSF1 NOE Cross Peaks.**

2.1.7 NMR Spin-Relaxation Experiments

Backbone dynamics were probed by NMR spin relaxation experiments. Three experiments were recorded to measure the following relaxation rates: ^{15}N R_1 rates, ^{15}N R_2 rates, and the steady-state $\{^1H\}$-^{15}N NOE. A 0.5 mM U-[^{15}N] GTSF1 sample was used for measurement of relaxation rates. The suite of relaxation experiments used the pulse sequences of Lakomek, et al. [57]. For the steady-state $\{^1H\}$-^{15}N NOE experiment, experimental duplicates were used for error estimation. For measurement of ^{15}N longitudinal and transverse relaxation rates, three duplicate relaxation delays were used for error estimation. Table 2.4 lists the relevant NMR experiments for side-chain chemical shift assignment as well as experimental details.

Spin Relaxation Expt.	**TD ($t_3 \times t_2 \times t_1$)**	**Relaxation Delays (ms)**	**Scans (NS)**
R_1 Relaxation	1024×256	24, 176, 336, 496, 656, 816, 976, 1200	16
R_2 Relaxation	1024×256	16, 32, 48, 65, 98, 130, 163, 196	16
$\{^1H\}$-^{15}N hetNOE	1024×512	10 second delay for ^1H saturation	48

Table 2.4: **Backbone ^{15}N Spin-Relaxation Experiments for GTSF1.**

2.1.8 RNA-Protein Titrations by NMR

RNA binding was assessed by monitoring two-dimensional ^{15}N HSQC and ^{13}C SOFAST-HMQC [58] spectra in the absence and presence of RNA. Typically, a U-[^{15}N] or U-[^{15}N,^{13}C] GTSF1 purified protein stock was split into two samples. Wilmad NMR tubes, which hold approximately 500 μL of NMR sample, were used for NMR titration experiments. Several μL volumes of RNA, which were refolded into NMR buffer at high temperature, were added to 500 uL GTSF1 protein stocks with a Hamilton syringe. RNA was added to give a final 1:1 stoichiometry for protein:RNA. These two samples were then used to record respective without/with RNA NMR spectra. In addition to two-dimensional heteronuclear NMR experiments, one-dimensional proton jump-return experiments were recorded to qualitatively evaluate RNA folding. ^1H Jump-return experiments are especially well-suited to detect nucleic acid imino proton peaks.

2.1.9 Processing and Data Analysis

One-, two-, and three-dimensional NMR spectra were processed using NMRPipe. Visualization of three-dimensional backbone assignment strips, peak picking, fragment building, and fragment matching was carried out using Sparky. All NMR spin relaxation experiments were analyzed in Sparky; peak intensities in ^{15}N R_1 and R_2 experiments were fit to a monoexponential decay, while peak intensities from saturated and unsaturated heteronuclear NOE experiments were exported from Sparky for visualization in Python.

2.2 GTSF1 RNA Binding Assays*

**Collaboration with Jon Ipsaro and Leemor Joshua-Tor (CHSL)*

GTSF1 RNA binding was assessed using two different experiments. The first was more qualitative; the overarching goal was to confirm co-purifying RNA binding to GTSF1 and test the nucleic acid binding capabilities of different length GTSF1 constructs. Five different GTSF1 constructs were used: residues 1-45 (Zn finger 1 only), residues 1-75 (Zn finger 1 and 2), residues 1-167 (full-length), residues 46-167 (includes Zn finger 2), and residues 76-167 (includes neither Zn finger). SF9 cells were transfected with plasmid encoding a GTSF1 construct, protein expression induced, and then cell lysate was applied to a Ni-NTA column. The sample was denatured on-column and the resulting flow-through monitored for nucleic acid elution. Finally, co-purifying nucleic acid was analyzed on a urea-PAGE gel.

2.3 RNA Library Construction and Bioinformatics Analysis*

**Collaboration with Jon Ipsaro and Leemor Joshua-Tor (CHSL)*

An initial attempt at RNA library construction started with ssRNA that co-purified with GTSF1 from protein expression in SF9 insect cells. This heterogenous RNA was used as a starting point for amplification, barcoding of reads, and sequencing. A second attempt focused on constructing libraries of GTSF1-binding RNA using an enhanced UV-crosslinking and immunoprecipitation (eCLIP). Compared to typical UV-crosslinking and immunoprecipitation, eCLIP experiments required an order of magnitude less amplification yet still retains single-nucleotide binding resolution. Most importantly, the UV-crosslinking step of an eCLIP experiment guarantees that identified RNA-protein interactions are the result of specific interactions between RNA and

GTSF1. A quick summary of an eCLIP experiment follows [59]. First, UV-crosslinking affects RNA-binding protein (RBP), which in this case is GTSF1, and any RNA that its bound to. Cells are then lysed and fragmented. 2% of the sample is saved as size-matched input, while the remaining 98% is used immunoprecipitated and washed. The sample is then dephosphorylated and ligated with a 3' adapter. A protein gel electrophoresis and membrane transfer, a portion of the gel is excised corresponding to the molecular weight of GTSF1 plus 75 kDa. This isolates any bound RNA up to ~220 nucleotides. The gel is then solubilized and treated with proteinase K to digest protein and leave only its partner RNA. Several steps are then carried out on this RNA to prepare paired-end sequencing libraries: remaining RNA from above is converted to DNA by incubation with reverse transcriptase and the RNA strand removed, a 3' adapter is ligated, and the final single strand DNA is PCR-amplified and size-selected to give a library structure of (5'-3') read 1 element, in-line barcode, DNA (corresponding to RNA fragment), random-mer, and read 2 element. RNA libraries were sequenced in duplicate. In-house Python scripts were used for analysis of raw read data. A more in-depth discussion of analysis of the generated RNA library will be included in the appendix.

Chapter 3

Study of Structural and RNA-binding Properties of GTSF1

3.1 Motivation

Besides the two predicted RNA-binding regions [50], little is known about the molecular structure of GTSF1 nor the functionality of the Zn fingers and C-terminal domain. Earlier studies attempting to crystallize GTSF1 found little success (Ipsaro, J.J., et al. *unpublished*); the relatively large disordered region of GTSF1 presumably inhibits formation of crystal contacts and consistent crystal lattice formation. As an alternative, subsequent research pivoted to characterization of GTSF1 using nuclear magnetic resonance spectroscopy (NMR). This change in experimental approach was motivated by three unique aspects of NMR spectroscopy. First, despite the lack of secondary structure, the study of disordered proteins is still possible with NMR spectroscopy [60, 61]. Nuclei populating so-called random coil states are easily identified by characteristic chemical shifts, narrow linewidths, and strong peak intensities. In a protein such as GTSF1, with a suspected combination of random coil segments and residues in secondary structure, elements distinguishing between the two types of resonances should be apparent even before NMR triple-resonance backbone assignment experiments are recorded. Second, NMR observables measured with two-dimensional and three-dimensional experiments should allow a description of both structure and dynamical function of GTSF1. Finally, site-specific experimental interrogation of GTSF1 allows confirmation of previous genetic and biochemical results. In this chapter, a structural study of the piRNA-pathway protein GTSF1 and the search for a specific RNA binding sequence is presented.

3.2 Optimization of GTSF1 U-[^{15}N] Sample for NMR Spectroscopy

From sequence-based secondary structure predictions and gene annotation, full-length GTSF1 is predicted to have two nucleic acid-binding Zn^+ finger domains. These two domains, which each comprise 27 residues, are the only secondary structure elements predicted; predictions for residues 1-13 and residues 76-167 are random coil. Thus, more than half the protein is disordered. The first attempt at recording ^{15}N HSQC for GTSF1.FL demonstrated just that. As seen in Figure 3.2.1, the overwhelming signal comes from the random coil region and no dispersed peaks appear. Furthermore, there exists a plateau of NMR signal in the random coil region, potentially indicative of anisotropic effects on the rotational correlation time of GTSF1. As a first step to resolve this, a series of truncated GTSF1 protein constructs was produced. Three protein constructs, truncated to residue 115, residue 125, and residue 135, were expressed and purified. Shorter constructs intended to isolate the Zn fingers, with truncations at residue 45 or residue 75, expressed very poorly in *E. coli* and were not pursued further by NMR study. Of the three tested constructs, ^{15}N HSQC spectra qualitatively looked very similar. Construct 1-115, referred to henceforth as GTSF1.115, was selected for further testing because of the ten less disordered residues and lower potential to obscure dispersed peaks. Also, of the three GTSF1 constructs tested, GTSF1.115 showed the highest protein yields (14 mg per liter of expression media) of protein expression and purification. With no precipitation visible in the NMR sample of GTSF1.115 and circular dichroism-measured indications of a folded protein yet still mediocre spectra, the initial buffer conditions (20 mM Tris, pH 8.0, and 200 mM NaCl) were suspected of requiring optimization.

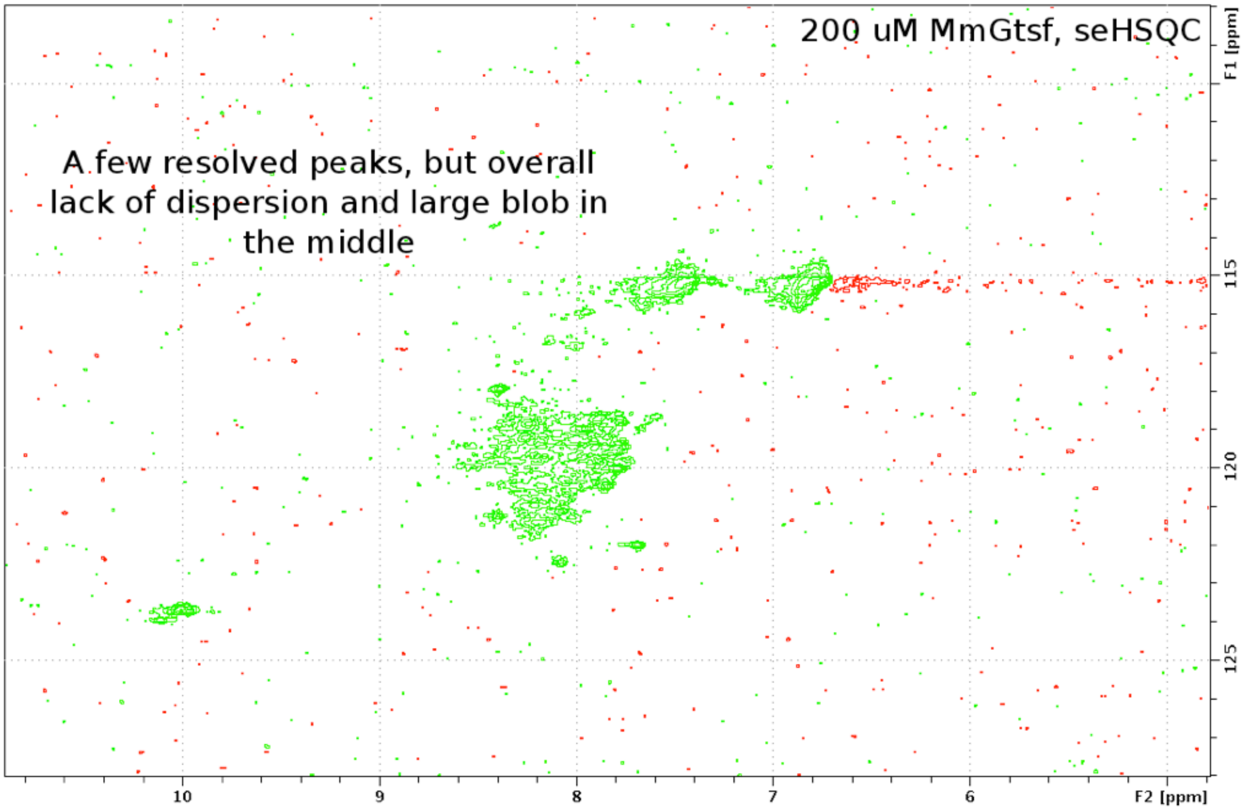
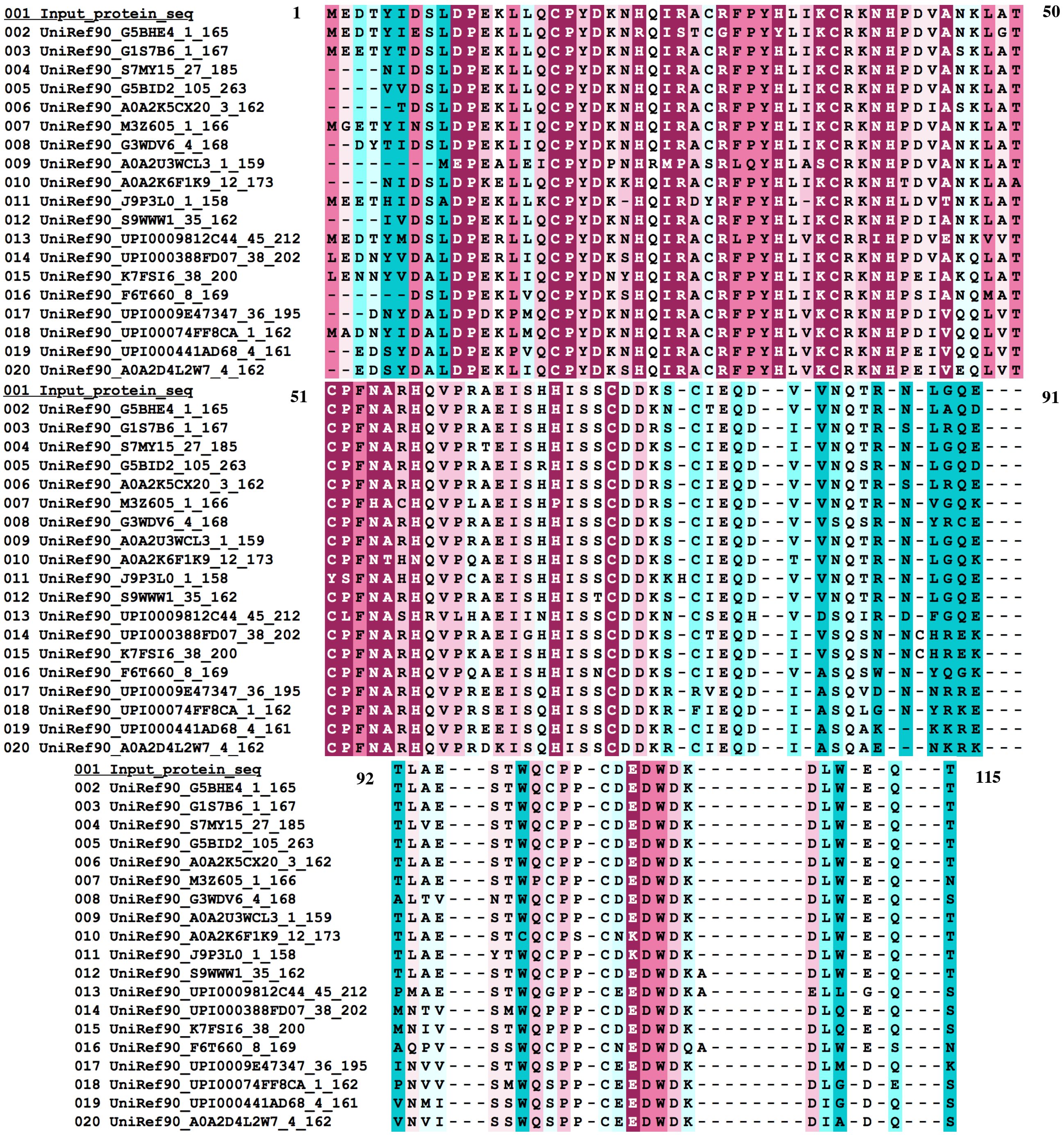


Figure 3.2.1: **^{15}N HSQC of Full-Length GTSF1.**
A ^{15}N HSQC experiment using the full-length construct of *Mm*GTSF1, which consists of 167 amino acids, showed i) no peak dispersion, ii) broad linewidths in the proton dimension, and iii) a blob of protein-derived signal presumably from the disordered C-terminal domain of GTSF1. These three issues immediately make the full-length construct unsuitable for NMR spectroscopy experiments.

To accomplish this, a modified version of the solvent screening method by Lepre and Moore was used. ^{15}N spectra were recorded using a 1.7 mm NMR tube, which only requires 33 μL of protein sample. This allowed testing of buffering agents between pH 8.0 and pH 5.5, and scoring according to peak dispersion and ratio of random coil dispersed peak intensities. Two more sample conditions were also varied: Zn^{+2} was adjusted, and TCEP was added at three concentrations. With its classification as a Zn^{+2} finger protein, the tendency of GTSF1.115 to bind metal ions was always suspected. Biophysical studies of Zn^{+2} finger proteins, though, typically rely on endogenously-bound zinc because of the historically high affinity Zn^{+2} finger proteins have for their metal cofactor [62, 63]. Since the protein still appeared only partially ordered once the pH was optimized, addition of zinc in 1:1 and 2:1 stoichiometric ratios of zinc:GTSF1.115 was attempted. Finally, the reducing agent tris(2-carboxyethyl)phosphine (TCEP) was added in three ratios to judge whether intermolecular associations mediated by cysteines affected the GTSF1.115 spectra.

The buffer pH screen was highly informative. The protein was observed to come out of solution at pH 6.0 and below. 2-(N-morpholino) ethanesulfonic acid (MES) buffer at pH 6.5 was judged as the optimal pH for GTSF1.115, with a combination of defined peaks in the random coil region and weak peaks appearing downfield (between 8.8 and 9.3 ppm) and upfield (between 7.5 and 6.5 ppm) in the spectrum. The average difference in peak intensities between random coil and dispersed peaks was only ~10^1. Titration of $ZnCl_2$ was even more successful. A ^{15}N HSQC spectrum of 1:1 Zn^{+2} to GTSF1 exhibited modest increases in intensity and sharpness in lineshape of dispersed peaks, whereas 2:1 Zn^{+2} to GTSF1 led to even more promising spectra. Further titrations were not attempted because additional spectral benefits were not observed and higher ratios of zinc to protein resulted in sample precipitation. Finally, addition of the reducing agent TCEP was observed to confer minimal effects on the ^{15}N HSQC. Regardless, 10 mM TCEP or d_{16}-TCEP (98%, Cambridge Isotope Laboratories) was included in all NMR samples. The final solvent optimized ^{15}N HSQC spectrum is visible in Figure 3.2.1.

Using the amide correlation spectrum as a diagnostic indicator, GTSF1.115 was judged to be a promising protein construct for further investigation, notably for NMR triple-resonance-based chemical shift assignment and structure calculations. However, initial triple-resonance assignment experiments (HNCACB, HN(CO)CACB, etc.) showed a lack of completeness in experimental crosspeaks. As an example, dispersed peaks in a HNCACB experiment lacked both C^{α} and C^{β} peaks for i^{th} residues and C^{α} and C^{β} residues for the $(i\text{-}1)^{th}$ residue. Similar results occurred with HNCA experiments. The remaining property of the GTSF1.115 protein construct that was still yet optimized were the four wild-type cysteine residues in the protein construct. As seen in the multiple sequence alignment of Figure 3.2.2, C28 and C76, which the two cysteines located in the region containing the Zn finger domains, were not strongly conserved across organisms and were not predicted as part of the $\mathbf{C}\text{-}X_3\text{-}\mathbf{H}\text{-}X_9\text{-}\mathbf{H}\text{-}X_3\text{-}\mathbf{C}$ motif. The hypothesis was that any or all of the four cysteines might be contributing to intermolecular associations between GTSF1.115 monomers, which can lead to problematic effects on magnetization transfer during the NMR pulse sequence and restrict the information content of the NMR experiment. Thus, constructs consisting of three (excluding C28) and four cysteine-to-serine mutations were used to express isotopically labeled GTSF1.115. Figure 3.2.4 shows ^{15}N HSQC spectra for both GTSF1.115.3CtoS and GTSF1.115.4CtoS. With the exception of four amide peaks, all other chemical shifts overlapped between the two mutant GTSF1.115 sequences. To mitigate any concerns over intermolecular associations, the final construct used for NMR studies was GTSF1.115.4CtoS in 50 mM MES (pH 6.5), 200 mM NaCl, 2:1 stochiometric addition of $ZnCl_2$, and 10 mM TCEP. A spectrum of the sequence- and solvent-optimized is shown in Figure 3.2.5.



```
001 Input_protein_seq            1  MEDTYIDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDVANKLAT  50
002 UniRef90_G5BHE4_1_165           MEDTYIESLDPEKLLQCPYDKNRQISTCGFPYYLIKCRKNHPDVANKLGT
003 UniRef90_G1S7B6_1_167           MEETYTDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDVASKLAT
004 UniRef90_S7MY15_27_185          ----NIDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDVANKLAT
005 UniRef90_G5BID2_105_263         ----VVDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDVANKLAT
006 UniRef90_A0A2K5CX20_3_162       -----TDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDIASKLAT
007 UniRef90_M3Z605_1_166           MGETYINSLDPEKLIQCPYDKNHQIRACRFPYHLIKCRKNHPDVANKLAT
008 UniRef90_G3WDV6_4_168           --DYTIDSLDPEKLIQCPYDKNHQIRACRFPYHLIKCRKNHPDVANKLAT
009 UniRef90_A0A2U3WCL3_1_159       --------MEPEALEICPYDPNHRMPASRLQYHLASCRKNHPDVANKLAT
010 UniRef90_A0A2K6F1K9_12_173      ----NIDSLDPKELLQCPYDKKHQIRACRFPYHLIKCRKNHTDVANKLAA
011 UniRef90_J9P3L0_1_158           MEETHIDSADPEKLLKCPYDK-HQIRDYRFPYHL-KCRKNHLDVTNKLAT
012 UniRef90_S9WWW1_35_162          ----IVDSLDPEKLLQCPYDKNHQIRACRFPYHLIKCRKNHPDIANKLAT
013 UniRef90_UPI0009812C44_45_212   MEDTYMDSLDPERLLQCPYDKNHQIRACRLPYHLVKCRRIHPDVENKVVT
014 UniRef90_UPI000388FD07_38_202   LEDNYVDALDPERLIQCPYDKSHQIRACRFPYHLIKCRKNHPDVAKQLAT
015 UniRef90_K7FSI6_38_200          LENNYVDALDPEKLIQCPYDNYHQIRACRFPYHLIKCRKNHPEIAKQLAT
016 UniRef90_F6T660_8_169           ------DSLDPEKLVQCPYDKSHQIRACRFPYHLIKCRKNHPSIANQMAT
017 UniRef90_UPI0009E47347_36_195   ---DNYDALDPDKPMQCPYDKNHQIRACRFPYHLVKCRKNHPDIVQQLVT
018 UniRef90_UPI00074FF8CA_1_162    MADNYIDALDPEKLMQCPYDKNHQIRACRFPYHLVKCRKNHPDIVQQLVT
019 UniRef90_UPI000441AD68_4_161    --EDSYDALDPEKPVQCPYDKNHQIRACRFPYHLVKCRKNHPEIVQQLVT
020 UniRef90_A0A2D4L2W7_4_162       --EDSYDALDPEKLIQCPYDKSHQIRACRFPYHLVKCRKNHPEIVEQLVT
001 Input_protein_seq           51  CPFNARHQVPRAEISHHISSCDDKS-CIEQD--V-VNQTR-N-LGQE---  91
002 UniRef90_G5BHE4_1_165           CPFNARHQVPRAEISHHISSCDDKN-CTEQD--V-VNQTR-N-LAQD---
003 UniRef90_G1S7B6_1_167           CPFNARHQVPRAEISHHISSCDDRS-CIEQD--I-VNQTR-S-LRQE---
004 UniRef90_S7MY15_27_185          CPFNARHQVPRTEISHHISSCDDKS-CIEQD--I-VNQTR-N-LGQE---
005 UniRef90_G5BID2_105_263         CPFNARHQVPRAEISRHISSCDDKS-CIEQD--V-VNQSR-N-LGQD---
006 UniRef90_A0A2K5CX20_3_162       CPFNARHQVPRAEISHHISSCDDRS-CIEQD--V-VNQTR-S-LRQE---
007 UniRef90_M3Z605_1_166           CPFHACHQVPLAEISHPISSCDDRS-CIEQD--V-VNQTR-N-VGQK---
008 UniRef90_G3WDV6_4_168           CPFNARHQVPRAEISHHISSCDDKS-CIEQD--V-VSQSR-N-YRCE---
009 UniRef90_A0A2U3WCL3_1_159       CPFNARHQVPRAEISHHISSCDDKS-CIEQD--V-VNQTR-N-LGQE---
010 UniRef90_A0A2K6F1K9_12_173      CPFNTHNQVPQAEISHHISSCDDKS-CIEQD--T-VNQTR-N-LGQK---
011 UniRef90_J9P3L0_1_158           YSFNAHHQVPCAEISHHISSCDDKKHCIEQD--V-VNQTR-N-LGQE---
012 UniRef90_S9WWW1_35_162          CPFNARHQVPRAEISHHISTCDDKS-CIEQD--V-VNQTR-N-LGQE---
013 UniRef90_UPI0009812C44_45_212   CLFNASHRVLHAEIINHISSCDDKN-CSEQH--V-DSQIR-D-FGQE---
014 UniRef90_UPI000388FD07_38_202   CPFNARHQVPRAEIGHHISSCDDKS-CTEQD--I-VSQSN-NCHREK---
015 UniRef90_K7FSI6_38_200          CPFNARHQVPKAEISHHISSCDDKS-CIEQD--I-VSQSN-NCHREK---
016 UniRef90_F6T660_8_169           CPFNARHQVPQAEISHHISNCDDKS-CIEQD--V-ASQSW-N-YQGK---
017 UniRef90_UPI0009E47347_36_195   CPFNARHQVPREEISQHISSCDDKR-RVEQD--I-ASQVD-N-NRRE---
018 UniRef90_UPI00074FF8CA_1_162    CPFNARHQVPRSEISQHISSCDDKR-FIEQD--I-ASQLG-N-YRKE---
019 UniRef90_UPI000441AD68_4_161    CPFNARHQVPREEISQHISSCDDKR-CIEQD--I-ASQAK---KKRE---
020 UniRef90_A0A2D4L2W7_4_162       CPFNARHQVPRDKISQHISSCDDKR-CIEQD--I-ASQAE---NKRK---
001 Input_protein_seq           92  TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T  115
002 UniRef90_G5BHE4_1_165           TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T
003 UniRef90_G1S7B6_1_167           TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T
004 UniRef90_S7MY15_27_185          TLVE---STWQCPP-CDEDWDK---------DLW-E-Q---T
005 UniRef90_G5BID2_105_263         TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T
006 UniRef90_A0A2K5CX20_3_162       TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T
007 UniRef90_M3Z605_1_166           TLAE---STWPCPP-CDEDWDK---------DLW-E-Q---N
008 UniRef90_G3WDV6_4_168           ALTV---NTWQCPP-CDEDWDK---------DLW-E-Q---S
009 UniRef90_A0A2U3WCL3_1_159       TLAE---STWQCPP-CDEDWDK---------DLW-E-Q---T
010 UniRef90_A0A2K6F1K9_12_173      TLAE---STCQCPS-CNKDWDK---------DLW-E-Q---T
011 UniRef90_J9P3L0_1_158           TLAE---YTWQCPP-CDKDWDK---------DLW-E-Q---T
012 UniRef90_S9WWW1_35_162          TLAE---STWQCPP-CDEDWDKA--------DLW-E-Q---T
013 UniRef90_UPI0009812C44_45_212   PMAE---STWQGPP-CEEDWDKA--------ELL-G-Q---S
014 UniRef90_UPI000388FD07_38_202   MNTV---SMWQPPP-CDEDWDK---------DLQ-E-Q---S
015 UniRef90_K7FSI6_38_200          MNIV---STWQPPP-CDEDWDK---------DLQ-E-Q---S
016 UniRef90_F6T660_8_169           AQPV---SSWQCPP-CNEDWDQA--------DLW-E-S---N
017 UniRef90_UPI0009E47347_36_195   INVV---STWQSPP-CEEDWDK---------DLM-D-Q---K
018 UniRef90_UPI00074FF8CA_1_162    PNVV---SMWQSPP-CEEDWDK---------DLG-D-E---S
019 UniRef90_UPI000441AD68_4_161    VNMI---SSWQSPP-CEEDWDK---------DIG-D-Q---S
020 UniRef90_A0A2D4L2W7_4_162       VNVI---SSWQSPP-CEEDWDK---------DIA-D-Q---S
```

Figure 3.2.2: **Sequence Alignment of GTSF1 Across Various Organisms.** The top 20 hits from a multiple sequence alignment (MSA) of 150 homologues of wild-type *Mm*GTSF1 (1-115). Conservation runs the spectrum from dark red-purple (highly conserved) to aqua-green (highly variable). The region encompassing the Zn fingers, specifically D10-D73, possessed a sustained stretch of highly conserved residues. These included Zn #1 ligands C17, H23, H33, and C37, and Zn #2 ligands C51, H57, H67, and C71. Note wild-type residues C28, C76, C100, and C103 are transparent or colored in light green, indicating an average or moderately variable character across MSA. This supported mutation of these non-essential cysteines to prevent problematic inter-molecular associations.

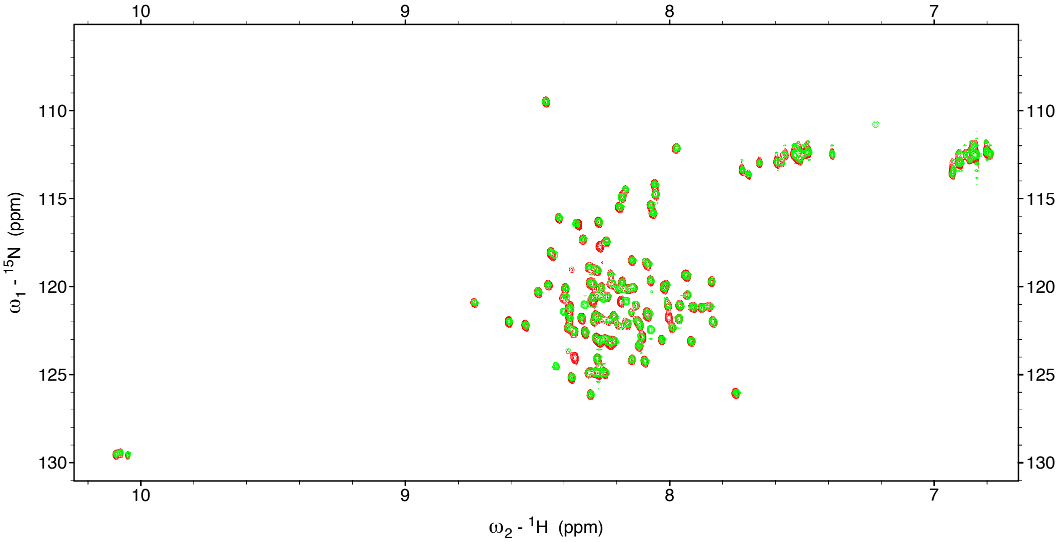


Figure 3.2.3: **^{15}N HSQC of GTSF1.115.3CtoS (green) and GTSF1.115.4CtoS (red).** Differences between spectra of GTSF1.115.3CtoS and GTSF1.115.4CtoS were minimal, with only small chemical shift perturbations observed in the apo-GTSF1 state (no $ZnCl_2$, no TCEP). It was decided to pursue further NMR experiments using the GTSF1.115.4CtoS construct to minimize the effects of non-Zn finger cysteine-mediated self-association.

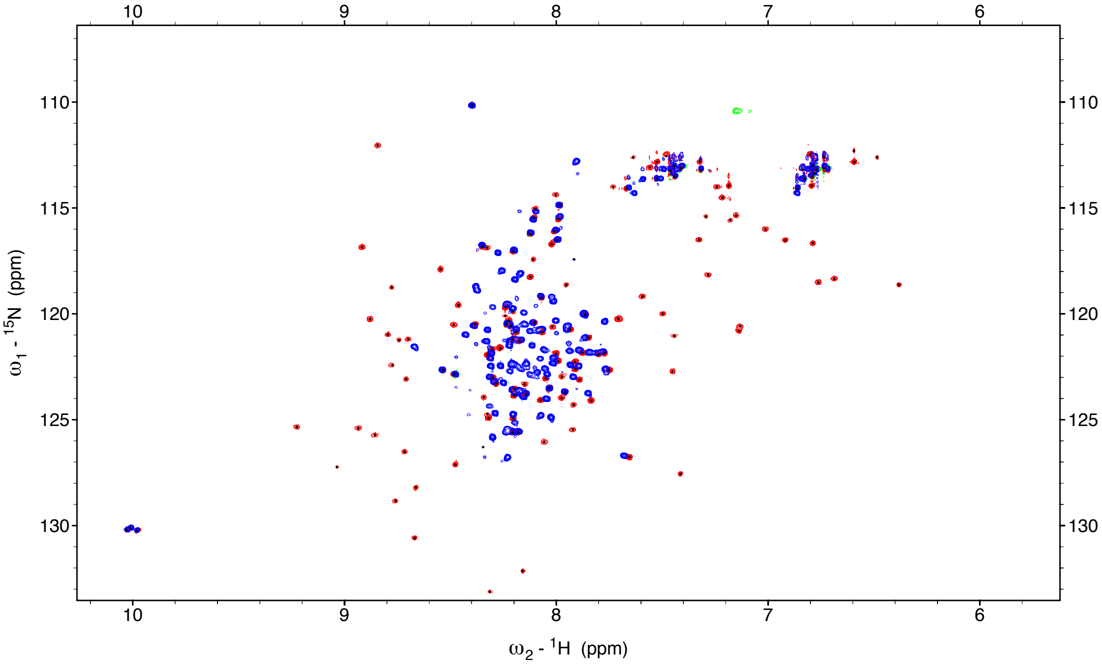


Figure 3.2.4: **^{15}N HSQC of Sequence- and Solvent-Optimized GTSF1 (GTSF1.115.4CtoS).** Apo-GTSF1.115.4CtoS (blue) corresponds to the protein in NMR buffer at pH6.5, while the dispersed holo-GTSF1.115.4CtoS (red) corresponds to NMR buffer plus 2:1 stoichiometric $ZnCl_2$ and 10mM TCEP. All further NMR experiments were recorded under these buffer conditions, and holo-GTSF1.115.4CtoS will be referred to as GTSF1 going forward.

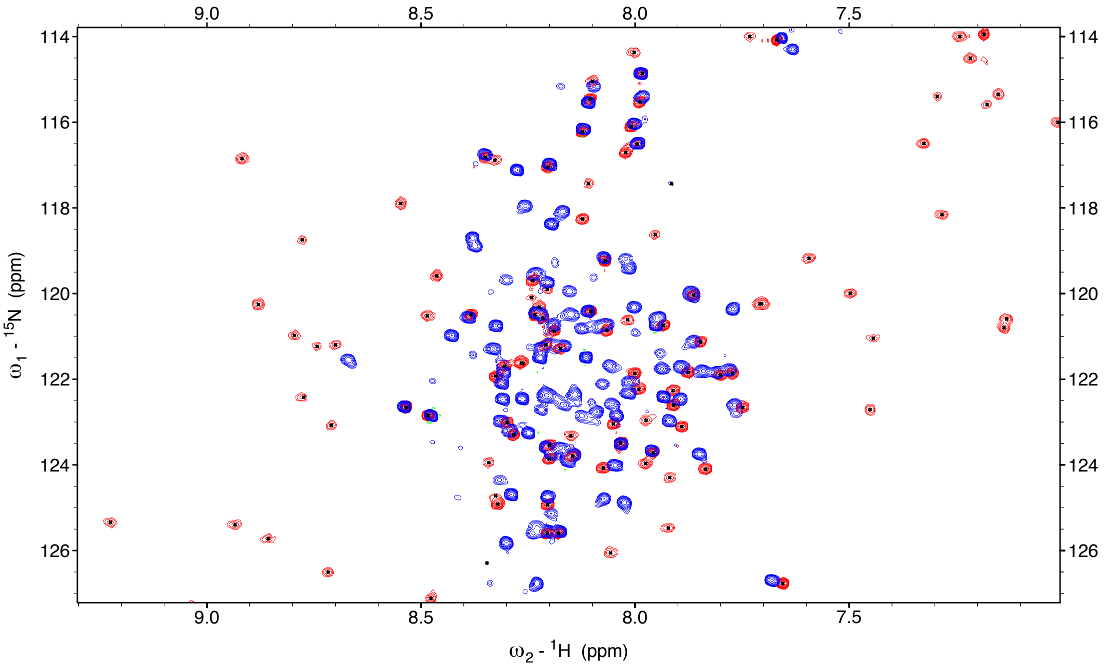


Figure 3.2.5: **Random Coil Region of GTSF1.115.4CtoS ^{15}N HSQC.** Addition of $ZnCl_2$ and TCEP promoted folding of apo-GTSF1.115.4CtoS (blue), which both lead to appearance of dispersed peaks and loss of random coil (8.3 ppm) peaks in GTSF1.115.4CtoS (red). However, peaks that will later be assigned to the C-terminal portion of GTSF1.115.4CtoS show minimal perturbation.

3.3 Towards an NMR Structure: GTSF1 Chemical Shift Assignment

3.3.1 Backbone Chemical Shift Assignment Experiments

NMR triple-resonance assignment experiments were used to assign backbone C^{α}, C^{β}, CO, H^{N}, and N chemical shifts. Initial analysis of assignment experiments proved difficult as purified protein samples showed microscopic degradation with no corresponding macroscopic precipitation. These effects manifested as the appearance of sharp peaks at random coil ^{1}H frequencies and downfield ^{15}N frequencies in the ^{15}N HSQC, as well as an increased number of dispersed peaks. More thorough purification and addition of a wider breadth of protease inhibitors, an increase in the anti-microbial sodium azide, and more timely experimental use (stability on the order of one-two months) of the NMR sample allowed for reproducible ^{15}N HSQC and three-dimensional spectra.

The final set of NMR experiments used for assignment of backbone chemical shifts and assignment of the ^{15}N HSQC spectrum consisted of the following: HNCA, HN(CO)CA, HNCO, HNCACB, CBCA(CO)NH, and CBCANH. The assigned ^{15}N HSQC spectrum of GTSF1 is shown in Figure 3.2.6. Of the 115 GTSF1 residues in this protein construct, all but E2 are assigned. Note that some peak overlap is visible (e.g. Q16 and Q79, etc.) in the ^{15}N HSQC spectrum. In triple-resonance experiments, two (or one, depending on the type of experiment) sets of peaks are visible for the third dimension (carbon) dimension, thus allow proper peak assignment. However, overlap in the ^{1}H-^{15}N dimensions occurs in only a handful of cases.

By itself, the assigned ^{15}N HSQC spectrum of GTSF1 offers interesting clues regarding GTSF1 behavior. First, the dispersed peaks (upfield and downfield of 8.3 ppm in ^{1}H) are almost exclusively assigned to either residues located in the two predicted Zn finger domains or the short ten residue fragment linking the two Zn fingers. In addition to the resultant dispersed chemical shifts, which implicates these residues as populating elements of secondary structure, the intensities of these dispersed peaks are smaller by a factor of 5-6 compared to intense peaks (residues G1-D10 and S76-T115).

In summary, triple-resonance backbone experiments permit near complete chemical shift assignment of ^{1}H-^{15}N correlations, C^{α}, C^{β}, and CO. For carbonyls shifts, assignment coverage was less than other nuclei types (only 90.4% for CO vs. 97% for C^{α}/C^{β}/H^{N}/N) because residues preceding the eight prolines lacked assignments. An HN(CA)CO experiment gave extremely low signal-to-noise, and thus was not able to remedy the missing assignments. Using the backbone chemical shifts, both C$^{\alpha}$ and CO secondary chemical shifts [64] indicate secondary structure for Zn finger 1 and 2, which comprise residues C17-C37 and C51-C71, respectively. Figure 3.2.7 shows these results.

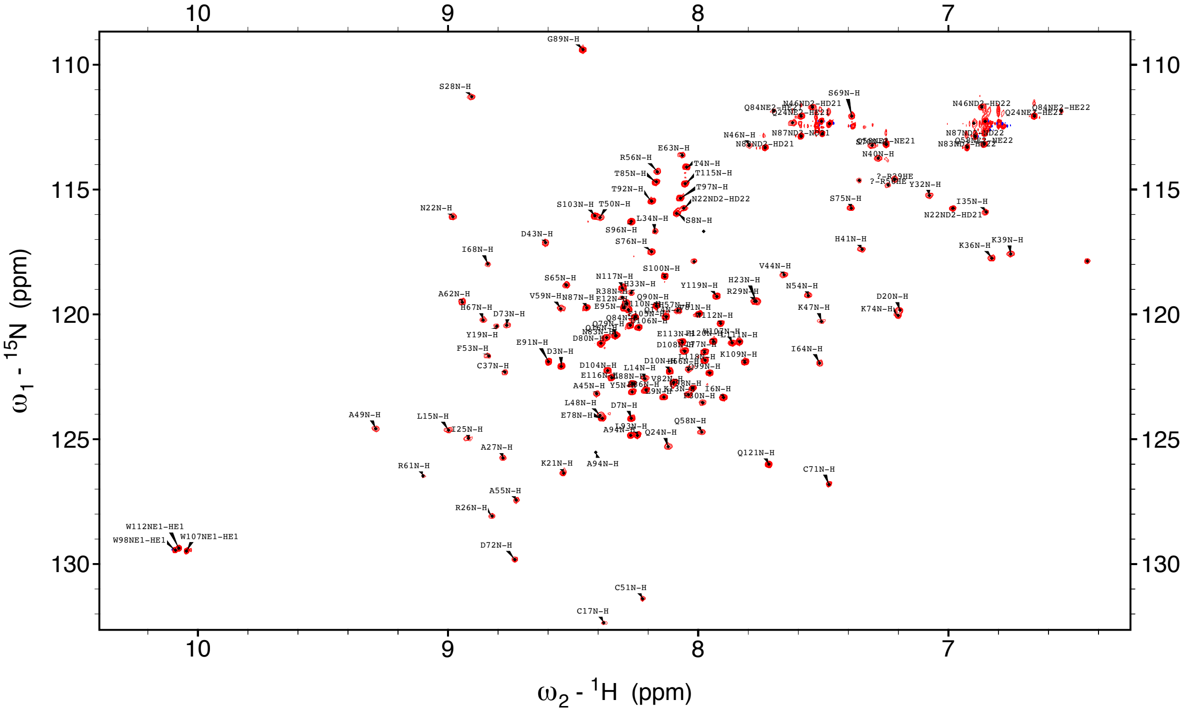


Figure 3.3.1: **^{15}N HSQC Showing Complete Amide Assignments for GTSF1.** Analysis of triple-resonance backbone assignment experiments resulted in complete assignment of ^{1}H-^{15}N correlations in a ^{15}N HSQC. The three-dimensional spectra acquired from the triple-resonance pulse sequences also permitted assignment of ~94% of backbone C$^{\alpha}$, C$^{\beta}$, CO chemical shifts.

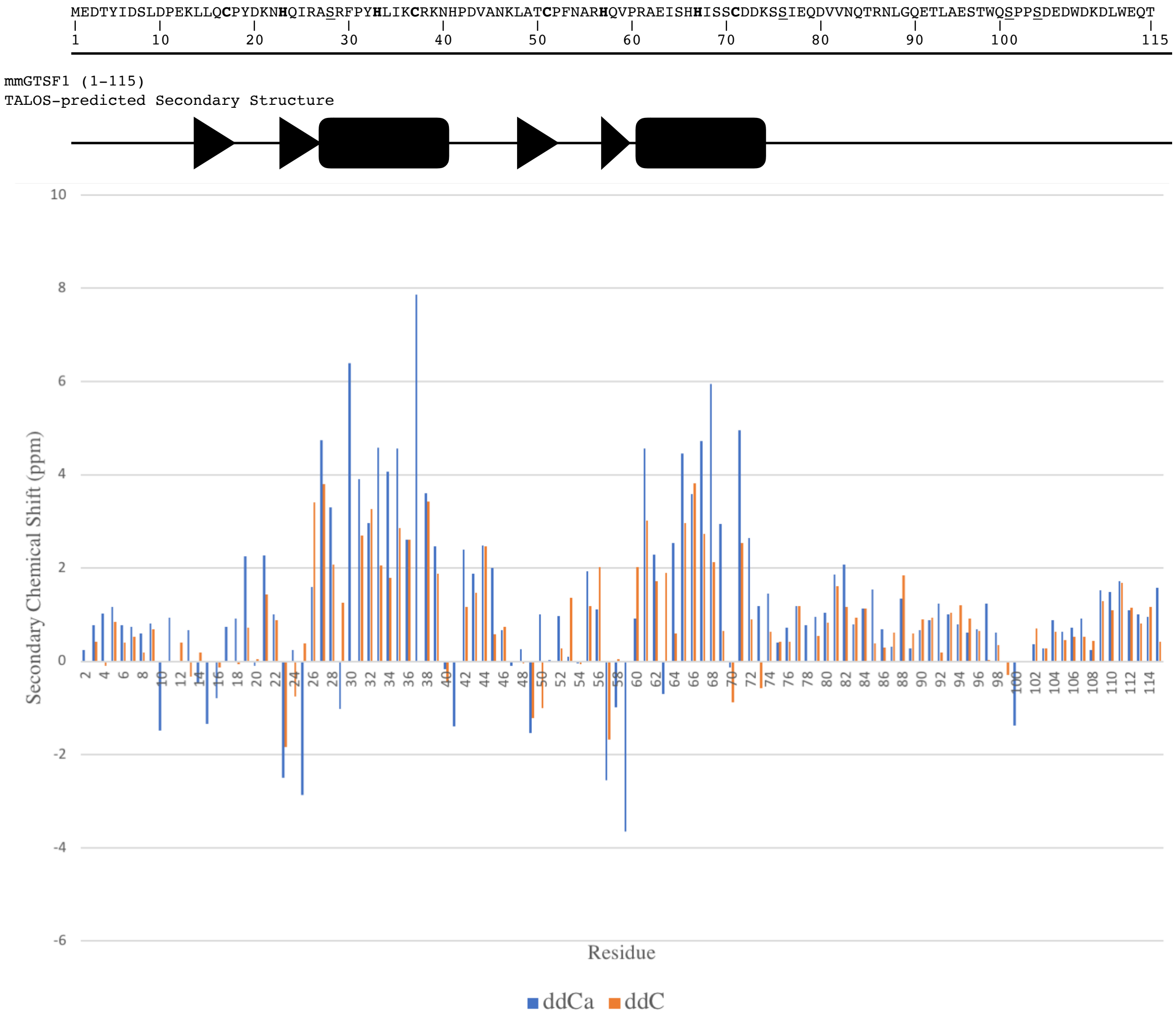


Figure 3.3.2: **C^{α} and CO Secondary Chemical Shifts for GTSF1.**
Secondary chemical shifts, which are computed by subtracting the residue-specific random coil shift from the observed chemical shift, are used for the identification of protein secondary structure. Positive secondary shifts are indicative of protein α-helical character, while negative secondary shifts indicate the presence of a β-sheet. For GTSF1, secondary shifts predict a similar secondary structure motif for both Zn fingers: two small β-sheets precede a longer (between eight and ten residue) α-helix. Additionally, the linker between Zn fingers is predicted to contain a four-residue α-helix (P42-A45). Finally, two regions in the C-terminal portion of GTSF1 (V80-V81 and K109-L111) show modestly strong (~2 ppm or slightly lower) positive secondary shifts.

3.3.2 Side-chain Chemical Shift Assignment Experiments

While an HCCH-TOCSY experiment was recorded, final side-chain ^1H and ^{13}C chemical shifts were assigned using a combination H(CCO)NH and CC(CO)NH experimental approach. This approach took advantage of the comparatively better ^{15}N resolution of GTSF1 spectra compared to the peak overlap present in the carbon dimension of the HCCH-TOCSY experiment. Furthermore, intensity differences between carbon correlations of dispersed peaks and those of random coil peaks consistently led to the former being obscured from confident assignment in the HCCH-TOCSY experiment. Both H(CCO)NH and CC(CO)NH experiments involve magnetization that originates on aliphatic protons and is mixed among protons or carbon, respectively, with a carbon isotropic mixing sequence. t_1 labeling is followed by polarization transfer to the carbonyl of the sequential residue and chemical shift evolution on the amide nitrogen and proton nuclei. The resulting three-dimensional spectrum allows similar analysis to that of backbone triple-resonance assignment spectra: moving along nitrogen slices, strips plots at the *i*th amide proton frequency give all the (*i*-1)th proton or carbon resonances. This assignment strategy for side-chain ^1H and ^{13}C chemical shifts gave good assignment coverage, especially for residues D10-D80. The CC(CO)NH experiment serves as an excellent cross-validation for backbone chemical shift assignment because of the unique side-chain ^{13}C footprint for each of the 20 amino acids. The one issue that emerged from GTSF1 assignment with these experiments was overlapping (degenerate) $H^{\delta1/\delta2}$, $H^{\gamma1/\gamma2}$, and $H^{\varepsilon1/\varepsilon2}$ peaks. Even with high digitization in the indirect ^1H dimension, overlap was still an issue. This is not regarded as a hindrance for Zn finger peaks given that other side-chain chemical shifts, especially ^1H-^{13}C methyl correlations, could all be assigned. An example is shown in Figure 3.2.8.

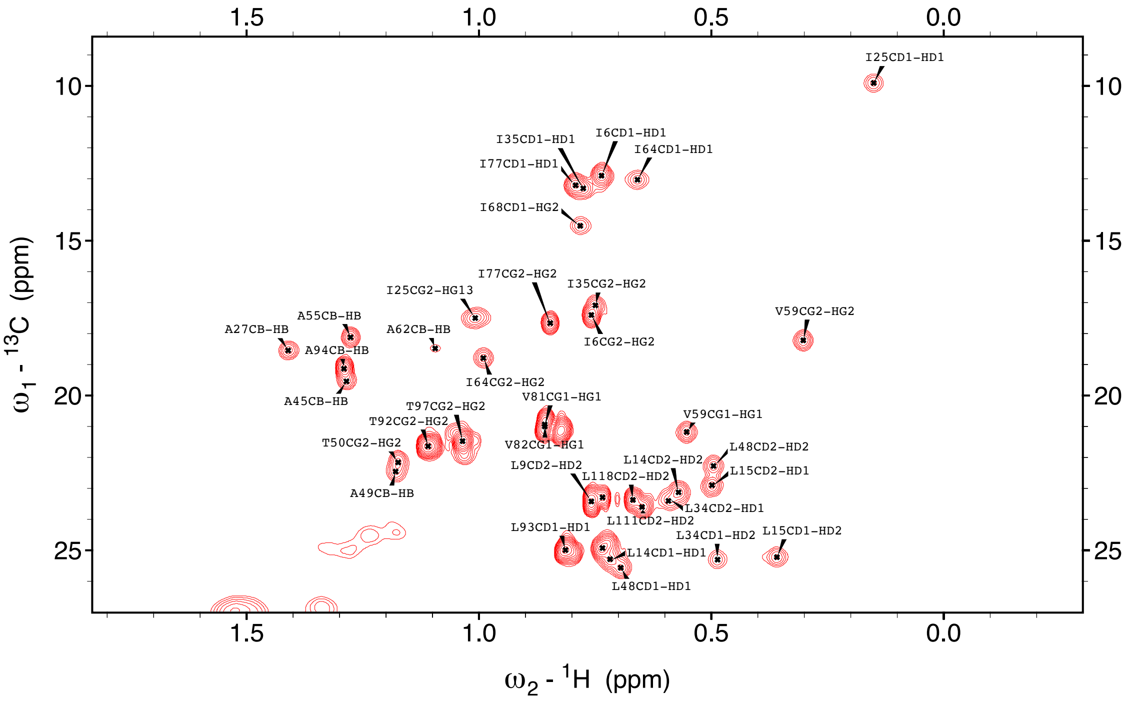


Figure 3.3.3: **Assigned ^{1}H-^{13}C SOFAST-HMQC Experiment of GTSF1.**
High resolution ^{13}C SOFAST-HMQC demonstrated dispersion for AILV ^{1}H-^{13}C methyl correlations. The methyl peaks originating from Zn finger AILV residues were all assigned, while peak overlap/degeneracy did occur for ^{1}H-^{13}C correlations in the disordered regions of GTSF1.

3.3.3 Aromatic Chemical Shift Assignment Experiments

Assignment of GTSF1 aromatic chemical shifts proved especially difficult. Non-constant and constant time ^{13}C HSQC spectra of GTSF1.115.4CtoS were dominated by high intensity aromatic ^{1}H-^{13}C correlations originating from random coil residues. Some aromatic peaks from the folded region were assigned, but the assignment of others was simply not possible (see Assignment Table). However, histidine aromatic H$^{\delta2}$-C$^{\delta2}$ and H$^{\varepsilon1}$-C$^{\varepsilon1}$ peaks were distinct. Because of the large chemical shift anisotropy of GTSF1 aromatic carbons, a TROSY-based pulse sequence for correlation of aromatic proton-carbon nuclei was recorded and yielded noticeably better spectra for the desired resonances (Figure 3.2.8).

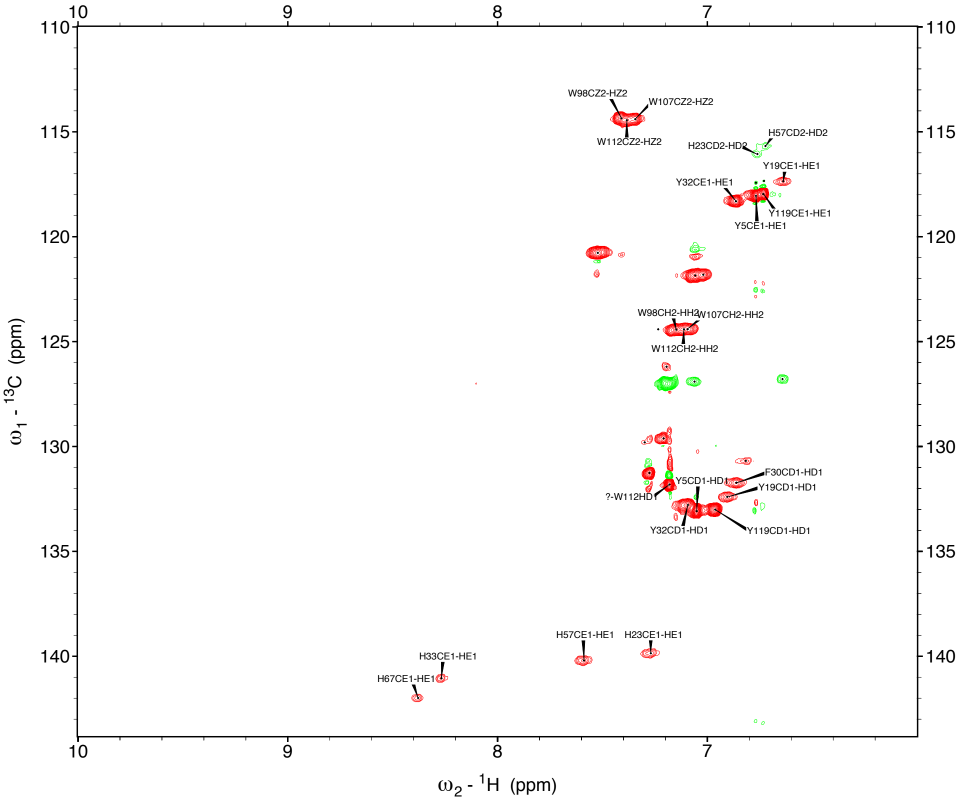


Figure 3.3.4: **TROSY constant-time ^{13}C HSQC for GTSF1.**
A ^{1}H-^{13}C spectrum that used TROSY coherence transfer that selected for the aromatic spectral region allowed observation of stronger peaks originating from the Zn finger aromatic residues. Most importantly, the $H^{\varepsilon}/C^{\varepsilon}$ and H^{δ}/C^{δ} peaks from histidine are visible for assignment. Even in a non-deuterated background, the ^{13}C chemical shift anisotropy (CSA) relaxation of aromatic carbons is large enough to serve as a counter to the dipole-dipole relaxation mechanism, giving a TROSY-like effect without the normally required isotopic labeling. In this experiment, peaks corresponding to signals modulated by odd *n*-multiples of active carbon couplings appear negative, while peaks corresponding to signals modulated by even *n*-multiples of active carbon couplings appear positive.

Elucidation of an accurate GTSF1 structure requires proper characterization of Zn ligands. Like other Zn finger proteins, the metal-coordinating elements are contributed by side-chain or aromatic elements. For the four GTSF1 Zn finger cysteine residues (C17 and C37 in Zn finger 1, and C51 and C71 in Zn finger 2), the lone pair of the side chain S^{γ} atom is assumed to ligate the Zn. Cysteine carbon chemical shifts back up this assumption: C17 and C51 showed modest C^{α} and C^{β} perturbations, while C37 and C71 give notable downfield C^{α} and upfield C^{β} shifts, which are characteristic of cysteine S^{γ} coordination (Table 3.2.1). On the other hand, the four histidine residues (H23 and H33 in Zn finger 1, and H57 and H66 in Zn finger 2) required further investigation to i) confirm Zn binding signature and ii) assign the correct tautomer for each histidine. Typically, ^{15}N HMQC experiment is recorded to correlate aromatic $H^{\varepsilon1}$ and $H^{\delta2}$ proton nuclei with $N^{\delta1}$ and $N^{\varepsilon2}$ nitrogen nuclei on the histidine imidazole side-chain. A ^{15}N HSQC-type experiment can't be recorded in this circumstance because of the potential lack of protonation on one of the two histidine nitrogen atoms as well as the effects of proton exchange on peak intensity. Therefore, a ^{15}N HMQC experiment is used that relies on the two-bond coupling (^{2}J = 20 Hz) between aromatic proton and nitrogen to transfer magnetization, t_1 label the nitrogen nuclei, and transfer back to aromatic proton for detection [4]. Recording a presaturation ^{15}N HMQC with a U-[^{13}C, ^{15}N] GTSF1.115.4CtoS sample in 99% D_2O, which is necessary to maximize water suppression, gave no signal. This prompted the search for an optimized pulse sequence capable of detecting the two-bond aromatic proton-nitrogen correlation in a double isotope labeled, deuterated solvent background (99% D_2O). Figure 3.5.X shows the results of this search. The final pulse sequence is similar to the presaturated- ^{15}N HMQC in the type of water suppression used, HMQC-like evolution, phase cycle, etc. However, instead of just converting single quantum ^{1}H into transverse double quantum ^{1}H^{15}N magnetization during the pulse sequence, double quantum ^{1}H^{13}C is created before evolution under the two-bond scalar coupling to a multi-quantum coherence, t_1 labeling of ^{15}N, and back to the aromatic ^{1}H for detection. Compared to the presat-^{15}N HMQC, this pulse sequence mitigated ^{13}C relaxation effects present in a dynamic macromolecular system with U-[^{13}C, ^{15}N] labeling on the histidine imidazole side-chain. The presat-^{15}N HMQC is typically recorded on U-[^{15}N]-only NMR samples; comparisons of this optimized ^{15}N HMQC with previous standard pulse sequences for detection of histidine aromatic ^{2}J proton-nitrogen correlations are still ongoing to judge the achievable sensitivity and resolution improvements of the new method.

Cysteine Residue	**C$^{\alpha}$ (ppm)**	**C$^{\beta}$ (ppm)**
C17	57.5	33.58
C37	64.63	28.92
C51	56.78	33.25
C71	61.71	28.22

Table 3.2.1: **Measurements of Cysteine C$^{\alpha}$ and C$^{\beta}$ Chemical Shifts Indicate S$^{\gamma}$ Zn Coordination.**

The newly developed pulse sequence showed all four histidine H$^{\varepsilon1}$ chemical shifts (Figure 3.2.9). At each aromatic proton position, there existed a pair of peaks: each H$^{\varepsilon1}$ proton is correlated to a N$^{\delta1}$ and a N$^{\varepsilon2}$, giving two separate peaks in the vertical dimension. At physiological pH, histidine side-chains normally exist in the neutral tautomer. This consists of either a protonated N$^{\delta1}$ and deprotonated N$^{\varepsilon2}$ (N$^{\delta1}$-H tautomer) or vice-versa (N$^{\varepsilon2}$-H tautomer). Protonated ^{15}N nuclei resonate at approximately 170 ppm, while lone pair ^{15}N nuclei resonate at approximately 250 ppm. For GTSF1 Zn finger histidine residues, four lone pairs and four protonated 15N nuclei are observed at the four H$^{\varepsilon1}$ resonances. Only one set of peaks appeared in the newly developed pulse sequence that corresponds to $H^{\delta2}$ resonances for H33 and H67, which indicated that both of these histidine residues coordinated the Zn ion through $N^{\delta1}$. Therefore, though our optimized ^{15}N HMQC experiment confirmed the signature of ligating-competent histidine aromatic side-chains and the coordination modes for H33 and H67, a determination of coordination for histidine side-chain nitrogens of H23 and H57 was still necessary.

To accomplish unambiguous determination of the lone pair histidine nitrogen atoms for GTSF1, an empirical correlation derived from analysis of a database of Zn finger ^{13}C chemical shifts of aromatic carbons was used. Barraurd, et al. observe that the $^{13}C^{\delta2}$ aromatic chemical shift is uniquely sensitive to the histidine coordination mode and tautomeric structure [65]. Furthermore, taking the difference of the $^{13}C^{\varepsilon1}$ and $^{13}C^{\delta2}$ aromatic chemical shifts of histidine allows a referencing-independent calculation to be applied since the $^{13}C^{\varepsilon1}$ chemical shift is relatively insensitive to coordination state. Use of the empirical correlation for determination of GTSF1 Zn finger histidine tautomeric state is shown in Figure 3.2.10. The cut-off for the chemical shift difference $^{13}C^{\varepsilon1}$ - $^{13}C^{\delta2}$ is 17 ppm. A difference greater than this indicates Zn coordination through histidine $N^{\delta1}$, while a chemical shift difference less than 17 ppm implies histidine $N^{\varepsilon2}$ coordination of Zn. Using the aromatic carbon chemical shifts of GTSF1, H23 and H57 are designated with $N^{\delta1}$ coordination (Table 3.2.2). As mentioned earlier, these qualitative findings serve a role as structural restraints on the GTSF1-Zn interaction.

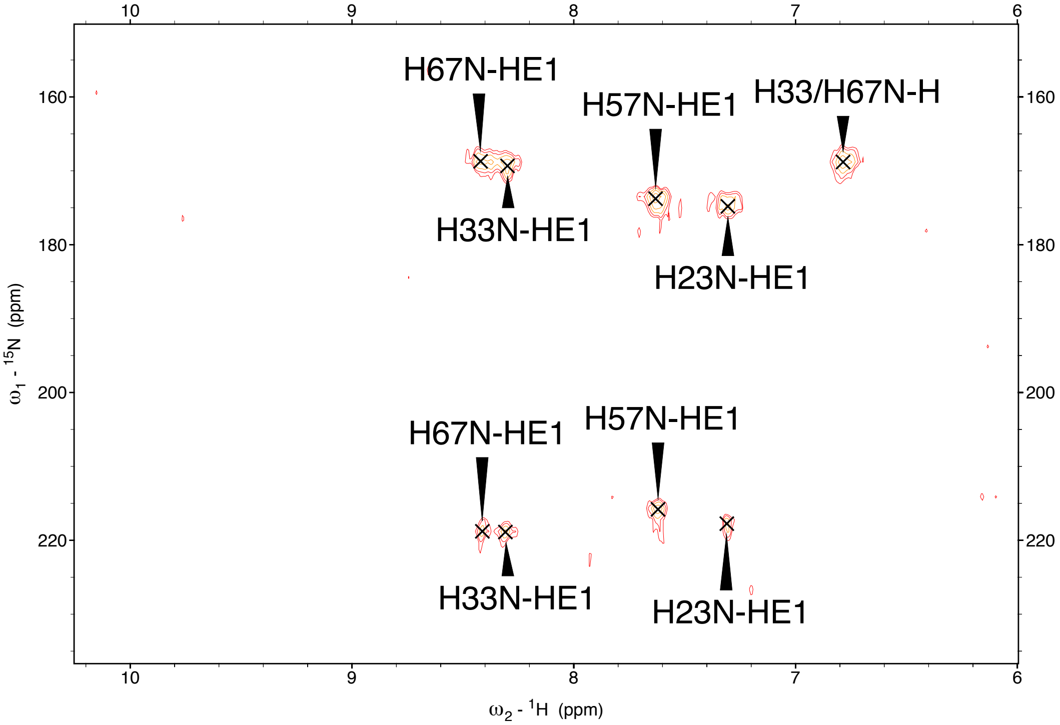


Figure 3.3.5: **A ^{15}N ^{2}*J* HMQC Experiment with Modified Handling of ^{13}C Spins.** Two-bond scalar couplings between aromatic protons ($H^{\varepsilon1}$ or $H^{\delta2}$) and imidazole nitrogen nuclei are typically used to identify tautomeric form of histidine. These same NMR experiments recorded using a GTSF1 sample gave no signal. Instead, a novel pulse sequence that evolved ^{15}N and ^{13}C transverse magnetization during the HMQC pulse sequence block gave peaks at each GTSF1 histidine $H^{\varepsilon1}$ proton. Unfortunately, only $H^{\delta2}$ peaks for H33 and H67 were observed in the spectrum. Since there was only one peak at the $H^{\delta2}$ slice, it is deduced that H33 and H67 both are $N^{\delta1}$ coordinated to Zn.

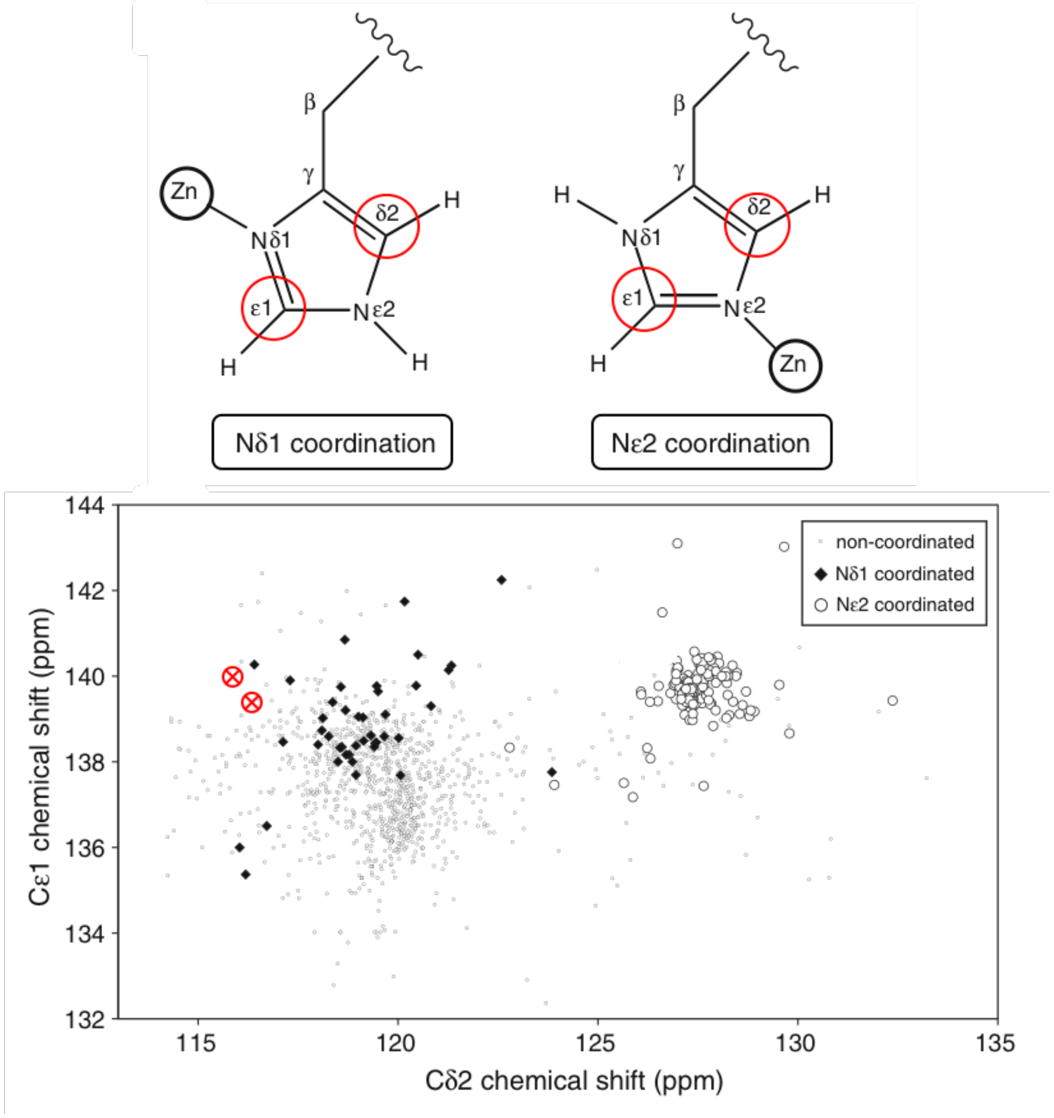


Figure 3.3.6: **Tautomer State of GTSF1 Identified by Database-Derived Empirical Correlation of Aromatic Carbon Chemical Shifts and Nitrogen Coordination.** Two types of neutral histidine tautomer exist: either histidine $N^{\delta 1}$ can coordinate Zn while $N^{\varepsilon 2}$ is protonated or $N^{\varepsilon 2}$ can coordinate Zn while $N^{\delta 1}$ is protonated. A past study correlated the aromatic ^{13}C chemical shifts with $N^{\delta 1}$/ $N^{\varepsilon 2}$ for Zn finger proteins with confirmed Zn binding. It was discovered that $\Delta(^{13}C^{\varepsilon 1} - {}^{13}C^{\delta 2})$ is a diagnostic for histidine Zn ligation. A difference in aromatic ^{13}C chemical shifts greater than 17 ppm predicts $N^{\delta 1}$-coordinated Zn. If the difference is instead less than 17 ppm, $N^{\varepsilon 2}$-coordination is predicted. The *xy*-plot above is taken from the paper, and shows the groups of different Zn coordination modes with GTSF1 H23, H33, H57, H67 positions indicated with red crosshairs. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Journal of Biomolecular NMR, 53, Barraud, P., Schubert, M., Allain, F. A strong 13C chemical shift signature provides the coordination mode of histidines in Zn-binding proteins. 93-101, Copyright 2012. [65]

Histidine Residue	**$\Delta(C^{\varepsilon1}\text{-}C^{\delta2})$ (ppm)**	**Nitrogen Coordination**
H23	23.7	$N^{\delta1}$
H57	24.3	$N^{\delta1}$

Table 3.2.2: **Histidine Tautomer Identified by Database-Derived Empirical Correlation of Aromatic Carbon Chemical Shifts and Tautomer State.**

3.3.4 NMR Spin Relaxation Experiments

Finally, ^{15}N backbone spin relaxation experiments were recorded using a U-[^{15}N] GTSF1 sample. The suite of experiments permitted measurement of the ^{15}N R_1 and R_2 relaxation rates, as well as the {^{1}H}-^{15}N heteronuclear NOE. Analysis of the heteronuclear NOE data indicated the formation of secondary structure within the dual Zn finger region of GTSF1. As viewed in Figure 3.2.11, an increase in the hetNOE is visible starting with E12 and plateaus at ~0.7 starting with L15. This value of hetNOE is generally regarded as indicative of secondary structure formation, and thus this measurement is expected for both Zn finger regions of GTSF1. A small decrease in hetNOE occurred in the seven-residue linker region between Zn fingers, which comprises H41 to K47. Finally, the hetNOE values begin to decrease starting with S75, which corresponds to the C-terminal end of Zn finger 2's α-helix. As expected for protein regions that are unstructured and considered random coil, residues C-terminal of D80 gave hetNOEs approaching zero or even slightly negative values. However, the stretch of residues between W107 and W112 exhibited hetNOEs ~0.30. This implies the presence of either the formation of nascent secondary structure order or dynamically-averaged secondary structure. {^{1}H}-^{15}N heteronuclear NOE measurements will be revisited later in the context of the structural model of GTSF1.

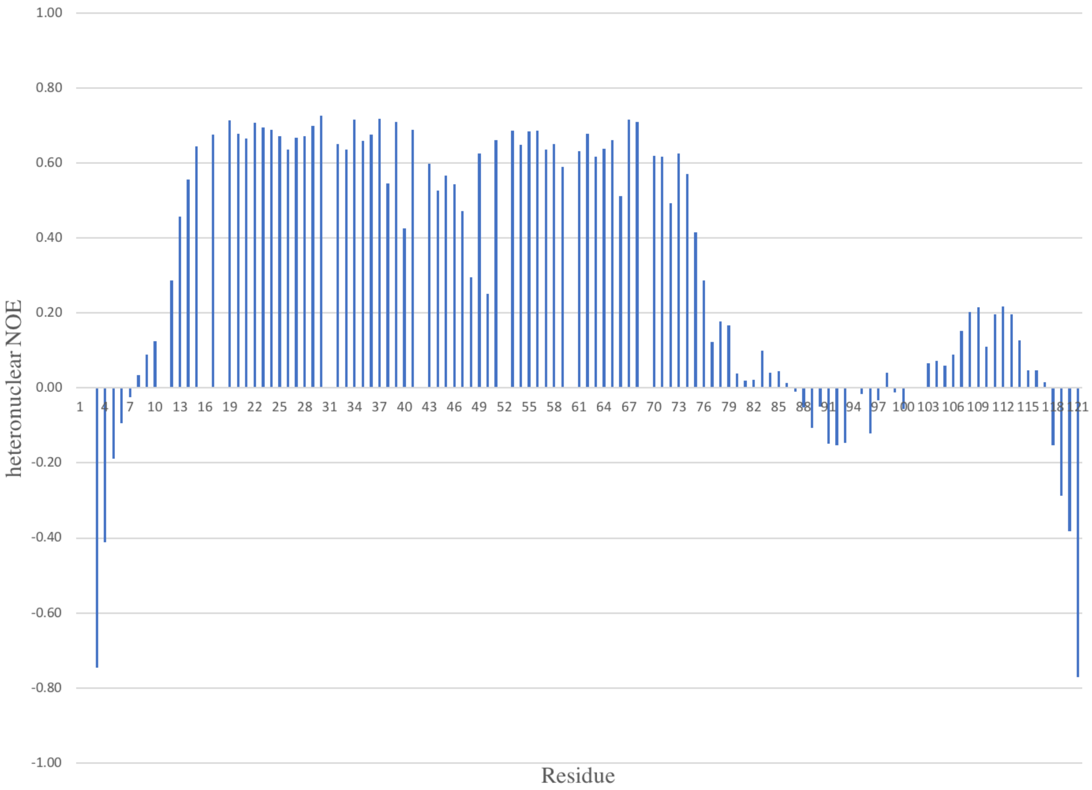


Figure 3.3.7: **{^{1}H}-^{15}N Heteronuclear NOE Measurements of GTSF1 Show Secondary Structure Formation in the Predicted Zn finger Regions.**

3.3.5 Discussion

Analysis of NMR experiments to assign backbone chemical shifts resulted in almost complete assignment of backbone chemical shifts (H^{N}, N, C^{α}, C^{β}, and CO) for residues 1-115 of GTSF1. Coverage of side-chain chemical shifts, which include aliphatic ^{1}H and ^{13}C, was slightly less than that of backbone shifts because of peak degeneracy, but the majority of nuclei were assigned in the region covering the two Zn fingers. Finally, only a subset of aromatic ^{1}H and ^{13}C assignments were made. This is attributed to the weak peak intensities stemming from relaxation effects, and the spectral overlap due aromatic correlations from non-structured residues (sharp peaks). Regardless, the combination of secondary chemical shifts, {^{1}H}-^{15}N hetNOE, and peak dispersion confirmed that the Zn finger regions were folded. Therefore, structure calculations using NOE cross-peak restraints was pursued for just the M1-D80 domain.

3.4 GTSF1 Structure Calculations Yield Two Defined Zn Finger Domains Connected by a Small Helix

3.4.1 ARIA Structure Calculations

Three NOESY-HSQC experiments were recorded to measure amide-, aliphatic-, and aromatic-mediated NOE cross-peaks. To guarantee the highest digital resolution possible in t_1, all three NOESY-HSQC experiments were run with 8 or 16 scans. Unfortunately, this compromised the signal-to-noise for certain peaks. Regardless, the Table gives the number of NOE cross-peaks extracted from each experiment as restraints for structure generation of the M1-D80 domain.

Restraint Type	#
^{1}H-^{1}H (^{15}N)	875
^{1}H-^{1}H (aliphatic ^{13}C)	1155
^{1}H-^{1}H (aromatic ^{13}C)	112
TALOS dihedrals	55
$^{3}J_{HNHA}$	20
Zn-coordination	8

Table 3.4.1: **Restraints Used for Structure Calculations.**

GTSF1 backbone chemical shifts were used to derive dihedral restraints through use of TALOS-N [66]. In total, 55 restraints for residues between M1 and D80 were used in structure calculations. TALOS-N predicted secondary structure elements are aligned along with the GTSF1 primary sequence and {^{1}H}-^{15}N heteronuclear NOE values in Figure 3.4.1. α-helix and β-sheet regions match with other indications of protein secondary structure.

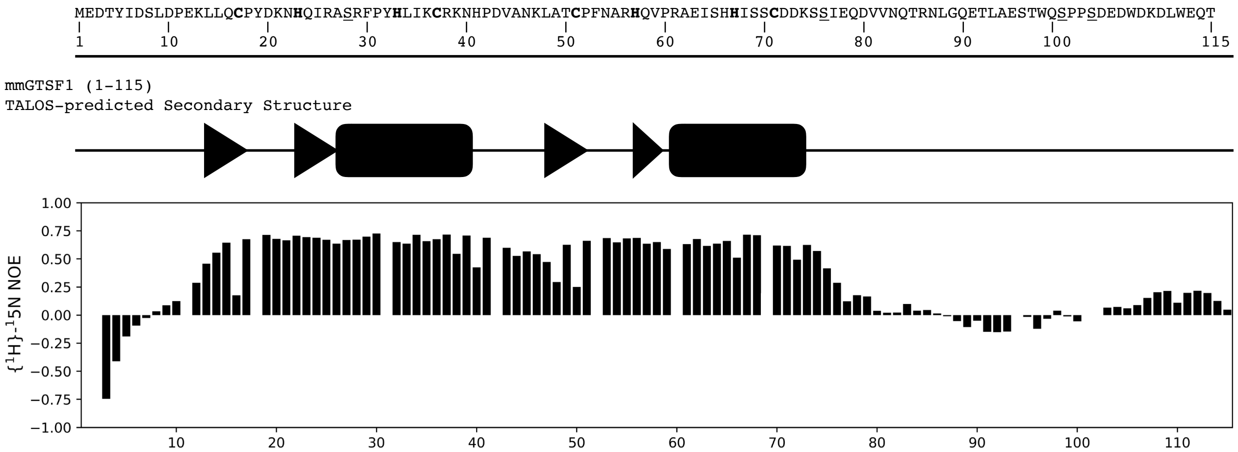


Figure 3.4.1: **TALOS-N Predicted Secondary Structure for GTSF1 Matches with hetNOE Measurements and Predicted Zn Finger Regions.**

Structure calculations were performed with ARIA (Ambiguous Restraints for Iterative Assignment) [67]. All three NOESY-HSQC spectra were manually assigned initially. Less than 10% of all assignments were labeled ambiguous after initial and final ARIA structure calculations, which gave support to the accuracy of NOESY cross-peak assignment. Structure calculations were run in an eight-step iterative scheme, with thresholds tightening after each subsequent iteration. A final ensemble of the 20 lowest-energy structures of GTSF1 is presented in Figure 3.4.2. Alignment structures over all residues (calculations were performed for M1-D80) proved very uninformative. The ensemble alignment is displayed in Figure 3.4.3 with backbone atoms only. Apparent from the structural models is the predicted β-β-α topography of each Zn finger. Taking a calculated structure from the ensemble at random, Figure 3.4.4 shows the molecular architecture of GTSF1 (M1-D80). Two Zn finger domains are bridged by a short linker α-helix, consisting of residues P42-D43-V44-A45-N46. Residues H41 and K47 are classified as possessing no secondary structure according to their backbone ϕ/ψ angles. Further investigation of this linker region explained the inability to perfectly align the 80-residue structure. As shown in Figure 3.4.5, alignment of two randomly chosen structures from the NMR ensemble over the i) L14-N40, ii) the 5-residue a-helix linker, and iii) L48-D72 individually aligns (less than 1 Å RMSD). However, the lack of restraints for both H41 and K47 allowed them to populate different orientations. Thus, the individual Zn finger domains are aligned differently among the ensemble and do not allow a clear superimposition. Finally, because of the hinge introduced by the α-helix linker, the two Zn finger domains are oriented in an anti-parallel fashion.

Structural alignment of each Zn finger domain proved to be a more informative approach. Presented in Figure 3.4.6, alignment is first restricted to the region L14-N40 (left), and then to residues L48-D72. While this window extends outside the beginning and end of the CHHC Zn fingers (C17/C51 and C37/C71, respectively), these alignment limits are based off the beginning and end of calculated secondary structure elements. RMS deviation for heavy atoms was 1.25 Å RMSD for Zn finger 1 and 1.0 Å RMSD for Zn finger 2. Only GTSF1 residues F53 and N54 were located in the Ramachandran disallowed region.

Two notable observations emerge from the NMR structural model of GTSF1 residues M1-D80. First, the conserved residues R38 and K39 demonstrated a localized positively charge surface on the α-helix of Zn finger 1. Likewise, the side-chain of R61 also projects in the same direction, though this residue position is not conserved for arginine. Presumably, GTSF1 interacts with RNA; the positioning of positively charged residues might provide interaction with the negatively charged phosphate backbone of nucleic acid. A second observation involved structural comparison between the two Zn finger domains. Whereas the last Zn ligand (C37) in GTSF1 Zn finger 1 is located in a ten-residue α-helix, the last Zn ligand in Zn finger 2 (C71) is instead located at the end of an α-helix. Though they are differentially populated in the structural ensemble, the three-residue segments D72-K74 and S76-Q79 after C71 form α-helices. Furthermore, comparison of ^{13}C chemical shifts showed a larger downfield chemical shift for C37 vs. C71 (64 ppm vs. 61 ppm, respectively), which indicates tighter binding of the Zn ion by C37. Together, the structural and chemical shift data indicate that GTSF1 Zn finger 2 has a less stable Zn finger architecture. This potentially has implications on its RNA binding capabilities.

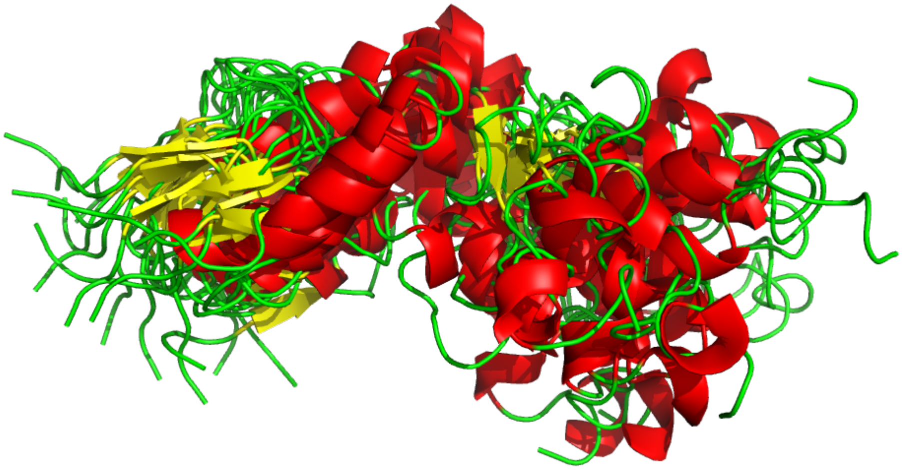


Figure 3.4.2: **Superposition of GTSF1 Structural Ensemble, Shown in Cartoon Form.**

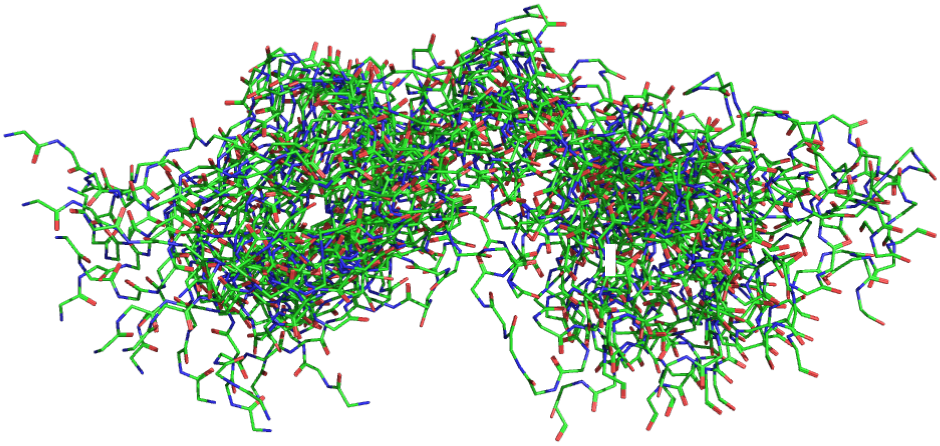


Figure 3.4.3: **Superposition of GTSF1 Structural Ensemble, Displayed with Backbone Atoms Only.**

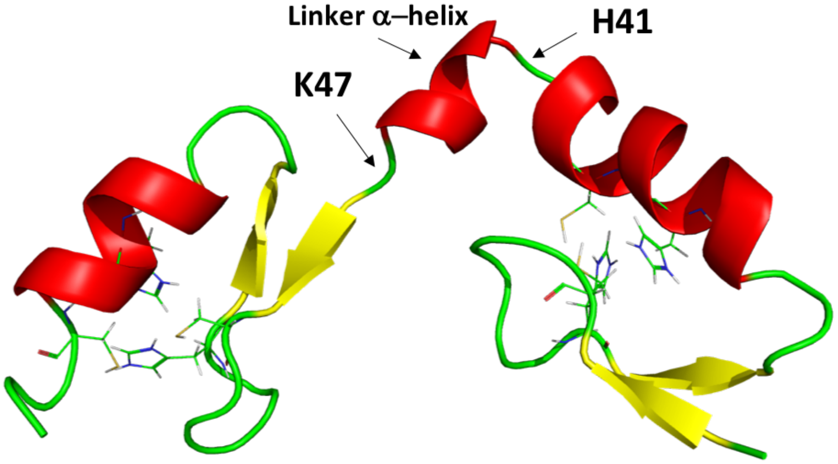


Figure 3.4.4: **Molecular Architecture of GTSF1: Zn fingers, Consisting of Two Anti-parallel β-sheets Followed by an α-helix, are Separated by a Hinge Linker Region.**

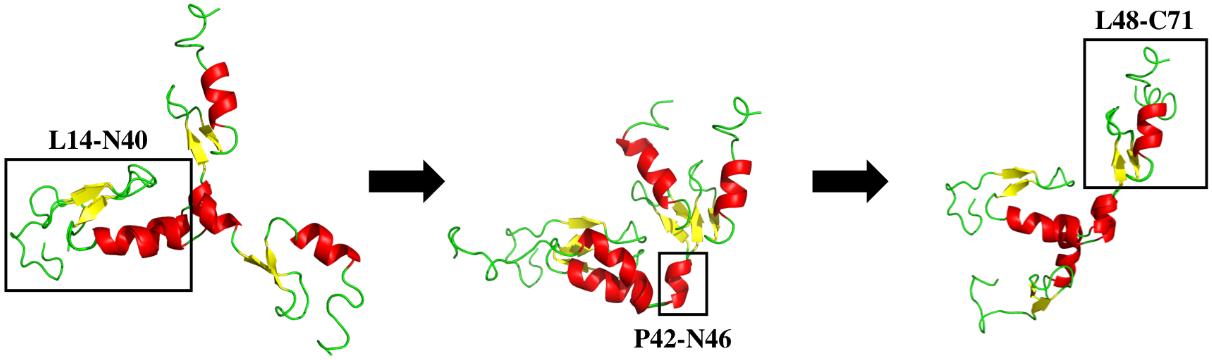


Figure 3.4.5: **The Flexibility of GTSF1 Residues H41 and K47 Prevent a Constant Orientation of Zn finger Domains and a Highly Superimposable GTSF1 Ensemble.**

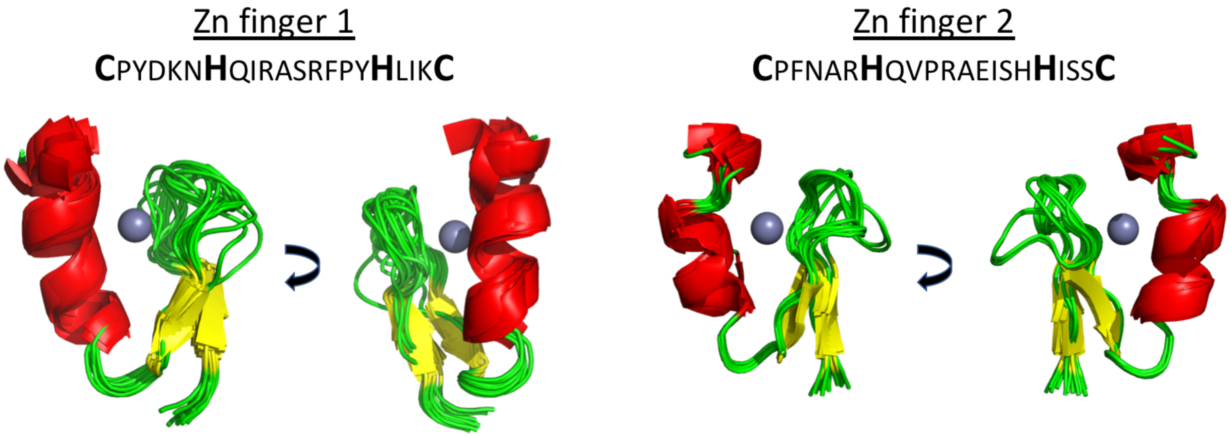


Figure 3.4.6: **Overlay of Individual GTSF1 Zn finger Domains 1 and 2 Shows Highly Superimposable Structures.**

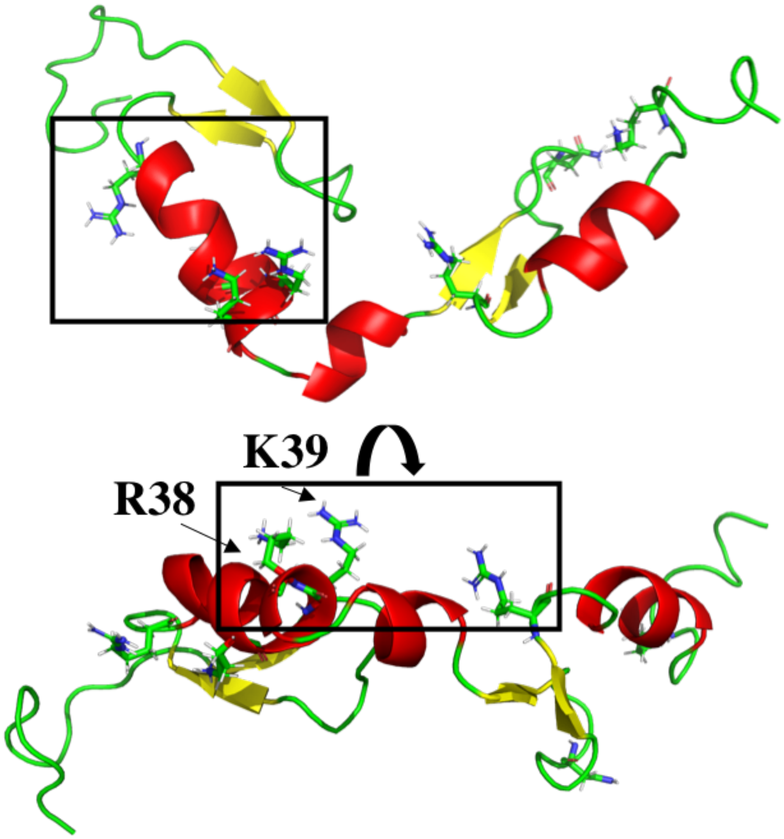


Figure 3.4.7: **Positively Charged Residues That are Conserved Across GTSF1 Sequences Project into a Region Capable of RNA Binding.**

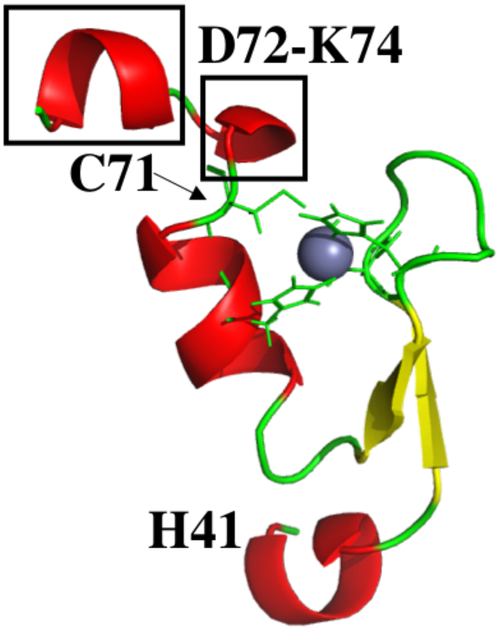


Figure 3.4.8: **GTSF1 Zn Finger 2 Shows a Peculiar Absence of Secondary Structure for Zn-ligating Residue C71.**

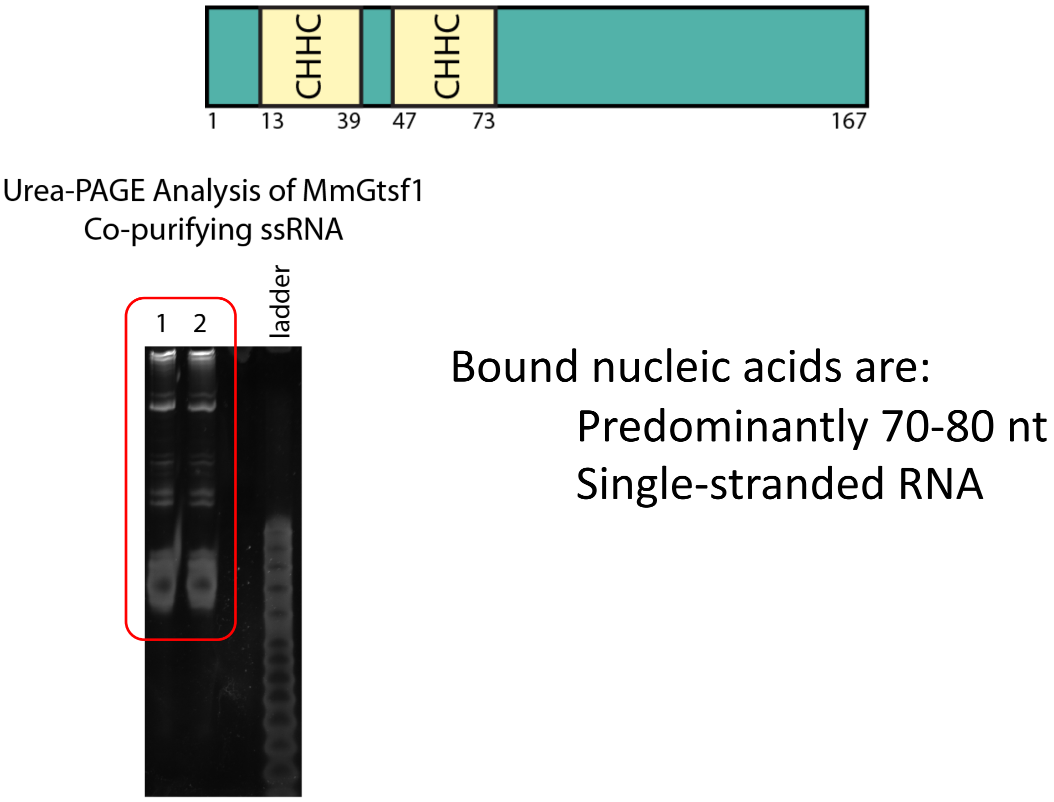
3.5 Hold on Tight: The Curious Case of GTSF1 Binding of RNA*

**Collaboration with Jon Ipsaro and Leemor Joshua-Tor (CHSL)*

3.5.1 Initial Observation and Confirmation of GTSF1 RNA Binding

The individual CHHC Zn finger domains of GTSF1 show relation to the CHHC Zn finger domains of two other protein families: U11-48K and TRM13. The former is a part of the minor spliceosome and theorized to stabilize base pairing that occurs at the U11-5' splice site [68], while the latter is a methyltransferase that binds tRNA for base-specific methylation [69]. These individual functional roles involve a common theme of RNA binding, and should have implications on GTSF1 function.

The initial clues of GTSF1 nucleic acid binding came from an unlikely source. When purifying GTSF1 protein constructs that were expressed in SF9 insect cells, single stranded RNA of ~70-80 nucleotides in length was observed to remain bound to GTSF1 until the ion chromatography purification step (Figure 3.5.1). Under increasing ionic strength, the RNA species dissociated from GTSF1. To better understand nucleic acid binding, five truncated GTSF1 constructs were expressed in SF9 cells, bound to Ni-NTA columns and denatured, and the presence of co-purifying ssRNA was assayed by Urea-PAGE. Figure 3.5.2 shows the results of this GTSF1-RNA truncation series. GTSF1 constructs comprising residues 1-45, 1-75 and 1-167 (full-length) all exhibit RNA binding. However, GTSF1 constructs consisting of residues 46-167 and 76-167 were observed to not co-purify with RNA in this experiment. From this dataset, the first CHHC Zn finger of GTSF1 appeared to be required and sufficient for RNA binding. The second CHHC Zn finger, though, appears dispensable for RNA binding given the absence of co-purifying RNA found in lanes corresponding to expression of GTSF1 constructs 46-167 and 76-167.

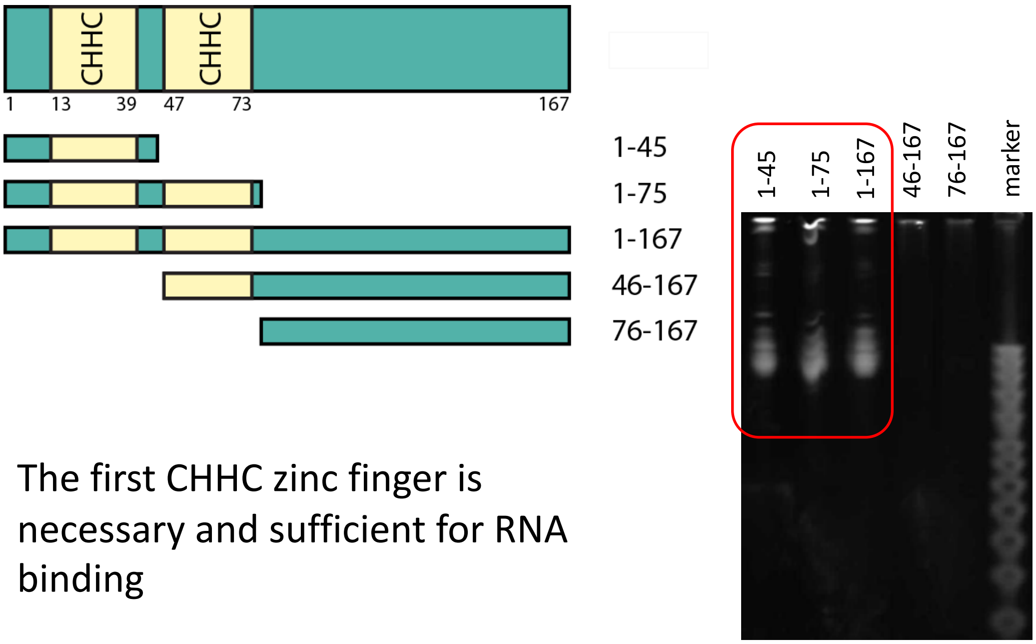


Urea-PAGE Analysis of MmGtsf1
Co-purifying ssRNA

Bound nucleic acids are:
Predominantly 70-80 nt
Single-stranded RNA

Figure 3.5.1: ***Mm*GTSF1 RNA-binding Characteristics.**

Single-stranded RNA (ssRNA) remained associated with GTSF1 protein until the ion chromatography step of recombinant protein purification. From urea-PAGE analysis, the nucleic acid appeared to be a heterogenous mixture with the majority of species 70-80 nucleotides long. This complex was observed with full-length GTSF1 expression from SF9 insect cells.



The first CHHC zinc finger is necessary and sufficient for RNA binding

Figure 3.5.2: **Truncations of *Mm*GTSF1 Reveal Zn Finger 1 as the RNA-binding Site.** To determine the minimal RNA-binding region of GTSF1, different size protein truncations (shown in the top right figure) were expressed in SF9 and the presence of co-purifying RNA was assessed urea-PAGE. The results were unequivocal: after tests of five different constructs, Zn finger 1 (C17-N40) was the region common to all three experiments that resulted in ssRNA co-purification with GTSF1.

3.5.2 First Attempt at Assaying Nucleic Acid Binding: A Heterogenous RNA Titration

A quick test of GTSF1 RNA binding was performed using the RNA isolated from SF9-based GTSF1 purification. The RNA was refolded into GTSF1 NMR buffer and added to a 200 µM U-[^{15}N] sample of GTSF1 at a 1:1 stoichiometric ratio. Using 1.7 mm NMR tubes, which required only ~33 µL of sample, ^{15}N HSQC spectra were recorded with and without RNA. The result is shown in Figure 3.5.3. Comparing the apo-GTSF1 spectrum and GTSF1 plus RNA spectrum, the dispersed Zn finger peaks disappear entirely. On the other hand, most GTSF1 non-structured peaks were left unperturbed. Clearly GTSF1 interacts with RNA. However, the ssRNA species doesn't appear to specifically interact with either of the Zn finger domains. The molecular weight complex is approximated at 35 kDa, which puts it beyond the limit of accessibility for HSQC-based correlation experiments in a non-deuterated background. In an effort to reduce the rotational correlation time of the complex, the sample temperature for the experiment was increased from 295 K to 305 K. While some signal from dispersed peaks presented at the higher temperature, the majority of Zn finger peaks were still unobserved. These observations are consistent with the hypothesis that the isolated ssRNA population is heterogenous, but possesses a similar binding motif capable of mediating interaction with GTSF1. Further probing of the underlying ssRNA subpopulations is necessary.

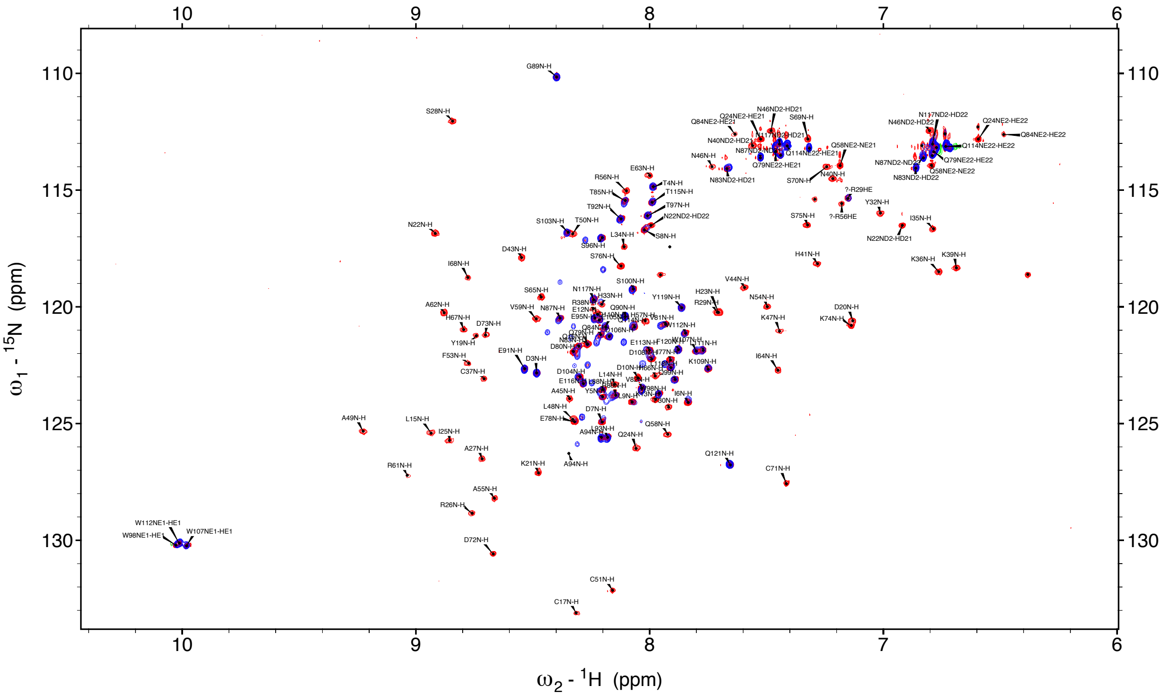


Figure 3.5.3: **Titration of GTSF1 with Heterogeneous Co-purifying ssRNA (295 K).** Addition of co-purifying ssRNA to 200 µM GTSF1 resulted in loss of Zn finger 1 and 2 amide peaks from a ^{15}N HSQC spectrum (red peaks correspond to apo-GTSF1, while blue peaks are GTSF1 plus ssRNA). Even reducing the spectrum's contour level to baseline, no peaks are observed. On the other hand, C-terminal GTSF1 residues V80-T115 show minimal chemical shift perturbation.

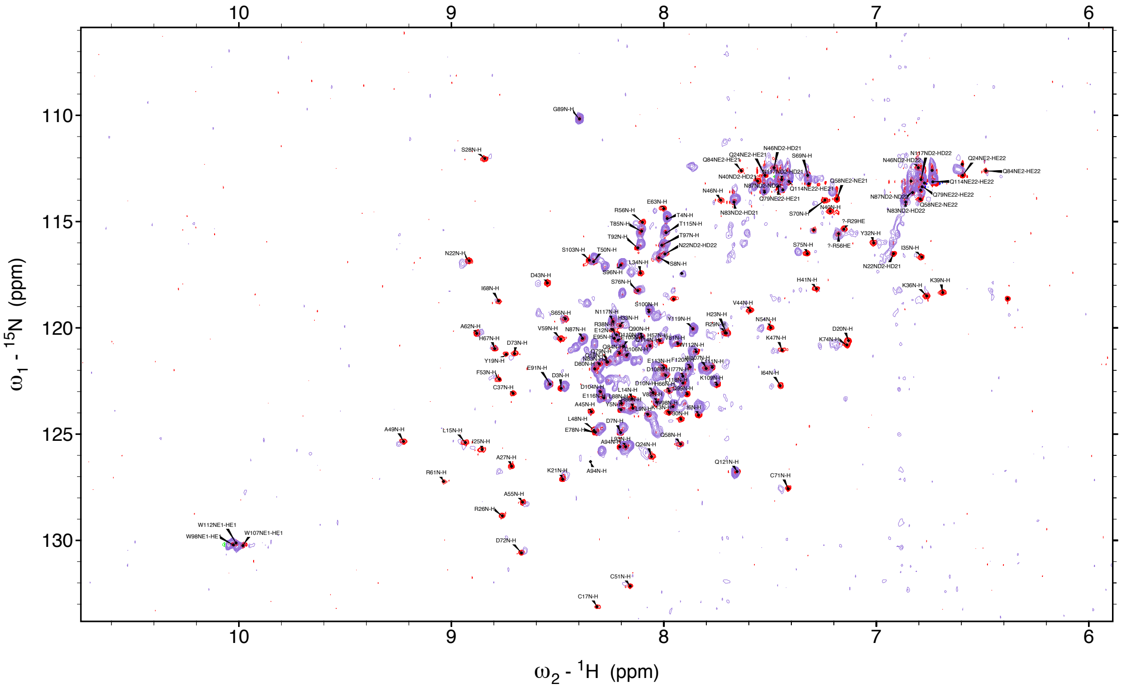


Figure 3.5.4: **Titration of GTSF1 with Heterogeneous Co-purifying ssRNA (305 K).** Recording the same experiment at 305 K, peaks corresponding to GTSF1 Zn finger 1 and 2 appeared in the ^{15}N HSQC spectrum with the contour level adjusted to barely above spectral noise (red peaks correspond to apo-GTSF1, while purple peaks are GTSF1 plus ssRNA). Chemical shift perturbations are visible in the spectrum, but RNA binding isn't confined to just a single Zn finger domain.

3.5.3 Assay of T-arm RNA Binding to GTSF1

Initial results from RNA library generation and sequencing analysis identified the six-nucleotide motif *GGUUC(G/A)* in the highest enriched set of RNA reads. While this repeat showed an extremely high p-value for enrichment, it is not immediately clear if the six-nucleotide motif definitely binds GTSF1 or if the motif just happens to occur in sequences with a more nuanced RNA motif that is responsible for association with GTSF1. Using ^{15}N-labeled protein samples of GTSF1, binding of a candidate read from the RNA library that contains the six-nucleotide motif was assessed by NMR spectroscopy. The RNA candidate synthesized and used for GTSF1 titration was a 19-nucleotide ssRNA that corresponded to a portion of the T-arm from tRNA. The sequence represented an enriched read in the RNA library and contained *GGUUCG*, with two caveats: the first uracil is a methylthymidine (rT), and the second uracil is a pseudouridine (~U). Thus, the six-nucleotide motif here is *GG(rT)(~U)CG*, which corresponds to the TωC loop of tRNA (Figure 3.5.5).

GTSF1 and T-arm RNA were mixed at a 1:1 stoichiometric ratio, with protein concentrations of 500 μM each. Due to material constraints, folding of the ssRNA was assessed following addition to the protein sample. ^{1}H jump-return experiments resolved guanine and uracil imino proton peaks in the 12-14 ppm range (Figure 3.5.5). For this RNA sequence, the ^{1}H spectrum allowed identification of: five of six hydrogen-bonded guanine H1' protons, all two of the uracil H3' protons, pseudouridine H1' and H3' protons, and one non-hydrogen bonded guanine H1' proton. These spectral observations confirm the presence of the T-arm RNA secondary structure elements. Despite this, the individual peak linewidths and bases of the 1H spectrum in Figure 3.5.5 appear slightly broadened. This issue is revisited later.

An overlay of apo-GTSF1 and RNA-bound GTSF1 ^{15}N HSQC is shown in Figure 3.5.6. The loss of dispersed Zn finger peaks is immediately visible. The few peaks that remain after T-arm RNA addition are assigned to residues in Zn finger 2. On the contrary, intense peaks corresponding to the non-structured regions of GTSF1 appear minimally perturbed. A deeper inspection of GTSF1 non-structured residues reveals that RNA titration selectively perturbs T85-T115, while residues S75-T85 and D3-D10 show less or no chemical shift perturbations. Regions of the ^{15}N HSQC that better demonstrate this are shown in Figure 3.5.7.

Altogether, GTSF1 titration with T-arm RNA demonstrates a definite interaction between the two species. However, as with the observation of heterogeneous ssRNA binding to GTSF1 (Chapter 3.5.2), amide correlations from both Zn fingers disappeared upon addition to 19-nucleotide T-arm RNA. Furthermore, the spectra in both instances shows RNA-binding induced global effects on GTSF1. This experiments observation stands in contrast to Chapter 3.5.1, where binding experiments isolated Zn finger 1 of GTSF1 as necessary and sufficient for RNA binding. Either the TΨC loop of tRNA is a non-specific binding site for GTSF1 (or not the GTSF1 binding site at all) or the 19-nucleotide T-arm RNA construct is not populating a fully binding-competent conformation. The ^{1}H jump return spectrum did show broadened linewidths and peak bases for G and U imino protons; the RNA population is potentially non-homogenous and consisting of improperly formed RNA dimer during refolding or just unstable T-arm RNA. To address this, a more stabilized stem-loop RNA construct was selected for synthesis.

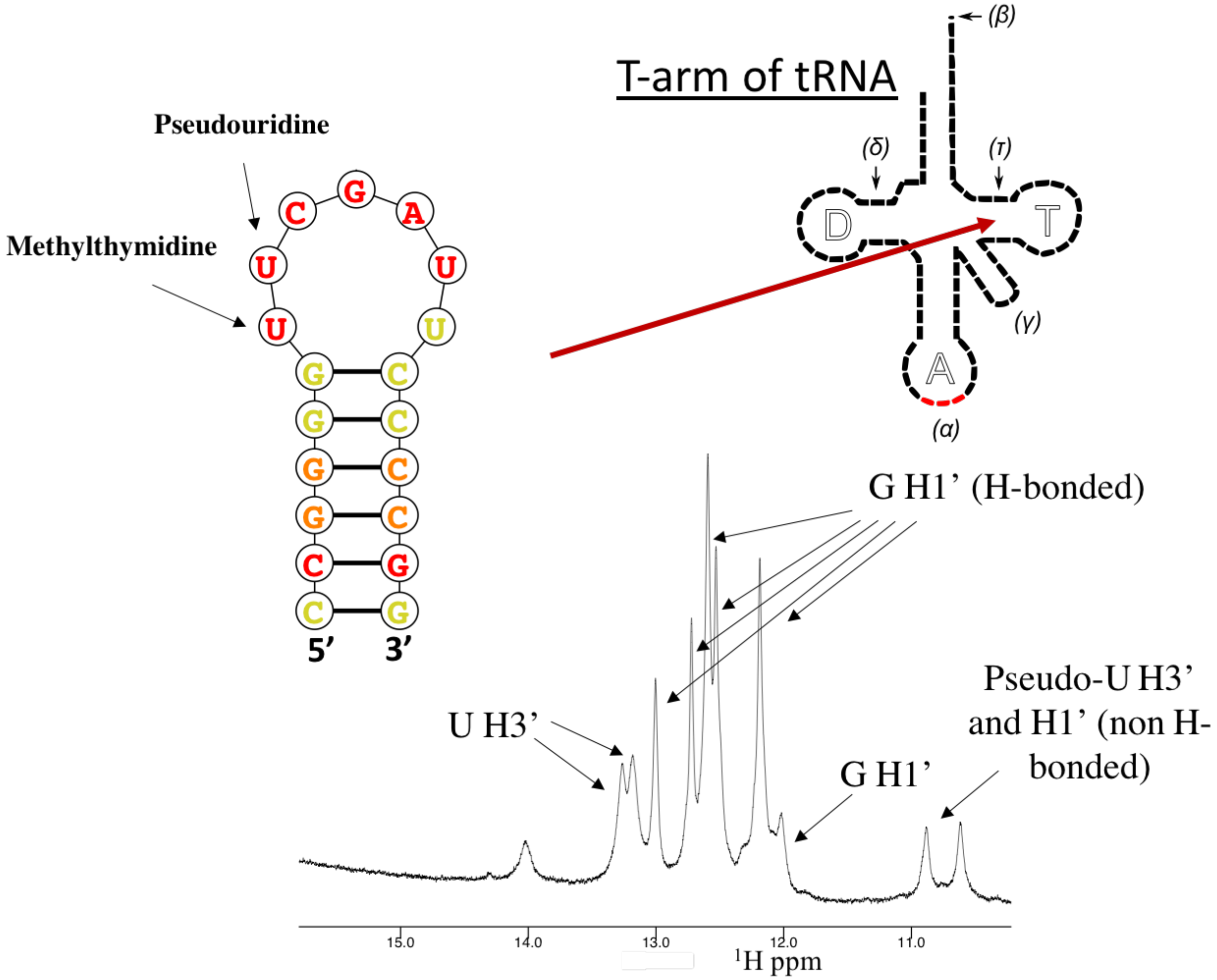


Figure 3.5.5: **Titration Experiments Assayed the Binding of T-arm tRNA to GTSF1.** The T-arm contains a known interaction motif named the TΨC loop, whose consensus sequence is *UUCGA* and occurs with similar posttranslational modifications as the identified RNA motif. A T-arm RNA construct (pictured to the right above) was added in equimolar amounts to GTSF1. A ^{1}H spectrum identified five hydrogen-bonded guanine nucleotide peaks, with the sixth guanine presumably undetected because of exchange broadening. Additionally, two uridine imino peaks, a non-hydrogen bonded guanine, and two pseudo-uridine non-hydrogen bonded peaks were identified.

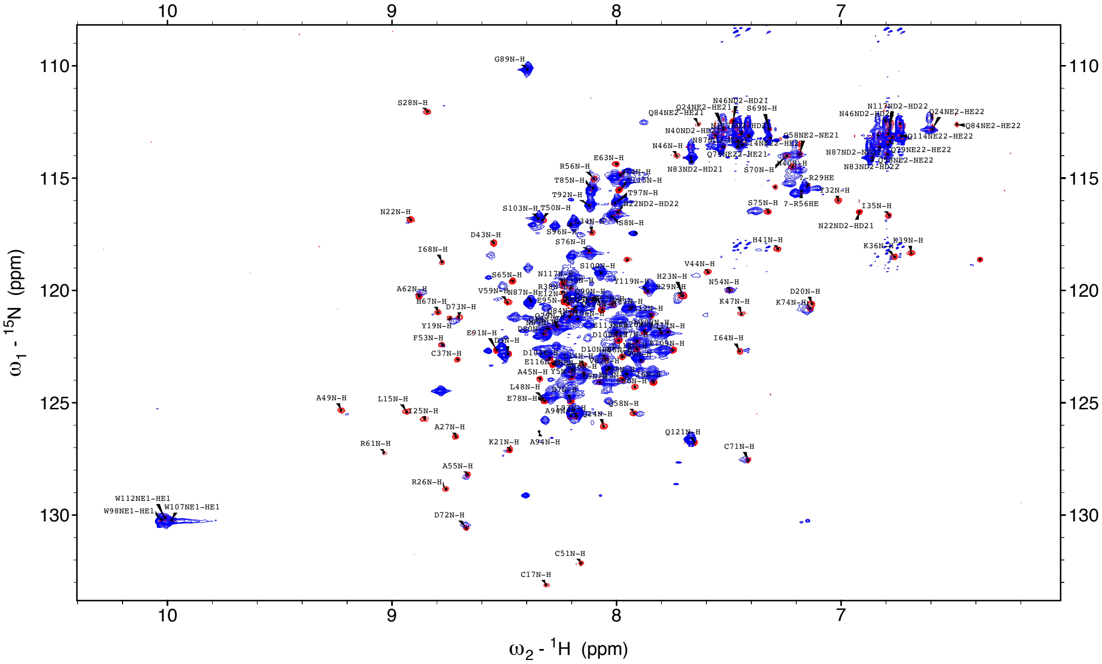


Figure 3.5.6: **^{15}N HSQC of Stoichiometric 1:1 GTSF1 and T-arm RNA.** Similar to the spectra of heterogeneous co-purifying ssRNA added to GTSF1, addition of T-arm RNA appeared to interact with both Zn fingers (red peaks correspond to apo-GTSF1, while GTSF1 plus RNA is represented by the blue peaks). With the exception of a few Zn finger 2 peaks (A49, A55, A62, C71, D72, D73, K74) peaks from both Zn finger domains were not visible in the ^{15}N HSQC. Again, RNA interactions with GTSF1 did not mimic the biochemical evidence implicating the first Zn finger as being responsible for RNA binding.

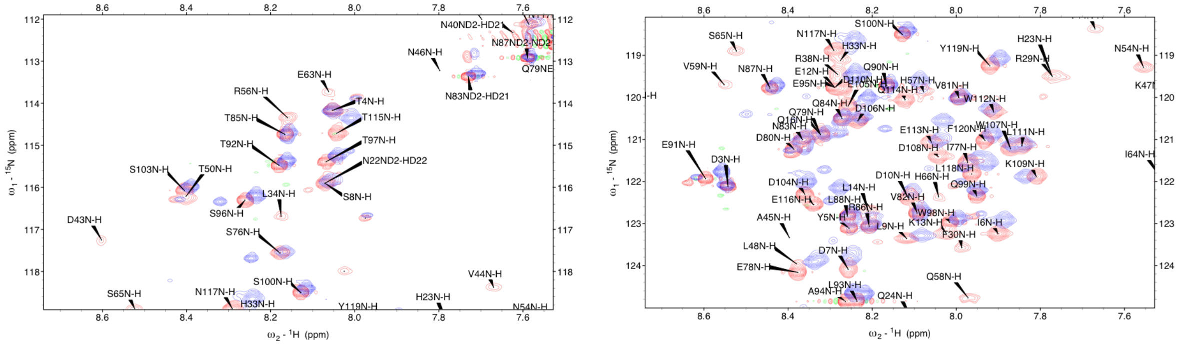


Figure 3.5.7: **^{15}N HSQC of Stoichiometric 1:1 GTSF1 and T-arm RNA (inset).** The titration involved 500 μM T-arm RNA added to 500 μM GTSF1. With the loss of the majority of peaks assigned to Zn finger residues, the remaining intense peaks show a clear trend (red peaks correspond to apo-GTSF1, while GTSF1 plus RNA is represented by the blue peaks). On average, residues T85-T115 demonstrated significant chemical shifts perturbations, while the first 10 residues of GTSF1 and residues S75-Q84 showed less perturbation.

3.5.4 A More Stabilized RNA Construct: T-arm mini-helix RNA Binding to GTSF1

A mini-helix version of the T-arm RNA from Section 3.5.2 was chosen as a second RNA construct for titration with GTSF1. Pictured in Figure 3.5.8, the T-arm mini-helix comprises the initial T-arm RNA plus an additional 16 nucleotides. With a nine base-pair C-G stem, the T-arm mini-helix is believed to be a more stabilized compared to the T-arm RNA construct. This is supported by predictions of ΔG of -24.6 kcal/mol for T-arm mini-helix compared to -11.23 kcal/mol for initial T-arm RNA construct (*RNA*fold Web Server). It was hypothesized that a more stable RNA species offers a more consistent binding surface, in the form of the TΨC loop of the T-arm, to bind with GTSF1. The ^{1}H spectrum towards the bottom of Figure 3.5.8 hints at this: imino proton peaks from base paired guanine and uridine showed narrower linewidths, which is indicative of a monodisperse molecular species and of more isotropic molecular tumbling. In total, nine imino proton peaks were observed. The stem of T-arm mini-helix is predicted to contain nine C-G base pairs, two A-U base pairs, and one non-canonical U-G base pair. Assuming a peak for the final A-U base pair of the stem cannot be observed due to exchange broadening, two base-paired imino peaks are missing from the ^{1}H spectrum. Finally, the spectrum shows only one pseudo-uridine non-bonded proton peak (H1' or H3') as opposed to the two peaks observed in the ^{1}H spectrum of T-arm RNA.

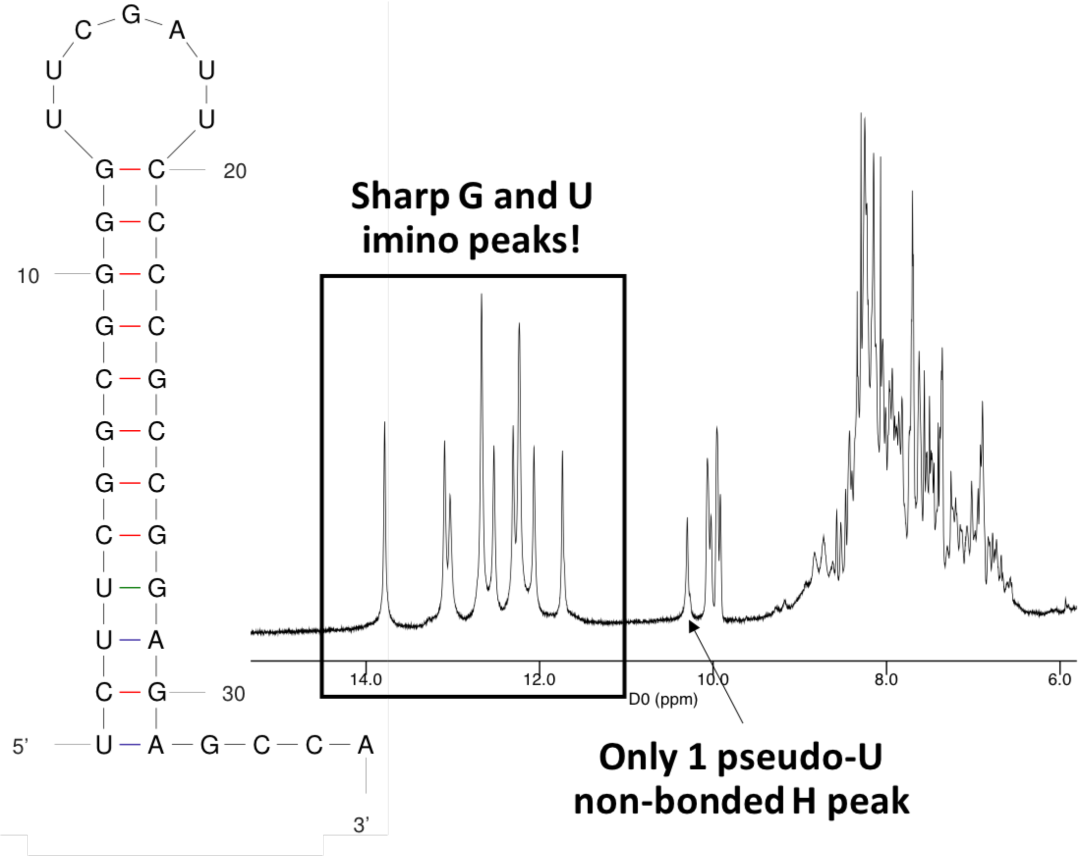


Figure 3.5.8: **A ^{1}H Spectrum of T-arm mini-helix Indicated A More Stable RNA Construct.** Additional C-G and U-A base pairs to the initial T-arm construct (left) make for a higher predicted stability (more stable by 13.37 kcal.mol) and gave ^{1}H jump return HE spectra with sharper and less broadened imino proton peaks (right).

Titration of GTSF1 with T-arm mini helix gave a very promising set of results. Again, 500 μM U-[^{13}C, ^{15}N] GTSF1 was mixed in a 1:1 ratio with T-arm mini-helix RNA (refolded and exchanged into GTSF1 NMR buffer). High resolution ^{15}N HSQC spectra were acquired (Figure 3.5.9). Superficially, comparison of the apo-GTSF1 and GTSF1-T-arm mini-helix complex resembled titration results using just the T-arm RNA construct. However, three aspects of the spectra overlay became apparent after a closer inspection. First, C-terminal disordered residues are exclusively unperturbed by RNA binding. As shown in Figure 3.5.10, chemical shifts for GTSF1 residues S75 through T115 are exactly incident between apo-GTSF1 and RNA-associated GTSF1. The lack of chemical shift perturbation for S75-T85 is notably different compared the effects observed with just T-arm RNA. Second, while at first it seems that all dispersed Zn finger amide peaks vanished after addition of T-arm mini-helix RNA, adjustment of the spectral contour levels reveals that Zn finger 2 peaks appear in their exact chemical shift positions with only a significant change to their peak intensity. Specifically, residues in Zn finger 2 from F53 to K74 are observed (see boxes in Figure 3.5.10). Also, the first ten residues of GTSF1, which are predicted to be disordered and showed no structural preference in GTSF1 NMR models, exhibited chemical shifts perturbations following the addition of T-arm mini-helix RNA. Together, these takeaways from the ^{15}N HSQC spectrum of GTSF1 bound to T-arm mini-helix RNA suggested the localization of the RNA-binding interface to the N-terminal 40 residues of GTSF1. To further probe this interaction, ^{13}C SOFAST HMQC [58] spectra were recorded separately for apo- and T-arm mini-helix-bound GTSF1. The total molecular weight of the ribonucleoprotein should be ~25kDa, which puts it at the end of observation spectrum for HSQC-based NMR experiments (without deuteration). However, SOFAST HMQC-type pulse sequences take advantage of the natural 3-fold increase in signal for methyl groups (three degenerate protons) and beneficial relaxation properties of multiple quantum transverse magnetization. Therefore, a ^{13}C SOFAST HMQC spectrum can serve as a substitute for detection of chemical shift perturbations when amide chemical shift perturbations can't be observed or resolved. Figure 3.5.11 shows the results of such an experiment. Fully assigned ^{1}H-^{13}C methyl correlations from GTSF1 are overlaid with the same correlations from GTSF1 bound to T-arm mini-helix. The results showed chemical shift perturbations for GTSF1 I25 δ1, L15 and L34 δ1, and A27 β methyl groups, while peaks for L14 and I35 δ1 methyl groups disappeared. Methyl groups in the linker region (A45, L48) and in Zn finger 2 (V59, I64, I68) displayed no chemical shift perturbations. The results of this final NMR titration of GTSF1 with T-arm mini-helix clearly recapitulated RNA-binding propensity of Zn finger 1 observed in the co-purifying ssRNA experiment (Figure 3.5.11, left).

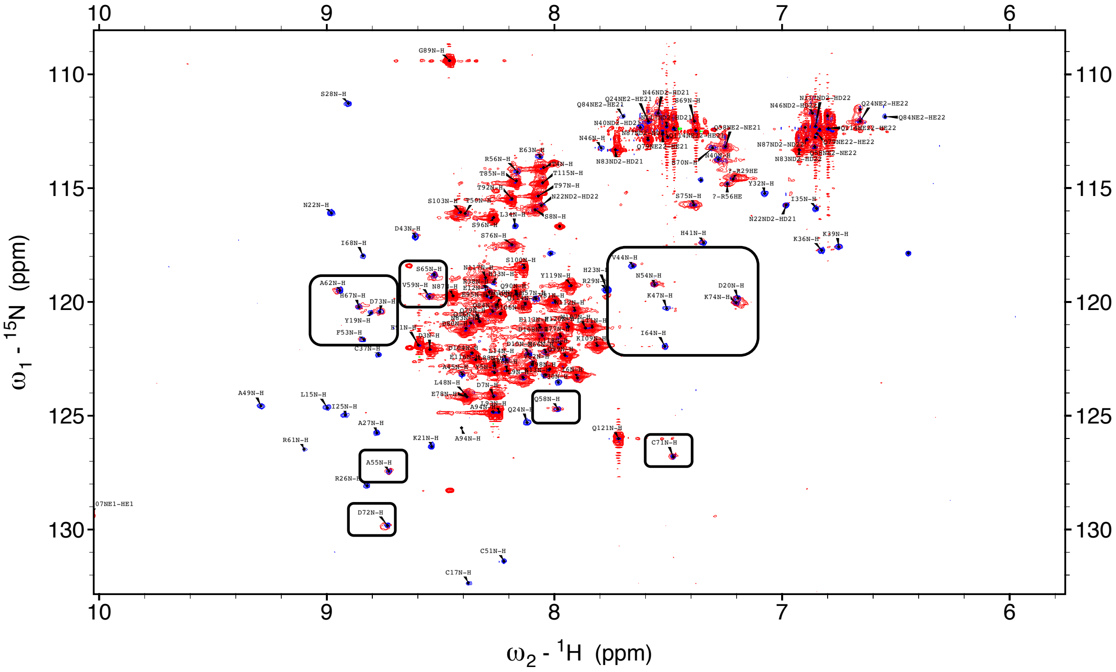


Figure 3.5.9: **Titration of GTSF1 with T-arm mini-helix RNA.**
Comparison of the apo-GTSF1 (blue) and GTSF1 (red) associated with T-arm mini-helix RNA spectra showed residual peak intensity for Zn finger 2 residues with minimal to no chemical shift perturbation. Boxed residues indicate these GTSF1 residues.

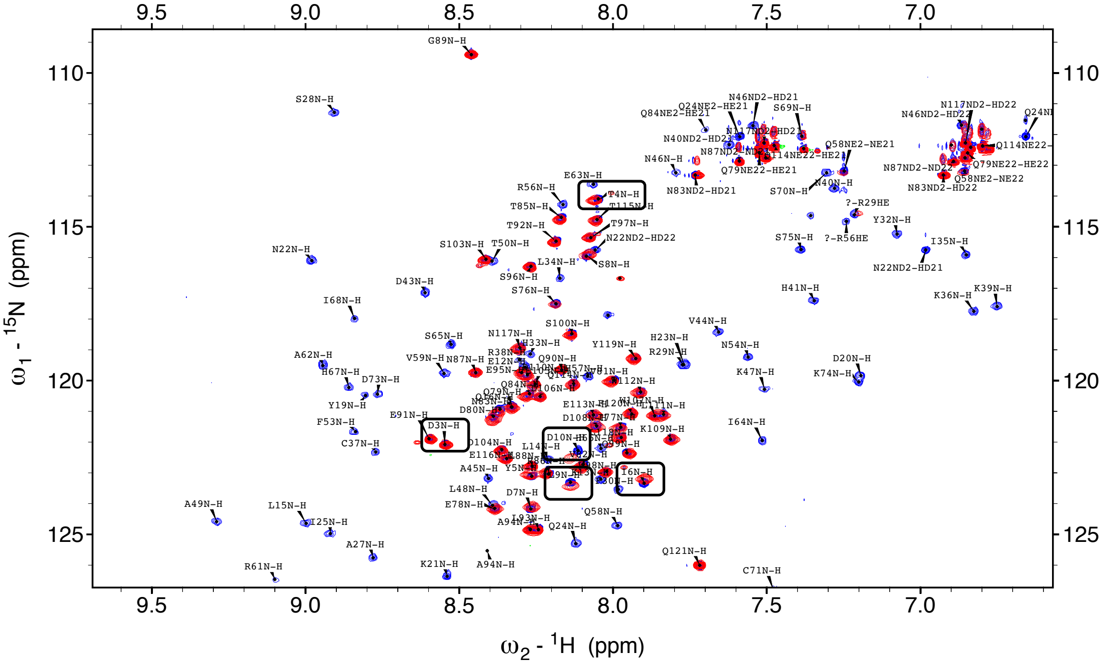
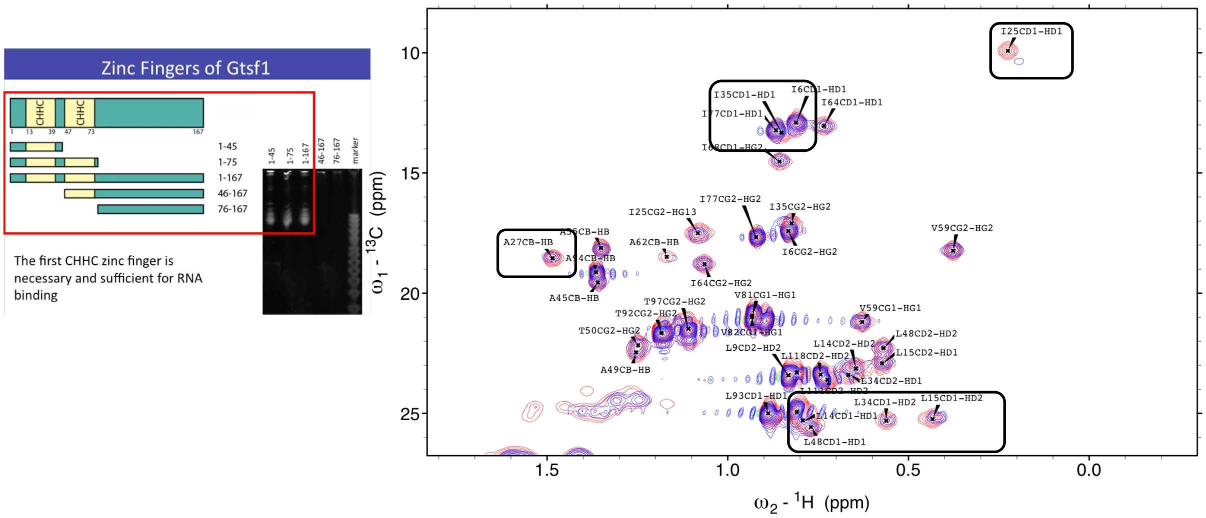


Figure 3.5.10: **T-arm mini-helix Interaction with GTSF1 is Confined to the N-terminal Domain.**

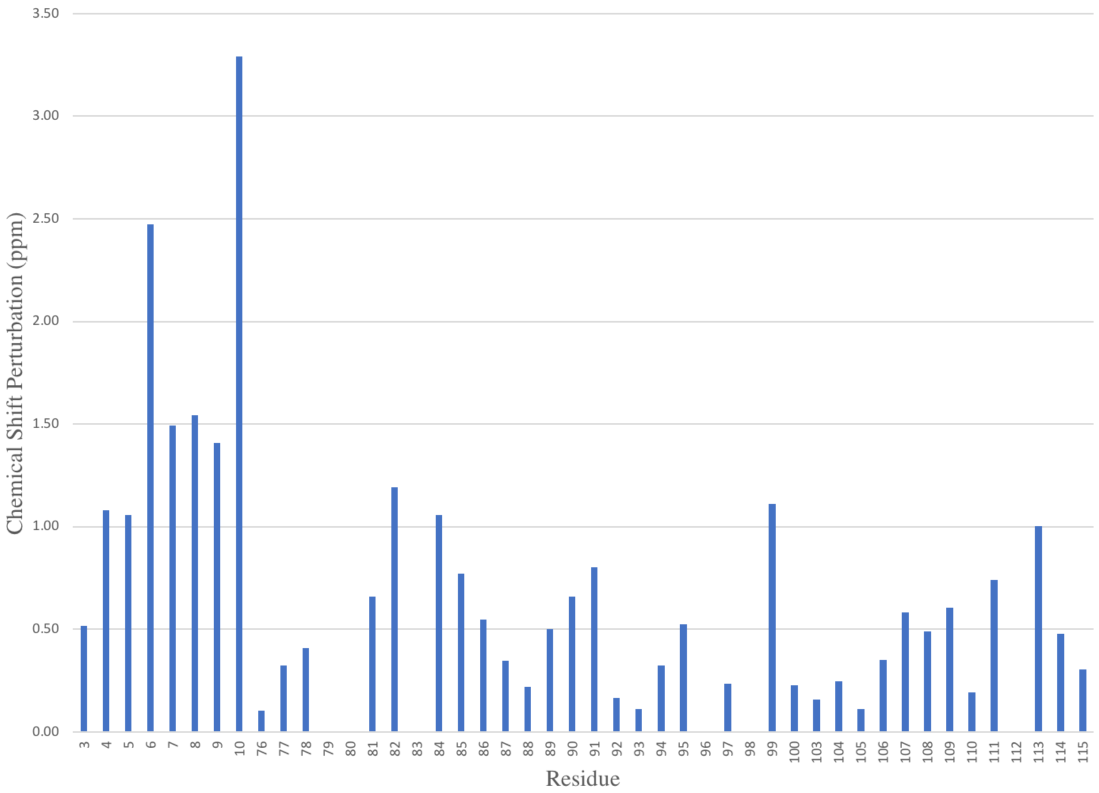
Residues S75-T115 demonstrated no chemical shift perturbations between apo-GTSF1 (blue) and RNA-bound GTSF1 (red). However, the non-structured N-terminal residues, which includes M1-D10, did show chemical shift perturbations (see boxes).



Zinc Fingers of Gtsf1

The first CHHC zinc finger is necessary and sufficient for RNA binding

Figure 3.5.11: **AILV Methyl Probes of RNA Interaction with GTSF1.** 13C SOFAST HMQC experiments were recorded to probe T-arm mini-helix binding to GTSF1. As shown in the boxed residues, addition of RNA (blue) coincided with peak movement or peak disappearance for AILV methyl correlations in Zn finger 1.



81

Figure 3.5.12: **T-arm mini-helix RNA Binding Selectively Affected the N-terminal of GTSF1.**

Chemical shift perturbations were calculated using CSP = $100\times[(0.15\times D_N)^2+(D_H)^2]^{1/2}$. Non-structured residues between D3 and D10 demonstrated larger CSPs compared to disordered C-terminal residues S76-T115. Note that CSPs were scaled by 100 to facilitate ease of viewing.

3.6 Discussion

The trajectory of Chapter 3 described the isolation of a GTSF1 protein construct amendable to study by NMR spectroscopy, GTSF1 backbone and side-chain chemical shift assignment, use of structural restraints from NOE cross peaks to determine the molecular architecture of the N-terminal 80 residues of GTSF1, and finally the identification of an *in vivo* RNA binding partner for GTSF1. To summarize, this Chapter has presented compelling data that GTSF1 Zn finger 1 specifically binds a motif in the T-arm mini-helix RNA. This motif, which from analysis of an

RNA library generated from co-purifying RNA was *GGUUC(G/A)*, presents as four of the seven nucleotides of the T-arm TΨC loop. Zn finger 2 is affected by the interaction with RNA, but data points to the loss of peak intensity resulting from the larger molecular weight and increased molecular tumbling time of the GTSF-RNA complex. Furthermore, careful inspection of the Zn finger 2 architecture and cysteine chemical shifts indicates that the second Zn finger of apo-GTSF1 is more dynamic than Zn finger 1. Potentially, the function of GTSF1 first requires binding of RNA to the more stable Zn finger 1 and only then is Zn finger 2 poised for binding to another molecular species. Without speculating too much, the accommodation of various orientations of the Zn finger domains by the linker region might help in binding two separate RNA species and bringing them in close proximity.

Biologically, the identification of tRNA-derived RNA as an *in vivo* binding partner of GTSF1 was initially confusing. Although knockdown of GTSF1 is associated with stalled ribosomes, genetic and biochemical experiments indicated a more intimate relationship between GTSF1 and the piRNA-induced silencing complexes (piRISCs). However, only recently has a glut of experimental evidence supporting the involvement of so-called tRNA-derived fragments (tRFs) in transcriptional- and post-transcriptional gene silencing. Both 5' and 3' cleavage products of tRNA have been identified in various organisms as functioning as tRFs. Recently, 18 and 22 nucleotide tRFs were demonstrated to silence retrotransposon activity by inhibition of reverse translation (18 nucleotide tRF) and post-transcriptional gene silencing (22 nucleotide tRF) [70]. This is presented in Figure 3.6.1. Pivoting back to GTSF1, there is a considerable evidence documented for the interplay between tRNAs and transposable elements. Certain tRNAs are used as primers for transcription or reverse transcription of transposons [71], while other tRNAs are needed for their specific charged-amino acid [72]. Moreover, piRNAs have been isolated that derive from tRNAs [73]. Regardless the source, tRFs can be regarded as an indicator of transposon activity. In this instance, cleavage of tRNA into binding-competent tRFs could be a signal to localize piRISC to genomic locations of active transposons. In Figure 3.6.2, a model is depicted where RNA-activated GTSF1 (Asterix) is a guiding signal for the piRNA-Piwi complex to promote histone methylation (by way of SETDB1), heterochromatin formation, and binding of HP1α. More work is being conducted to more precisely determine how RNA-activated GTSF1 binds with piRISC and/or active genomic transposon sites.

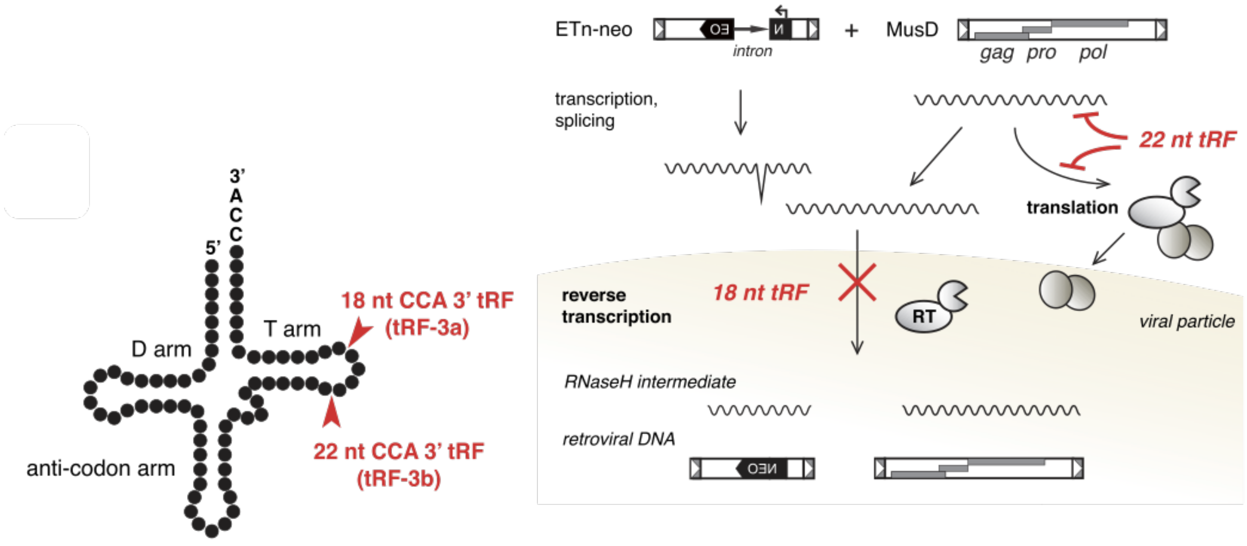


Figure 7. 22 nt tRFs Mediate Post-transcriptional Gene Silencing while 18 nt tRFs Interfere with Reverse Transcription Including Non-coding, Mobile Elements

Model of retrotransposon silencing by 3′ tRFs: 22 nt tRFs target coding-competent LTR-retrotransposons (here MusD) at the level of retroviral protein production. 18 nt tRFs inhibit reverse transcription of any element with perfect complementarity at the PBS, including non-autonomous elements (here ETn). tRFs specifically promote silencing of retrotransposition-competent elements that maintain a functional PBS.

Figure 3.6.1: **3' tRNA-derived Fragments (tRFs) Have Been Demonstrated to Function in Retrotransposon Silencing.**

tRFs of 18 nt and 22 nt in length have been confirmed to function in retrotransposon silencing through inhibition of reverse transcription and promotion of post-transcriptional gene silencing. The 22 nt tRF contains the *GGUUC(G/A)* motif identified from analysis of the generated RNA library, and thus mimics the T-arm mini-helix RNA used for titration with GTSF1. Reprinted by permission from Elsevier: Cell, 170, Schorn, A.J., Gutbrod, M.J., LeBlanc, C., Martienssen, R. LTR-retrotransposon control by tRNA-derived small RNAs. 61-71, Copyright 2017. [70]

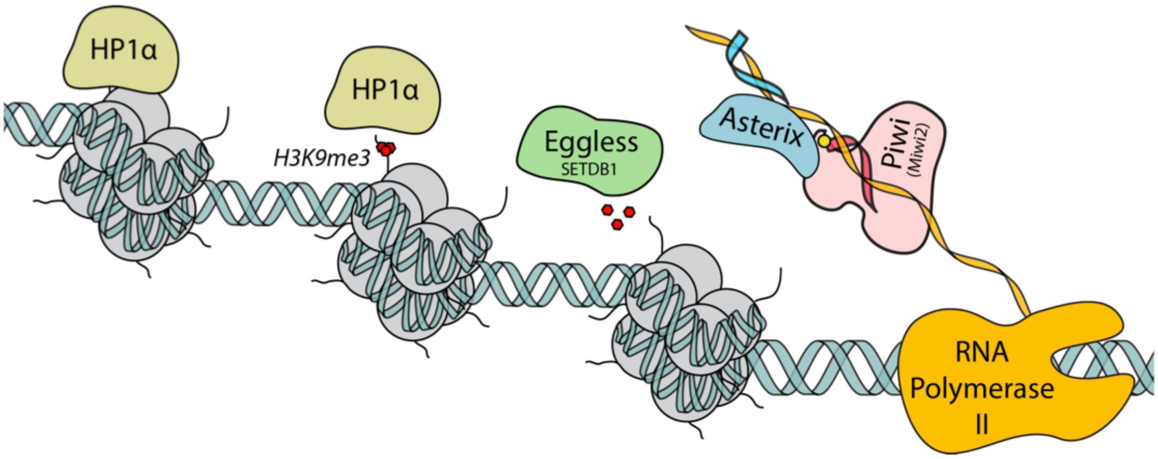


Figure 3.6.2: **Proposed Model of GTSF1's Role in the piRNA Effector Step.** Results from this study implicate GTSF1 as a sensor of the tRF population. However, it can't be ruled out that the RNA motif discovered in this work is not present in other non-coding RNA molecules that function in the piRNA pathway or RNAi in general. As was shown by Keam et al., certain piRNAs are derived from tRNAs in somatic cells, which hints at different functional roles between somatic and germline environments.

Assignment Table A: **GTSF1 Backbone Chemical Shifts.**

Residues	C	CA	CB	H	N
E2	175.8	56.41	30.36	-	-
D3	176.2	54.53	41.11	8.534	122.1
T4	174.3	62.04	-	8.037	114.1
Y5	175.6	58.27	38.71	8.252	123.1
I6	175.6	60.94	38.9	7.893	123.3
D7	176.3	54.5	41.31	8.258	124.1
S8	174.4	58.48	63.73	8.072	115.9
L9	176.9	55.12	42.24	8.127	123.3
D10	-	52.28	41.17	8.105	122.3
P11	-	63.62	32.38	-	-
E12	176.2	56.15	30.78	8.283	119.5
K13	175.2	56.53	33.34	8.028	123.2
L14	176.4	53.84	42.95	8.205	122.6
L15	-	52.98	43.5	8.997	124.7

Q16	175.2	54.54	30.25	8.324	120.9
C17	-	57.5	33.58	8.371	132.4
P18	176.1	63.61	32.53	-	-
Y19	175.5	59.35	39.05	8.802	120.5
D20	175.8	53.66	42.26	7.198	119.9
K21	177	58.14	32.08	8.532	126.3
N22	175.5	53.9	38.83	8.972	116.1
H23	172.8	53.22	32.71	7.763	119.5
Q24	174.6	55.57	29.7	8.107	125.3
I25	175.6	57.28	40.18	8.92	125
R26	179	57.25	30.57	8.822	128.1
A27	180.4	56.7	18.59	8.776	125.8
S28	176.3	61.2	62.89	8.903	111.3
R29	176.8	54.63	30.82	7.764	119.5
F30	-	63.48	38.44	7.981	123.5
P31	178.9	66.59	30.93	-	-
Y32	178	60.07	37.83	7.069	115.3
H33	176.7	60.29	29	8.259	119.2
L34	178	58.39	41.69	8.166	116.7
I35	178.1	64.72	38.25	6.853	115.9
K36	178.2	58.48	32.57	6.825	117.8
C37	-	64.63	28.92	8.769	122.3
R38	179	59.24	30.77	8.3	119.4
K39	177.4	58.33	32.14	6.75	117.6
N40	174.2	52.73	38.94	7.281	113.8
H41	-	54.31	29.32	7.343	117.4
P42	177.3	65.09	31.64	-	-
D43	177.2	55.64	40.02	8.606	117.2
V44	177.7	63.73	33.28	7.654	118.4
A45	177.2	53.96	19.59	8.401	123.2
N46	175.3	53.57	38.46	7.792	113.3
K47	-	55.77	32.41	7.503	120.3
L48	176.1	54.59	43.91	8.382	124.1
A49	175.4	50.42	22.47	9.286	124.6
T50	173.4	62.01	70.1	8.385	116.1
C51	-	56.78	33.26	8.219	131.4
P52	176.4	63.66	31.77	-	-
F53	176	57.19	39.38	8.842	121.6

N54	174.6	52.85	37.96	7.553	119.3
A55	177.8	53.9	18.21	8.717	127.5
R56	177.6	56.76	29.71	8.159	114.3
H57	172.9	53.16	31.28	8.077	119.9
Q58	175.4	54.34	30.79	7.974	124.7
V59	-	57.59	34.25	8.544	119.8
P60	178.2	63.6	32.11	-	-
R61	178.6	60.2	29.63	9.088	126.5
A62	178.4	54.24	18.54	8.936	119.5
E63	177.7	55.47	31.44	8.061	113.6
I64	175.8	62.7	38.19	7.515	122
S65	177.2	62.34	62.02	8.53	118.8
H66	178.4	59.29	29.56	8.033	122.3
H67	177.3	60.43	28.81	8.859	120.2
I68	177.3	66.11	38.33	8.83	117.9
S69	174.9	60.83	64.14	7.389	112.2
S70	173.3	57.76	65	7.302	113.3
C71	176.9	61.71	28.33	7.472	126.8
D72	176.7	56.4	41.13	8.724	129.8
D73	175.2	54.95	41.94	8.755	120.5
K74	176.2	57.32	32.36	7.203	120
S75	174.6	58.29	-	7.399	115.7
S76	174.6	58.61	63.76	8.173	117.5
I77	176.4	61.34	38.79	7.97	121.5
E78	-	56.95	30.14	8.374	124.2
Q79	175.8	56.29	29.64	8.316	120.8
D80	176.6	54.8	41.02	8.381	121.2
V81	176.9	63.11	32.55	7.995	120
V82	176.4	63.33	32.45	8.092	122.7
N83	175.5	53.7	38.78	8.363	121
Q84	176.4	56.46	29.46	8.267	120.5
T85	174.7	62.54	69.72	8.164	114.7
R86	175.9	56.34	30.85	8.205	123
N87	175.2	53.21	38.66	8.443	119.8
L88	178	55.66	42.21	8.256	122.8
G89	174.4	45.47	-	8.458	109.4
Q90	176.2	56	29.5	8.16	119.7
E91	176.8	57.06	30.33	8.594	121.9

T92	174.5	62.24	69.65	8.178	115.5
L93	177.2	55.33	42.27	8.234	124.9
A94	177.8	52.75	19.2	8.278	125
E95	176.7	56.8	30.52	8.287	119.8
S96	174.9	58.58	63.73	8.258	116.3
T97	174.4	62.24	69.63	8.067	115.4
W98	175.8	57.36	29.53	8.015	123
Q99	175	55.32	29.78	7.951	122.4
S100	-	56.52	-	8.124	118.5
P101	-	-	-	-	-
P102	176.9	63.05	32.05	-	-
S103	174.5	58.17	64.05	8.401	116.1
D104	176.4	54.63	41.06	8.353	122.3
E105	176.3	56.82	30.37	8.244	120.2
D106	176.3	54.48	40.96	8.23	120.5
W107	176	57.67	29.5	7.856	121.2
D108	176.2	54	40.61	8.043	121.4
K109	176.8	57.39	32.8	7.805	121.9
D110	176.9	55.24	40.91	8.275	119.9
L111	177.9	56.03	41.96	7.833	121.1
W112	176.6	57.85	29.41	7.908	120.4
E113	176.6	57.19	30.24	8.057	121.2
Q114	176.5	56.28	29.25	8.124	120.1
T115	174.8	62.58	69.68	8.048	114.8

Assignment Table B: **GTSF1 Side-Chain ^{1}H Chemical Shifts.**

Resi	HA	HA2	HB	HB2	HB3
E2	4.358	-	-	2.062	1.91
D3	4.578	-	-	2.634	2.582
T4	4.225	-	4.111	-	-
Y5	4.501	-	-	2.984	2.954
I6	4.008	-	1.752	-	-
D7	4.506	-	-	2.677	2.606
S8	4.344	-	-	3.871	3.8
L9	4.3	-	-	1.618	1.564
D10	4.79	-	-	2.708	2.498

P11	4.373	-	-	2.26	1.946
E12	4.126	-	-	2.102	1.916
K13	4.065	-	-	1.726	-
L14	4.389	-	-	1.627	0.9492
L15	4.369	-	-	0.8918	0.453
Q16	4.316	-	-	2	-
C17	4.655	-	-	-	-
P18	4.471	-	-	2.155	1.457
Y19	4.153	-	-	2.96	1.93
D20	4.378	-	-	1.703	-
K21	4.041	-	-	1.714	1.699
N22	4.693	-	-	2.875	2.641
H23	4.904	-	-	3.716	2.396
Q24	4.886	-	-	1.995	1.681
I25	4.393	-	1.878	-	-
R26	4.255	-	-	1.915	1.664
A27	3.871	-	1.476	-	-
S28	4.204	-	-	4.007	3.857
R29	4.728	-	-	2.27	1.811
F30	4.323	-	-	3.232	3.037
P31	4.03	-	-	2.285	1.55
Y32	4.155	-	-	3.12	3.04
H33	3.995	-	-	3.284	-
L34	3.639	-	-	1.526	1.396
I35	3.47	-	1.765	-	-
K36	3.849	-	-	1.589	1.577
C37	3.926	-	-	2.537	2.092
R38	2.929	-	-	1.684	-
K39	3.853	-	-	1.788	1.77
N40	4.642	-	-	2.674	2.617
H41	5	-	-	3.315	2.547
P42	4.309	-	-	2.216	1.985
D43	4.485	-	-	2.753	2.671
V44	4.11	-	2.228	-	-
A45	4.292	-	1.348	-	-
N46	4.584	-	-	2.866	-
K47	4.402	-	-	1.928	1.823
L48	4.573	-	-	1.681	1.114

A49	4.638	-	1.245	-	-
T50	4.491	-	3.826	-	-
C51	4.186	-	-	-	-
P52	4.286	-	-	2.001	1.365
F53	4.714	-	-	3.119	3.082
N54	4.382	-	-	2.268	1.798
A55	4.056	-	1.342	-	-
R56	4.149	-	-	1.925	1.598
H57	4.781	-	-	3.676	2.293
Q58	5.004	-	-	1.904	1.814
V59	4.699	-	1.977	-	-
P60	4.525	-	-	1.703	-
R61	3.735	-	-	1.816	1.74
A62	4.093	-	1.171	-	-
E63	4.57	-	-	2.549	1.886
I64	4.042	-	1.724	-	-
S65	4.189	-	-	3.892	-
H66	4.35	-	-	3.304	3.255
H67	3.998	-	-	3.353	3.315
I68	3.728	-	2.047	-	-
S69	4.248	-	-	3.938	-
S70	4.53	-	-	3.766	3.571
C71	3.684	-	-	2.649	2.457
D72	4.431	-	-	2.727	2.635
D73	4.729	-	-	2.793	2.543
K74	2.768	-	-	1.296	-
S75	4.372	-	-	3.837	-
S76	4.426	-	-	3.878	-
I77	4.208	-	1.889	-	-
E78	4.221	-	-	2.023	1.923
Q79	4.239	-	-	2.064	1.98
D80	4.571	-	-	2.698	2.621
V81	4.025	-	2.096	-	-
V82	3.962	-	2.051	-	-
N83	4.648	-	-	2.818	2.755
Q84	4.319	-	-	2.133	1.993
T85	4.235	-	-	-	-
R86	4.289	-	-	1.805	1.74

N87	4.675	-	-	2.828	2.708
L88	4.28	-	-	1.654	1.585
G89	-	3.93	-	-	-
Q90	4.295	-	-	2.103	1.953
E91	4.283	-	-	2.038	1.948
T92	4.279	-	4.165	-	-
L93	4.3	-	-	1.631	1.573
A94	4.236	-	1.356	-	-
E95	4.156	-	-	2	1.905
S96	4.407	-	-	3.819	3.761
T97	4.273	-	4.148	-	-
W98	4.616	-	-	3.225	-
Q99	4.216	-	-	1.908	1.744
S100	4.476	-	-	3.787	3.696
P101	4.467	-	-	-	-
P102	4.362	-	-	2.181	1.847
S103	4.395	-	-	3.88	3.796
D104	4.526	-	-	2.637	2.587
E105	4.161	-	-	1.879	1.77
D106	4.563	-	-	2.657	2.551
W107	-	-	-	3.256	3.207
D108	4.439	-	-	2.402	2.37
K109	3.901	-	-	2.239	1.673
D110	4.5	-	-	2.568	-
L111	4.125	-	-	1.444	1.372
W112	4.537	-	-	3.296	3.197
E113	4.087	-	-	1.949	1.865
Q114	4.249	-	-	2.094	1.962
T115	4.219	-	4.191	-	-

Resi	HD1	HD2	HD3	HE1	HE2	HG	HG1	HG2	HG3
E2	-	-	-	-	-	-	-	2.262	-
D3	-	-	-	-	-	-	-	-	-
T4	-	-	-	-	-	-	-	1.108	-
Y5	7.051	-	-	6.766	-	-	-	-	-
I6	0.802	-	-	-	-	-	-	0.8239	-
D7	-	-	-	-	-	-	-	-	-

S8	-	-	-	-	-	-	-	-	-
L9	0.8807	0.8186	-	-	-	-	-	-	-
D10	-	-	-	-	-	-	-	-	-
P11	-	3.782	-	-	-	-	-	1.994	-
E12	-	-	-	-	-	-	-	2.356	2.297
K13	-	1.639	-	-	2.93	-	-	1.392	1.352
L14	0.7868	0.6362	-	-	-	1.577	-	-	-
L15	0.4347	0.5665	-	-	-	1.248	-	-	-
Q16	-	-	-	-	-	-	-	2.274	-
C17	-	-	-	-	-	-	-	-	-
P18	-	4.164	3.901	-	-	-	-	1.72	1.048
Y19	6.907	-	-	6.637	-	-	-	-	-
D20	-	-	-	-	-	-	-	-	-
K21	-	1.746	-	-	2.905	-	-	1.378	1.265
N22	-	-	-	-	-	-	-	-	-
H23	-	6.764	-	7.27	-	-	-	-	-
Q24	-	-	-	-	-	-	-	2.2	2.179
I25	0.2164	-	-	-	-	-	-	1.069	-
R26	-	-	3.137	-	-	-	-	1.766	1.667
A27	-	-	-	-	-	-	-	-	-
S28	-	-	-	-	-	-	-	-	-
R29	-	-	3.144	-	-	-	-	1.776	1.663
F30	6.867	-	-	-	-	-	-	-	-
P31	-	4.034	3.699	-	-	-	-	2.047	2.012
Y32	7.092	-	-	6.861	-	-	-	-	-
H33	-	6.64	-	-	-	-	-	-	-
L34	0.5624	0.6569	-	-	-	1.177	-	-	-
I35	0.8405	-	-	-	-	-	-	0.8196	-
K36	-	1.59	-	-	2.837	-	-	1.298	1.167
C37	-	-	-	-	-	-	-	-	-
R38	-	-	-	-	-	-	-	-	-
K39	-	1.615	-	-	2.911	-	-	1.545	1.354
N40	-	-	-	-	-	-	-	-	-
H41	-	-	-	-	-	-	-	-	-
P42	-	-	3.328	-	-	-	-	-	-
D43	-	-	-	-	-	-	-	-	-
V44	-	-	-	-	-	-	1.109	-	-
A45	-	-	-	-	-	-	-	-	-

N46	-	-	-	-	-	-	-	-	-
K47	-	1.694	-	-	3.026	-	-	1.523	1.45
L48	0.7575	0.5636	-	-	-	1.586	-	-	-
A49	-	-	-	-	-	-	-	-	-
T50	-	-	-	-	-	-	-	1.241	-
C51	-	-	-	-	-	-	-	-	-
P52	-	3.453	-	-	-	-	-	1.561	0.6342
F53	-	-	-	-	-	-	-	-	-
N54	-	-	-	-	-	-	-	-	-
A55	-	-	-	-	-	-	-	-	-
R56	-	3.177	-	-	-	-	-	1.678	1.595
H57	-	6.716	-	7.586	-	-	-	-	-
Q58	-	-	-	-	-	-	-	2.192	2.133
V59	-	-	-	-	-	-	0.6187	0.37	-
P60	-	-	-	-	-	-	-	2.061	2.013
R61	-	3.146	-	-	-	-	-	1.629	1.593
A62	-	-	-	-	-	-	-	-	-
E63	-	-	-	-	-	-	-	2.182	2.125
I64	0.7266	-	-	-	-	-	-	1.059	-
S65	-	-	-	-	-	-	-	-	-
H66	-	-	-	-	-	-	-	-	-
H67	-	-	-	-	-	-	-	-	-
I68	0.8509	-	-	-	-	-	-	1.084	-
S69	-	-	-	-	-	-	-	-	-
S70	-	-	-	-	-	-	-	-	-
C71	-	-	-	-	-	-	-	-	-
D72	-	-	-	-	-	-	-	-	-
D73	-	-	-	-	-	-	-	-	-
K74	-	1.516	1.457	-	2.842	-	-	1.057	0.9262
S75	-	-	-	-	-	-	-	-	-
S76	-	-	-	-	-	-	-	-	-
I77	0.8595	-	-	-	-	-	-	0.9123	-
E78	-	-	-	-	-	-	-	2.269	2.234
Q79	-	-	-	-	-	-	-	2.341	-
D80	-	-	-	-	-	-	-	-	-
V81	-	-	-	-	-	-	0.9196	0.8837	-
V82	-	-	-	-	-	-	0.9273	0.8905	-
N83	-	-	-	-	-	-	-	-	-

Q84	-	-	-	-	-	-	-	2.359	2.277
T85	-	-	-	-	-	-	-	1.183	-
R86	-	3.138	-	-	-	-	-	1.605	1.564
N87	-	-	-	-	-	-	-	-	-
L88	0.8898	0.828	-	-	-	1.59	-	-	-
G89	-	-	-	-	-	-	-	-	-
Q90	-	-	-	-	-	-	-	2.31	-
E91	-	-	-	-	-	-	-	2.36	2.279
T92	-	-	-	-	-	-	-	1.176	-
L93	0.8817	0.8245	-	-	-	-	-	-	-
A94	-	-	-	-	-	-	-	-	-
E95	-	-	-	-	-	-	-	2.262	2.224
S96	-	-	-	-	-	-	-	-	-
T97	-	-	-	-	-	-	-	1.101	-
W98	7.199	-	-	10.08	-	-	-	-	-
Q99	-	-	-	-	-	-	-	2.12	2.073
S100	-	-	-	-	-	-	-	-	-
P101	-	-	-	-	-	-	-	-	-
P102	-	3.599	3.433	-	-	-	-	1.884	1.853
S103	-	-	-	-	-	-	-	-	-
D104	-	-	-	-	-	-	-	-	-
E105	-	-	-	-	-	-	-	2.169	-
D106	-	-	-	-	-	-	-	-	-
W107	7.213	-	-	10.04	-	-	-	-	-
D108	-	-	-	-	-	-	-	-	-
K109	-	1.693	-	-	3.906	-	-	1.304	1.246
D110	-	-	-	-	-	-	-	-	-
L111	0.7909	0.716	-	-	-	-	-	-	-
W112	7.189	-	-	10.07	-	-	-	-	-
E113	-	-	-	-	-	-	-	2.159	2.108
Q114	-	-	-	-	-	-	-	2.325	2.302
T115	-	-	-	-	-	-	-	1.164	-

Assignment Table C: **GTSF1 Side-Chain ^{13}C Chemical Shifts.**

Resi	CD	CD1	CD2	CE	CE1	CG	CG1	CG2	CH2	CZ2
E2	-	-	-	-	-	-	-	-	-	-
D3	-	-	-	-	-	-	-	-	-	-

T4	-	-	-	-	-	-	-	21.67	-	-
Y5	-	133.2	-	-	118.2	-	-	-	-	-
I6	-	12.93	-	-	-	-	27.35	17.42	-	-
D7	-	-	-	-	-	-	-	-	-	-
S8	-	-	-	-	-	-	-	-	-	-
L9	-	24.97	23.47	-	-	26.98	-	-	-	-
D10	-	-	-	-	-	-	-	-	-	-
P11	50.74	-	-	-	-	27.25	-	-	-	-
E12	-	-	-	-	-	36.56	-	-	-	-
K13	29.13	-	-	42.02	-	25.03	-	-	-	-
L14	-	25.34	23.16	-	-	26.85	-	-	-	-
L15	-	25.25	22.93	-	-	27.02	-	-	-	-
Q16	-	-	-	-	-	33.51	-	-	-	-
C17	-	-	-	-	-	-	-	-	-	-
P18	51.31	-	-	-	-	26.04	-	-	-	-
Y19	-	132.5	-	-	117.5	-	-	-	-	-
D20	-	-	-	-	-	-	-	-	-	-
K21	29.33	-	-	42.26	-	24.15	-	-	-	-
N22	-	-	-	-	-	-	-	-	-	-
H23	-	-	116.2	-	139.9	-	-	-	-	-
Q24	-	-	-	-	-	33.67	-	-	-	-
I25	-	9.901	-	-	-	-	26.72	17.52	-	-
R26	43.72	-	-	-	-	27.74	-	-	-	-
A27	-	-	-	-	-	-	-	-	-	-
S28	-	-	-	-	-	-	-	-	-	-
R29	42.73	-	-	-	-	26.74	-	-	-	-
F30	-	131.9	-	-	-	-	-	-	-	-
P31	50.16	-	-	-	-	28.87	-	-	-	-
Y32	-	133	-	-	118.4	-	-	-	-	-
H33	-	-	126.8	-	-	-	-	-	-	-
L34	-	25.42	23.41	-	-	26.18	-	-	-	-
I35	-	13.31	-	-	-	-	28.86	17.13	-	-
K36	29.2	-	-	42.02	-	24.7	-	-	-	-
C37	-	-	-	-	-	-	-	-	-	-
R38	44.44	-	-	-	-	27.04	-	-	-	-
K39	29.2	-	-	42.2	-	25.23	-	-	-	-
N40	-	-	-	-	-	-	-	-	-	-
H41	-	-	-	-	-	-	-	-	-	-

P42	-	-	-	-	-	27.51	-	-	-	-
D43	-	-	-	-	-	-	-	-	-	-
V44	-	-	-	-	-	-	21.32	-	-	-
A45	-	-	-	-	-	-	-	-	-	-
N46	-	-	-	-	-	-	-	-	-	-
K47	-	-	-	-	-	24.77	-	-	-	-
L48	-	25.58	22.34	-	-	26.73	-	-	-	-
A49	-	-	-	-	-	-	-	-	-	-
T50	-	-	-	-	-	-	-	22.15	-	-
C51	-	-	-	-	-	-	-	-	-	-
P52	50.68	-	-	-	-	25.83	-	-	-	-
F53	-	-	-	-	-	-	-	-	-	-
N54	-	-	-	-	-	-	-	-	-	-
A55	-	-	-	-	-	-	-	-	-	-
R56	43.21	-	-	-	-	27.81	-	-	-	-
H57	-	-	115.8	-	140.2	-	-	-	-	-
Q58	-	-	-	-	-	33.98	-	-	-	-
V59	-	-	-	-	-	-	21.22	18.28	-	-
P60	-	-	-	-	-	28.4	-	-	-	-
R61	43.35	-	-	-	-	27.44	-	-	-	-
A62	-	-	-	-	-	-	-	-	-	-
E63	-	-	-	-	-	36.54	-	-	-	-
I64	-	13.05	-	-	-	-	29.16	18.86	-	-
S65	-	-	-	-	-	-	-	-	-	-
H66	-	-	-	-	-	-	-	-	-	-
H67	-	-	-	-	-	-	-	-	-	-
I68	-	14.57	-	-	-	-	30.12	17.6	-	-
S69	-	-	-	-	-	-	-	-	-	-
S70	-	-	-	-	-	-	-	-	-	-
C71	-	-	-	-	-	-	-	-	-	-
D72	-	-	-	-	-	-	-	-	-	-
D73	-	-	-	-	-	-	-	-	-	-
K74	29.36	-	-	41.82	-	23.64	-	-	-	-
S75	-	-	-	-	-	-	-	-	-	-
S76	-	-	-	-	-	-	-	-	-	-
I77	-	13.26	-	-	-	-	27.37	17.69	-	-
E78	-	-	-	-	-	36.33	-	-	-	-
Q79	-	-	-	-	-	33.78	-	-	-	-

D80	-	-	-	-	-	-	-	-	-	-
V81	-	-	-	-	-	-	20.99	-	-	-
V82	-	-	-	-	-	-	21	21.32	-	-
N83	-	-	-	-	-	-	-	-	-	-
Q84	-	-	-	-	-	33.81	-	-	-	-
T85	-	-	-	-	-	-	-	21.9	-	-
R86	43.4	-	-	-	-	27.15	-	-	-	-
N87	-	-	-	-	-	-	-	-	-	-
L88	-	25.14	23.48	-	-	27.06	-	-	-	-
G89	-	-	-	-	-	-	-	-	-	-
Q90	-	-	-	-	-	33.79	-	-	-	-
E91	-	-	-	-	-	36.45	-	-	-	-
T92	-	-	-	-	-	-	-	21.72	-	-
L93	-	24.99	23.72	-	-	27.08	-	-	-	-
A94	-	-	-	-	-	-	-	-	-	-
E95	-	-	-	-	-	36.32	-	-	-	-
S96	-	-	-	-	-	-	-	-	-	-
T97	-	-	-	-	-	-	-	21.52	-	-
W98	-	-	-	-	-	-	-	-	124.5	114.4
Q99	-	-	-	-	-	33.56	-	-	-	-
S100	-	-	-	-	-	-	-	-	-	-
P101	-	-	-	-	-	-	-	-	-	-
P102	50.43	-	-	-	-	27.39	-	-	-	-
S103	-	-	-	-	-	-	-	-	-	-
D104	-	-	-	-	-	-	-	-	-	-
E105	-	-	-	-	-	36.22	-	-	-	-
D106	-	-	-	-	-	-	-	-	-	-
W107	-	-	-	-	-	-	-	-	124.5	114.5
D108	-	-	-	-	-	-	-	-	-	-
K109	29.23	-	-	42.15	-	24.55	-	-	-	-
D110	-	-	-	-	-	-	-	-	-	-
L111	-	24.69	23.66	-	-	26.96	-	-	-	-
W112	-	-	-	-	-	-	-	-	124.5	114.5
E113	-	-	-	-	-	36.27	-	-	-	-
Q114	-	-	-	-	-	33.87	-	-	-	-
T115	-	-	-	-	-	-	-	21.91	-	-

Chapter 4

Development of TROSY Pulse Sequence for Simultaneous Characterization of Backbone Dynamics in Deuterated Proteins

This chapter was published, in part:

O'Brien, P.A., and Palmer, A.G., TROSY pulse sequence for simultaneous measurement of the 15N R1 and {1H}-15N NOE in deuterated proteins. Journal of Biomolecular NMR, 2018. **70** (4): p. 205-209. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature. Copyright 2018.

4.1 Introduction

The ^{15}N R_1 and R_2 relaxation rate constants and the steady-state {^{1}H}-^{15}N nuclear Overhauser enhancement (NOE) provide vital probes of the picosecond-nanosecond dynamics of backbone amide moieties in proteins [9]. Typically, initial measurement of protein dynamics with these three parameters leads to characterization of the spectral density function $J(\omega)$ at the following frequencies: 0, the nitrogen Larmor frequency (ω_N), the proton Larmor frequency (ω_H), and linear combinations of these two Larmor frequencies ($\omega_H \pm \omega_N$). Reduced spectral density mapping is often used to approximate the ^{1}H single quantum and ^{15}N single quantum frequencies as a single effective frequency, thereby reducing the number and type of variables to be fit [74]. Experimental protocols and approaches for data interpretation are well-established; however, particularly as measurements at multiple static magnetic fields become more commonplace in characterization of backbone protein dynamics, maximizing the efficiency of experimental methods remains essential [75-77]. In this chapter, we discuss design, optimization, and validation of a novel NMR pulse sequence for use in measuring the ^{15}N R_1 relaxation rate constant and steady-state {^{1}H}-^{15}N NOE. We conclude with thoughts about its potential for future use with larger biomolecules and protein complexes.

4.2 Individual TROSY Experiments Measuring Backbone Dynamics

4.2.1 The {^{1}H}-^{15}N Heteronuclear NOE Experiment

Measurement of the steady-state {^{1}H}-^{15}N NOE requires two experiments: one in which the ^{15}N steady-state magnetization is measured after irradiation of the ^{1}H spins ("saturated"), and a second in which the Boltzmann equilibrium ^{15}N magnetization is measured in the absence of ^{1}H irradiation ("unsaturated" or "control"). The ratio of the saturated intensity to the unsaturated intensity is the steady-state {^{1}H}-^{15}N NOE. Conceptually, the steady-state {^{1}H}-^{15}N NOE experiment gives information on amide bond motion that is faster than the molecular tumbling time. The {^{1}H}-^{15}N NOE is defined as follows:

$$NOE = 1 + \left(\frac{d^2}{4}\right)\left(\frac{\gamma_H}{\gamma_N}\right)[5J(0.87\omega_H)](T_1)$$

Since the gyromagnetic ratio (γ_N) of nitrogen-15 is negative, the second term in the above equation is negative and steady-state {^{1}H}-^{15}N NOE will be always be less than one. Experimentally, this means the peak intensity from a saturated {^{1}H}-^{15}N NOE spectrum will always be less than the intensity from the accompanying unsaturated spectrum. N- and C-terminal protein residues, as well as residues that do not engage in secondary structure, will exhibit $\{^{1}H\}$-^{15}N NOE ratios anywhere from 0.5 to -1.0.

Although seemingly straightforward, the steady-state $\{^{1}H\}$-^{15}N NOE experiment suffers from several factors complicating easy implementation. Chief among these is the inherent lower experimental sensitivity. In both the unsaturated and saturated heteronuclear NOE experiments, magnetization originates on the low gyromagnetic ^{15}N nuclei, instead of the higher sensitivity ^{1}H, as in the R_1 and R_2 relaxation experiments. In addition, the recycle delay between transients must be set longer than is typical for most NMR experiments. To ensure complete recovery of equilibrium magnetization in the unsaturated control experiment, this delay should be set > 7-8/R_1 (^{15}N). Finally, avoidance of saturation transfer through chemical exchange of labile protons with protons from the water "sink" requires careful manipulation of the solvent magnetization. Addressing these concerns to ensure accurate steady-state $\{^{1}H\}$-^{15}N NOE values results in very long experimental times, especially as replicate experiments are required for precise estimates of experimental error.

4.2.2 The Inversion Recovery Experiment for Measurement of the R_1 Relaxation Rate

Measurement of the ^{15}N R_1 longitudinal relaxation rate requires the recording of $^{15}N_z$ intensity decay at increasing delay values. Referred to as an inversion recovery experiment, the pulse sequence starts with either Boltzmann N_z magnetization or, as is more commonplace, polarization transfer-enhanced $^{15}N_z$ magnetization. A *z*-filter is used for isolation of "pure" N_z magnetization through use of sequential orthogonal pulses or shaped pulses. Proton 180° inversion pulses are applied and the decay of N_z magnetization is recorded at increasing time points. The ^{15}N R_1 is the exponential rate that determines the reduced intensity of N_z magnetization at longer time points. The ^{15}N R_1 is defined as:

$$R_1 = \left(\frac{d^2}{4}\right)[3J(\omega_N) + 7J(0.921\omega_H)] + c^2J(\omega_N)$$

Careful handling of the water magnetization is also a significant concern when measuring the ^{15}N R_1 relaxation rate. Pre-saturation of the water resonance and hard ^{1}H 180° pulses are typically avoided in favor of shaped pulses that selectively invert the amide proton region. Any attenuation of the water signal will lead to artificially lower peak intensities at longer relaxation delays, which results in systematically overestimated R_1 rates. Even with implementation of shaped 180° pulses, incomplete inversion of water and partial water signal saturation can still occur. The former issue is potentially unavoidable because moving the resonance frequencies of amide ^{1}H-^{15}N correlations close to the water peak is an intractable problem. On the other hand, the latter issue is dealt with by increasing the recycle delay between transients.

4.3 Design of the Combined NOE-R_1 Pulse Sequence

Initial characterization of backbone protein dynamics relied on natural abundance ^{13}C or isotopically labeled ^{15}N samples. Subsequent advances in deuterium labeling of recombinantly expressed biomolecules, though, led to the widespread implementation of transverse relaxation-optimized spectroscopy (TROSY), which confers enhanced spectral sensitivity and resolution. Both multidimensional NMR experiments (for chemical shift assignment) and spin relaxation experiments have benefited from this advancement. A recent set of TROSY-based backbone spin relaxation experiments are the standard for measurement of the ^{15}N R_1 relaxation rate, the ^{15}N R_2 relaxation rate, and the steady-state {^{1}H}-^{15}N NOE.

In the TROSY-based {^{1}H}-^{15}N NOE pulse sequence [78], the unsaturated experiment shows a unique feature: following the t_1 labeling period, a reverse-INEPT block transfers transverse ^{15}N magnetization to ^{1}H for subsequent detection, and simultaneously transfers longitudinal ^{1}H magnetization to longitudinal ^{15}N magnetization [79]. In principle, this polarization-enhanced ^{15}N magnetization could serve as the starting point for another relaxation experiment following a detection period. In the present work, a second pulse sequence block was designed to use this "free" ^{15}N magnetization to measure the ^{15}N R_1 relaxation rate constant. Separate storage of decay curves originating with positive or negative values of the initial longitudinal ^{15}N magnetization also allows the steady-state ^{15}N intensity to be obtained by extrapolation during data analysis; hence, the experiment allows calculation of the {^{1}H}-^{15}N NOE. Thus, an independent measurement of the saturated {^{1}H}-^{15}N NOE spectrum is not necessary.

The pulse sequence for the proposed experiment is shown in Figure 4.1 and is based on the TROSY pulse sequences of Bax and coworkers [57]. The first part of the experiment is the control measurement of equilibrium ^{15}N magnetization. At the same time, this sequence transfers longitudinal ^{1}H magnetization to longitudinal ^{15}N magnetization at the start of the first t_2 acquisition period. Following the t_2 period (100 ms acquisition period), two orthogonal ^{1}H 90° hard pulses are used to purge remaining ^{1}H magnetization and guarantee only ^{15}N$_z$ magnetization is available for the remainder of the experiment. A composite 0° or 180° ^{15}N pulse is then used to generate positive or negative ^{15}N longitudinal magnetization. The inversion-recovery relaxation period consists of repeated [Δ - 180°(^{1}H)- Δ] blocks with Δ = 11 ms [80]. The ^{1}H 180° pulses are cosine-modulated I-BURP2 pulses to maintain water magnetization on the +z-axis [81], while maximally affecting the amide proton region. These ^{1}H 180° pulses serve two purposes: (i) to remove cross-correlated relaxation effects and (ii) to obtain ^{15}N steady-state magnetization by amide proton saturation. Following the relaxation delay, the partially relaxed ^{15}N magnetization is frequency labeled (t_1) and transferred to ^{1}H for detection during the second t_2 period. The experiment is performed twice for each R_1 relaxation delay: the unsaturated (control) ^{15}N free induction decays (FID) are co-added while the R_1 FIDs with positive initial ^{15}N longitudinal magnetization ($I_+(t)$), or negative initial ^{15}N longitudinal magnetization ($I_-(t)$) are stored separately. If the inversion-recovery experiment is performed with N relaxation time points and M scans per complex point in t_1 for each relaxation time point, then the final data unsaturated measurement consists of $2MN$ co-added FIDs per complex point in t_1. For each amino acid residue, the inversion-recovery curves are fit simultaneously to the equations:

$$I_+(T) = I_{ss} + (I_+(0) - I_{ss})e^{-R_1T}$$

$$I_-(T) = I_{ss} + (-I_-(0) - I_{ss})e^{-R_1T}$$

in which I_{ss} and R_1 correspond to the steady-state ^{15}N magnetization intensity and the longitudinal relaxation rate constant, respectively. The initial magnetizations are treated as local variables, while I_{ss} and R_1 are treated as global parameters. The NOE is calculated as the ratio of I_{ss} and the average intensity of the unsaturated experiment, I_0.

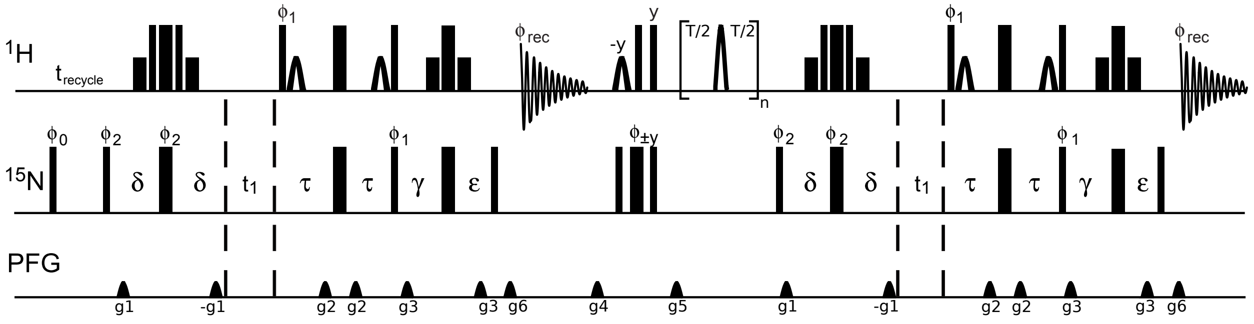


Figure 4.1: **Novel NOE-R_1 Pulse Sequence.**
Pulse sequence for simultaneous measurement of the steady state $\{^1H\}$-^{15}N NOE and the ^{15}N R_1 relaxation rate constant. Thin and thick solid bars represent high-power 90° and 180° pulses, respectively; short open shapes represent water-selective 90° sinc pulses; short thin solid bars represent rectangular water-selective 90° pulses; thick solid bars flanked by two thin bars indicate a 90_x-210_y-90_x pulse scheme for selection of the anti-TROSY component; and the open shaped pulse during *T* is a cosine-modulated I-BURP2 180° rectangular pulse crafted to leave the water magnetization unperturbed (T = 2 × Δ). Gradients were applied as square pulses and are used for artifact suppression. Delays: $t_{recycle}$ = 11s, $\delta = \tau = 2.65$ ms, $\gamma + \varepsilon = 2.91$ ms. Gradients: G1 (600μs, 2.17 G/cm), G2 (300μs, 0.68 G/cm), G3 (1000μs, 3.29 G/cm), G4 (500μs, 0.68 G/cm), G5 (1000μs, 1.32 G/cm), G6 (121.60μs, 2.17 G/cm). All pulses are *x*-phase unless indicated otherwise. Phase cycling: ϕ_2 = y, -y, and ϕ_{rev} = y, -y. [82]

4.4 Validation of the Combined NOE-R_1 Experiment using [^2H,^{15}N] UBQ

The proposed pulse sequence was used to measure ^{15}N R_1 and steady-state $\{^1H\}$-^{15}N NOE for a 1.0 mM sample of [U-^2H,^{15}N] ubiquitin. This system was chosen for experimental validation because of the vast amount of previously measured ubiquitin relaxation rates, as well as the lack of chemical exchange effects, which might hamper initial development of this pulse sequence. As shown in Figure 4.2, previous measurements of ubiquitin $\{^1H\}$-^{15}N NOE values show clustering around 0.7, which is typical of amides bonds located in elements of protein secondary structure. Sample temperature was calibrated to 298 K using 98% $^2H_4$-methanol [56]. Relaxation data were collected at 14.1 T on a Bruker DRX600 console equipped with a triple-resonance z-axis gradient cryogenic probe. Spectra were recorded using 4 scans per complex t_1 point, and $t_2 \times t_1$ of 512 × 128 complex points. The spectral width was set to 9.3 kHz × 2.4 kHz. Relaxation delays were set to $T = n \times 22$ ms, where $n = \{2, 4, 10, 18, 44, 88, 240, 464\}$, giving $T = \{0.044, 0.220, 0.396, 0.968, 1.936, 5.280, 10.208\}$ s. Interleaved spectra were separated and processed using NMRPipe [83], and in-house Python scripts were used to fit relaxation curves for calculation of R_1 and I_{ss} values. The proposed pulse sequence was validated by comparison with results from independent conventional measurements of ^{15}N R_1 and steady-state $\{^1\text{H}\}$-^{15}N NOE using TROSY-based pulse sequences[57]. The conventional NOE experiment used 16 scans per complex t_1 point for the control and the saturated spectra; hard 180° ^{1}H pulses were used to saturate amide proton magnetization. The conventional R_1 experiment used a recycle delay of 3.5 s, 8 scans per complex t_1 point, and 8 relaxation delays spaced between 0 and 1.6 s. Total acquisition times were 32, 13, and 20 hours for the proposed experiment, conventional NOE experiment, and conventional R_1 experiment, respectively. Experimental uncertainties in relaxation parameters are the standard error in the mean of three replicates.

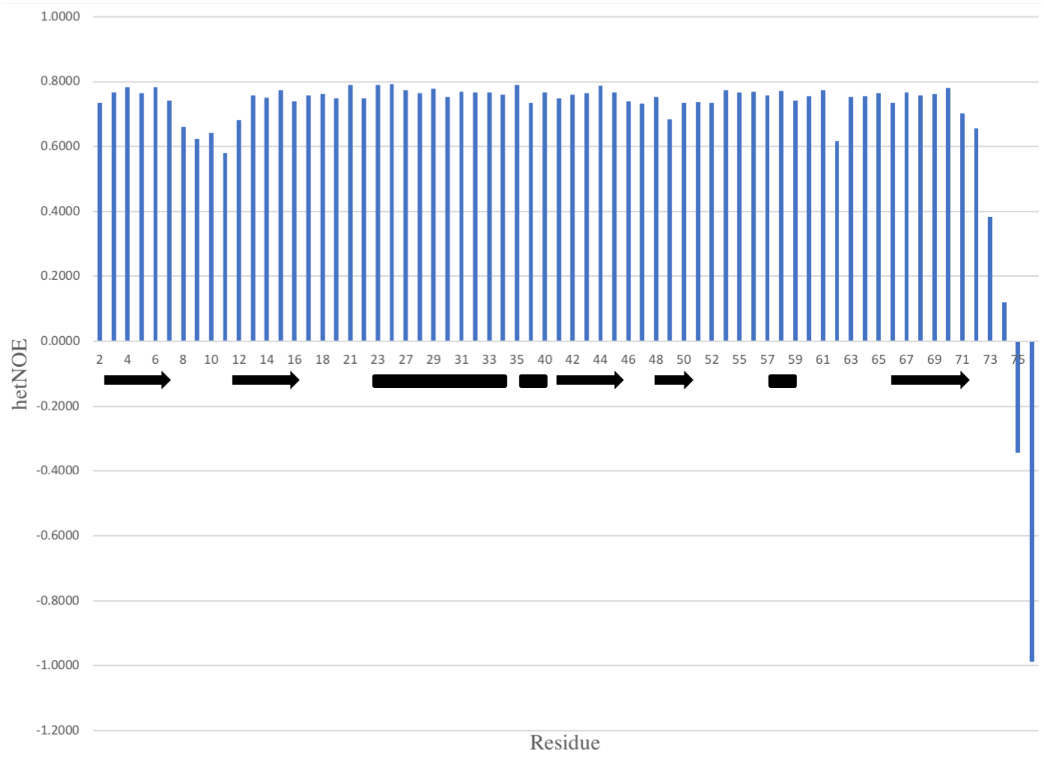


Figure 4.2: **Steady-State {^{1}H}-^{15}N NOE Values for Ubiquitin.**
As an example, this plot shows past measurements of hetNOE ratios for ubiquitin. For elements of protein secondary structure, the average heteronuclear NOE ratio falls between 0.7 and 0.75. Regions of flexibility in proteins, specifically turns or regions of random coil, lead to decreased values of the hetNOE (0.5-0.6). Very dynamic N- and C-terminal protein regions can even display negative hetNOEs, due to the negative sign of the ^{15}N gyromagnetic ratio.

Figure 4.3 shows representative R_1 relaxation decay curves for Val 5, Thr 9, Ile 30, and Gly 75, which are located in α-sheet, turn, β-helix, and C-terminal regions of ubiquitin, respectively. The absolute value of the initial magnetization for each set of the R_1 curves differed by approximately 2% intensity, with the $I_+(T)$ curve yielding larger intensity values compared to $I_-(T)$. As shown in Figure 4.4, the R_1 and NOE values from the combined experiment are in excellent agreement with reference TROSY-detected experiments for ubiquitin. Comparison of ^{15}N R_1 values between the two experiments gives a correlation coefficient of 0.974 and an RMSD of 0.032, while comparison of steady-state {^{1}H}-^{15}N NOE values gives a correlation coefficient of

0.998 and an RMSD of 0.016. Corrected for differences in acquisition times, the NOE values for the conventional and combined experiments have very similar root-mean-square uncertainties of 0.00925 and 0.0102, respectively. Figure 4.5 shows the residues used in validation of our pulse sequence mapped onto the structure of ubiquitin.

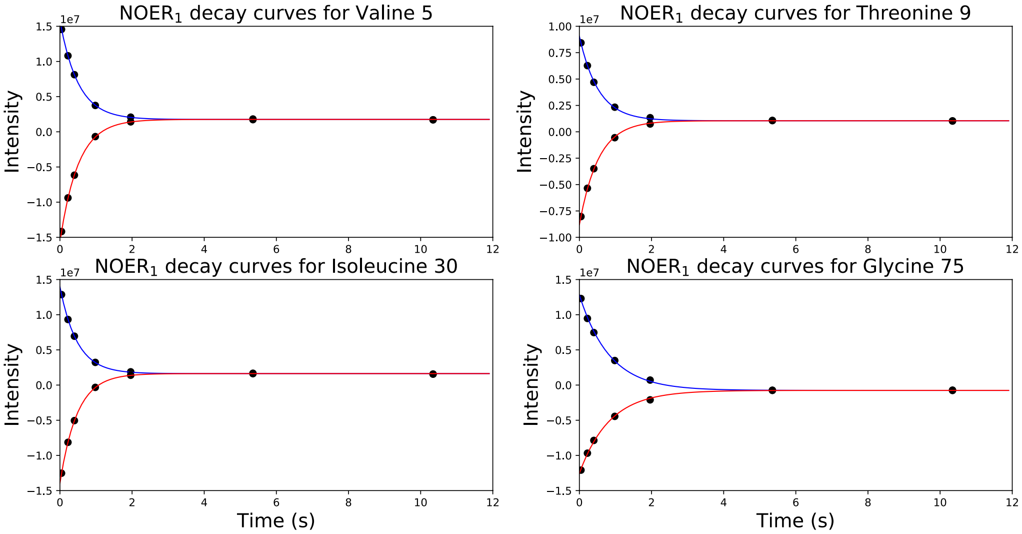


Figure 4.3: **^{15}N R_1 Decay Curves Recorded with the NOE-R_1 Pulse Sequence.**
^{15}N R_1 decay curves for Val 5, Thr 9, Ile 30, and Gly 75 in ubiquitin. Data points and fitted decay curves are shown for (blue) positive ($I_+(T)$) and (red) negative ($I_-(T)$) initial ^{15}N magnetization. Data were fit to Eq. 1 using a Levenberg-Marquardt algorithm. [82]

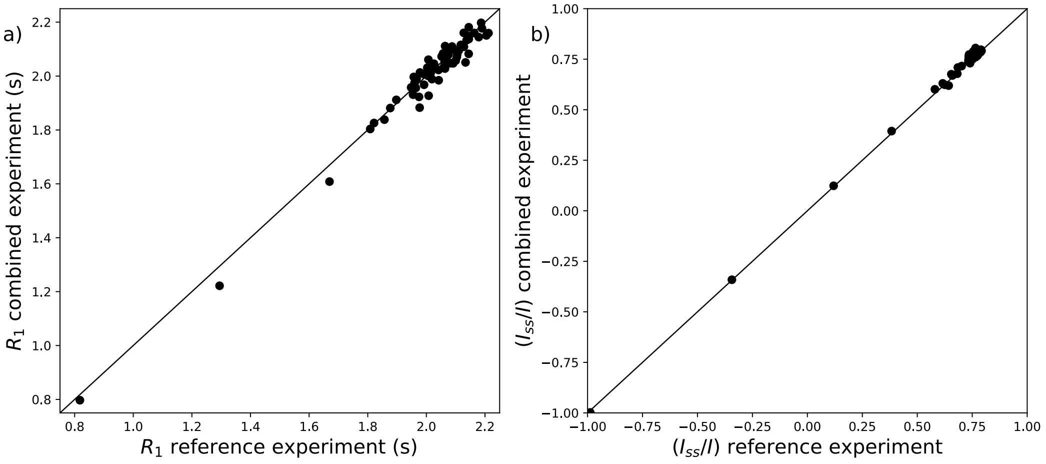


Figure 4.4: **Comparison of NOE-R_1 and Reference Measurements of a) ^{15}N R_1 relaxation rate (R^2 = 0.974, RMSD = 0.032) and b) steady state {^{1}H}-^{15}N NOE (R^2 = 0.997, RMSD = 0.020).** [82]

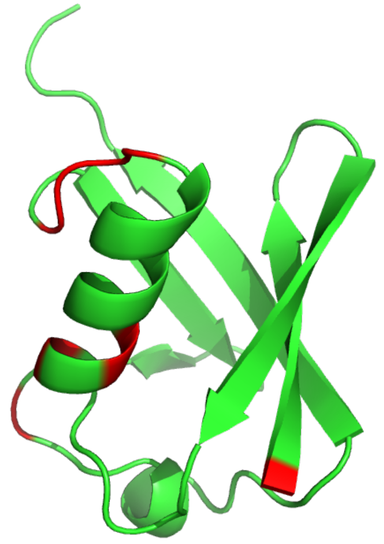


Figure 4.5: **Ubiquitin with Residues Included in NOE-R_1 Validation Labeled in Green.**

4.5 Discussion

As displayed in Figure 4.5, the relaxation rates extracted from our combined NOE-R_1 experiment compare favorably with those measured individually in reference experiments. However, we note that the proposed combined experiment differs from separately acquired NOE and R_1 datasets in three ways: (i) The resonance intensities for the "saturated" spectrum, needed to calculate the NOE, are not directly measured, but are extrapolated during the fitting of the positive and negative R_1 relaxation curves; (ii) longitudinal ^{1}H magnetization is inverted for a period $\tau + t_1$ prior to the TROSY transfer period; and (iii) non-equilibrium longitudinal ^{15}N magnetization relaxes during the first acquisition t_2 period prior to the start of the R_1 relaxation delay. Monte Carlo simulations of these effects were simulated for a protein with tumbling time 5 ns (similar to the τ_c of ubiquitin) and 15 ns (characteristic of a 40 kDa protein), using a 600 MHz spectrometer (full protonation was assumed for simplicity, which should give an upper bound on sensitivity differences). Simulations assumed that the conventional NOE experiments were recorded with 32 scans per FID for each of the control and saturated spectra. The combined experiment used 64 scans to measure the control spectrum and simultaneously used N = 8 time points to measure R_1, with 4 scans for positive initial ^{15}N magnetization and 4 scans for negative initial magnetization. The simulations indicate that the two different NOE measurements have virtually identical sensitivity per unit time for the 5 ns and 15 ns rotational diffusion correlation times. The R_1 measurement has a lower precision by 15% and ~10% compared with a conventional experiment measured with the same number of scans and time points, for 5 ns and 15 ns tumbling times, respectively. This difference arises from the small relaxation losses noted above. The precision of the R_1 experiment is ~1%, so these small differences are not a major effect on subsequent data analysis. Notably the simulations show that the values of the NOE and R_1 for the combined experiment are uncorrelated, with $R^2 < 0.01$. The theoretical calculations are in approximate agreement with the experimental results reported for ubiquitin. The combined experiment yields both the NOE and R_1; thus, R_1 is obtained essentially for free. The time savings afforded by the combined experiment depends on the relative lengths of the conventional NOE and R_1 experiments. For example, savings in experimental time would be 20 or 33% if the conventional R_1 experiment were to be recorded in 1/4 or 1/2 of the experimental time for the conventional NOE measurement. As the protein system under investigation increases in rotational correlation time, and hence in molecular weight, the R_1 relaxation rate is expected to change as well. A decreased R_1 in larger molecular weight systems leads to longer relaxation delays to properly measure the rate.

We have described a TROSY pulse sequence for simultaneous measurement of the intensity of the Boltzmann equilibrium ^{15}N magnetization and the ^{15}N longitudinal relaxation rate constant R_1 in a single experiment. Long-time plateau values from fitting of the R_1 data serve as proxies for the usual saturated $\{^{1}H\}$-^{15}N intensity spectrum to allow calculation of the steady-state $\{^{1}H\}$-^{15}N NOE. Measurements of ^{15}N R_1 and steady-state $\{^{1}H\}$-^{15}N NOE on a ubiquitin sample using both the combined experiment and conventional independent experiments are in very good agreement. The greater efficiency of the proposed pulse sequence should become even more useful for larger biomolecular systems, for which maximizing sensitivity per unit time is critical.

Chapter 5

Enhanced Spectral Density Mapping through Combined Multi-field Deuterium $^{13}CH_2D$ Methyl Spin Relaxation NMR Spectroscopy

This chapter was published, in part:

Hsu, A., et al., Enhanced spectral density mapping through combined multiple-field deuterium 13CH2D methyl spin relaxation NMR spectroscopy. Methods, 2018. **138-139**: p. 76-84. Reprinted by permission from Elsevier.

5.1 Introduction

Intramolecular dynamics of proteins increasingly are recognized as playing important roles in biological function, including allostery, catalysis, molecular recognition, regulation, and signaling [84-86]. Spin relaxation in NMR spectroscopy is a powerful probe of dynamics and kinetics of biological macromolecules with atomic resolution and over a broad range of time scales [9, 87]. In particular, NMR spin relaxation studies have used the dynamics of amino acid side-chains to determine the contribution to overall entropy from protein conformational entropy (ΔS_{conf}) in several protein-ligand binding systems [88]. For high-affinity interactions, results have demonstrated different magnitudes and favorability for ΔS_{conf} depending on the system studied. Notably, this work has highlighted the utility of deuterium (^{2}H or D) relaxation, which offers several advantages compared to other NMR-active nuclei in characterizing the dynamic properties of molecules in solution. First, ^{2}H relaxation is dominated by the quadrupolar mechanism and contributions from other mechanisms such as chemical shift anisotropy (CSA) and chemical exchange are negligible in comparison. Second, the quadrupolar Hamiltonian is axially symmetric and the principal axis is oriented along the carbon-deuterium bond, which simplifies interpretation of relaxation measurements. In contrast, CSA tensors may be rhombic and have principal axes systems that are rotated relative to convenient reference frames. Together, these make NMR deuterium relaxation experiments a very powerful technique for study of pico- to nanosecond side-chain dynamics in biomolecules.

5.2 Motivation

Kay and coworkers established methods for measuring relaxation rates for ^{2}H spins in CH_2D methyl group isotopomers in proteins in solution [89, 90]. These methods subsequently were extended to CD, CHD, and CHD_2 moieties in proteins [91, 92] and CD and CHD moieties in RNA [93, 94]. R_1, R_{1Q}, R_Q, R_{AP}, and R_{DQ} constitute five quantifiable ^{2}H relaxation rate constants for longitudinal magnetization (D_z), transverse magnetization (D^+), quadrupolar order ($3D_z^2 - 2$), antiphase coherence ($D^+D_z + D_zD^+$), and double-quantum coherence (D^{+2}), respectively. At a given static magnetic field strength, relaxation of the density operator is dependent on the values of the spectral density function, $J(\omega)$, at three frequencies: 0, ω_D, and $2\omega_D$, in which ω_D is the ^{2}H Larmor frequency. Because the five relaxation rate constants measured for the deuterium spin are defined by the spectral density functions of three deuterium frequencies, the spectral density values can be determined if three or more relaxation rate constants are measured. Inversion of the relaxation equations to obtain the values of the spectral density function is called spectral density mapping [74, 95].

The present work examines three approaches for spectral density mapping when relaxation data have been acquired at more than one static magnetic field strength. In the simplest approach, termed the *independent method*, each static magnetic field is treated independently [96]. If data sets are obtained using two NMR spectrometers whose static magnetic field differ by a factor of two, then the double-quantum frequency, $2\omega_D$, at the lower field is equal to the single quantum frequency at the higher field [97]. In this case, the relaxation data recorded at the two static magnetic fields can be analyzed simultaneously to obtain $J(\omega)$ at frequencies of $0, \omega_D, 2\omega_D, 4\omega_D$, in which ω_D is the Larmor frequency at the lower static magnetic field. This approach is termed the *joint method*. To overcome the limited possible pairs of static magnetic fields available with existing NMR spectrometers, we developed a strategy that interpolates the data sets for any two static fields to obtain a data set that can be paired with a third data set. In this paper, we demonstrate this combined field-pairing and interpolated analysis, termed the *joint-interpolated method*, by obtaining relaxation rate data sets using a 400 MHz/ 800 MHz spectrometer pair and a 450 MHz/ 900 MHz spectrometer pair and analyzing them globally. In this scenario, the 450 MHz data set was interpolated from data sets originating from the 400 MHz and 500 MHz spectrometer pair. In total, this yields spectral density values at seven frequencies, $\omega/2\pi$: 0, 61.4, 69.1, 122.8, 138.2, 245.6, and 276.4 MHz with increased statistical degrees of freedom for model evaluation.

The different approaches for spectral density mapping were tested using relaxation data acquired for an *E. coli* ribonuclease H (RNase H) U-[^{2}H, ^{15}N] ILV (Ile δ2, Ile γ2, Leu δ1, Leu δ2, Val γ1, Val γ2) [^{13}CH$_2$D] sample. RNase H (EC 3.1.26.4) is a conserved endonuclease responsible for cleaving the RNA strand of DNA/RNA hybrids in various biological processes, including reverse transcription of the viral genome in retroviral reverse transcriptases and Okazaki fragment processing during the DNA replication of the lagging strand [98]. RNase H also belongs to a broader superfamily of nucleotidyl-transferases with conserved structure and mechanism, including retroviral integrases, Holliday junction resolvases, and transposases [99, 100]. RNase H has been the subject of many investigations of folding, structure, and dynamics [101-103].

5.3 Methods

5.3.1 Preparation of Selectively ^{2}H-labeled RNase H NMR Sample

Samples of *E. coli* RNase H [U-^{2}H, ^{15}N] ILV-[^{13}CH$_2$D] were prepared using mostly standard overexpression and purification schemes, with only several modifications needed for compliance with ^{2}H relaxation experiments. BL21(DE3) cells were transformed with a pAED4 plasmid encoding a cysteine-free version of *E. coli* ribonuclease HI gene and selected for using an ampicillin resistance marker. Optimal isotope labeling for ^{2}H relaxation experiments includes a uniformly deuterated background at protein aliphatic, aromatic, and amide sites. Thus, protein expression in 99% D_2O media is required. Bacterial cultures were initially grown in M9 minimal media containing 100% H_2O, before performing a series of adaptation steps: at each stage, an increasingly higher composition D_2O based M9 minimal media was inoculated with culture from the previous stage to give an OD_{600} ~ 0.08-0.1, which corresponds to late lag or early exponential growth phase for *E. coli*. At a 10x smaller volume compared to our final expression media, cultures were adapted to 25%, 50%, 75%, and 99% D_2O-based M9 minimal media. The final 99% D_2O-based M9 minimal expression media, which included ^{15}N-ammonium chloride and ^{2}H$_7$-glucose, was grown at 37°C to $OD_{600} = 0.7$ before induction with 0.5 mM IPTG. Selective ^{13}CH$_2$D group labeling was achieved by supplementing the minimal growth media with 50 mg/mL of 2-

ketobutyric acid-4-^{13}C, 4-d$_1$(99% ^{13}C, 97% ^{2}H, Isotec, Sigma Aldrich) and 80 mg/mL of 2-keto-3-(methyl-^{13}C, d$_1$)-butyric acid-3,4,4,4-d$_4$ (99% ^{13}C, 98% ^{2}H, Isotec, Sigma Aldrich) an hour before induction [104, 105]. Note that position three in 2-ketobutyric acid-4-^{13}C, 4-d$_1$ (α-ketobutyric acid) contains two protons (-CH$_2$), which is not ideal and contaminates ^{2}H relaxation rates obtained from corresponding NMR experiments. To mitigate this, α-ketobutyric acid precursor was incubated at pH 10.5 in D$_2$O for approximately 12-14 hours. This allows quantitative exchange of the 3-H protons for deuterons. α-ketobutyric acid was then ready for inclusion in D$_2$O-based M9 minimal expression media. Protein expression was allowed to proceed for 4 hours, again at 37° C.

RNase H was purified as described previously [106-108]. Bacterial pellets were lysed into 20 mM sodium phosphate (pH 7.0), 50 mM NaCl, 10 mM MgCl$_2$, DNase, RNase, and cOmplete protease inhibitor (Roche). Lysate was sonicated and cleared in a centrifuge at 14K rpm for 45 minutes. The cleared supernatant was loaded on a tandem, pre-equilibrated Q-HP plus Heparin-HP HiTrap column (GE Life Sciences). The Q-HP portion of the column binds assorted cellular proteins; RNase H specifically binds to the Heparin-HP column. After detaching it from the Q-HP column, the Heparin-HP column was washed with five column volumes of Buffer A, 20 mM sodium phosphate (pH 7.0), and eluted with a linear gradient from 0% to 100% of 20 mM sodium phosphate (pH 7.0) and 1 M NaCl (Buffer B). Heparin-HP elution volumes are analyzed by SDS-PAGE, and pooled elution volumes containing RNase H were buffer exchanged with either a dialysis bag or spin concentrator into 50 mM sodium acetate (pH 5.5) with 50 mM NaCl. The sample was then loaded on a SP-HP column (GE Life Sciences) and five column volumes of Buffer A (50 mM sodium acetate, pH 5.5, and 50 mM NaCl) were used to wash the column. RNase H was then eluted with a linear gradient from 0% to 100% of 50 mM sodium acetate (pH 5.5) and 1 M NaCl (Buffer B). SDS-PAGE analysis identified the fraction(s) containing RNase H, and a UV-

VIS spectrophotometer was used to measure the final concentration of RNase H (A_{280} 2.29 g/L). Depending on the sample purity observed from the SDS-PAGE gel, a Superdex 200 10/300 GL (GE Life Sciences) was occasionally used for further purification of RNase H. The low salt SP-HP column buffer served as the size-exclusion buffer, and a flowrate of 0.2 mL/min was used. In all instances where size-exclusion chromatography was used as a last purification step, a single monodisperse was observed in the chromatogram and a final SDS-PAGE gel showed no visible band besides that corresponding to RNase H.

5.3.2 ^{2}H NMR Spectroscopy

NMR samples contained 500 μM RNase H, 100 mM $^2H_3$-sodium acetate, pH = 5.5, and 99% ^2H_2O in a 5 mm Shigemi NMR tube. Relaxation experiments were acquired at 9.4 T (400 MHz for ^{1}H spins) on a Varian Inova spectrometer with a room temperature probe and at 11.75 T (500 MHz), 16.45 T (700 MHz), 18.8 T (800 MHz), and 21.1 T (900 MHz) on Bruker *AVANCE* spectrometers with triple resonance z-axis gradient CryoProbes. The static magnetic field strengths correspond to ^{2}H Larmor frequencies of $\omega_D/2\pi$ = 61.4, 76.8, 107.4, 122.8, and 138.2 MHz. Calibration of sample temperature to 298 K was accomplished using 98% $^2H_4$-methanol [56] and small variations between spectrometers adjusted by matching the chemical shifts. The relaxation rate constants R_1, $R_{1\rho}$, R_Q, and R_{AP} were measured at five static B_o fields using a modified pulse sequence with non-constant time ^{13}C chemical shift evolution in the indirect dimension. Initial implementations of ^{2}H relaxation pulse sequences used a constant time approach [89] for ^{13}C chemical shift evolution because of non-specific ^{13}C isotope labeling of the target protein. However, use of a-keto acid precursors restricted ^{13}C labeling to the terminal methyl groups in Ile, Leu, and Val. Thus, no carbon-carbon scalar couplings were active during 13C chemical shift evolution since all aliphatic side-chain carbons were ^{12}C labeled. The double-quantum relaxation rate constant requires correction for interactions with remote spins and was not utilized in the present work. For all four ^{2}H relaxation experiments, spectra were recorded with 32 scans per t_1 increment and 1024 × 200 complex points for $t_2 \times t_1$. Spectral width for all relaxation experiments was 12 ppm × 20 ppm ($t_2 \times t_1$). The carriers for ^{1}H and ^{13}C were set to 4.70 ppm and 15.0 ppm, respectively. Relaxation delays were set to {1 ms, 10 ms, 20 ms, 30 ms, 40 ms, 50 ms} for measurement of ^{2}H R_1 and R_Q, {1 ms, 4 ms, 8 ms, 12 ms, 16 ms, 20 ms} for measurement of ^{2}H R_{AP}, and {0.5 ms, 3 ms, 5 ms, 10 ms, 15 ms, 20 ms} for measurement of ^{2}H $R_{1\rho}$. Two or three duplicate measurements within each relaxation delay series were recorded for purposes of error estimation.

5.3.3 Spectral Processing, Assignments of RNase H ILV Methyl Correlations, and Calculation of Relaxation Parameters

NMR spectra were processed using NMRPipe [83]. Identification of ecRNH methyl correlations from a ^{1}H-^{13}C constant-time HSQC experiment utilized previously assigned *E. coli* ribonuclease HI ^{1}H-^{13}C chemical shift assignments [109]. In all, 48 methyl peaks were observed: 7 Ile, 20 Leu, and 21 Val (Figure 5.1). An equivalent number of peaks was also observed for the shortest relaxation delay in each of the four ^{2}H relaxation experiments. Spectra were visualized in Sparky [110] and peak heights determined at assigned peak positions. The unweighted peak intensities were fit to mono-exponential decay functions using the trust-region non-linear least squares algorithm implemented in the MATLAB (version R2016b) exponential library model ‘exp1’. Uncertainties in fitted relaxation rate constants were determined by Monte Carlo simulations, using estimates of Gaussian noise obtained from variation in duplicate measurements, and by jackknife simulations. The larger of the two error estimates was used as the experimental uncertainty for subsequent spectral density mapping.

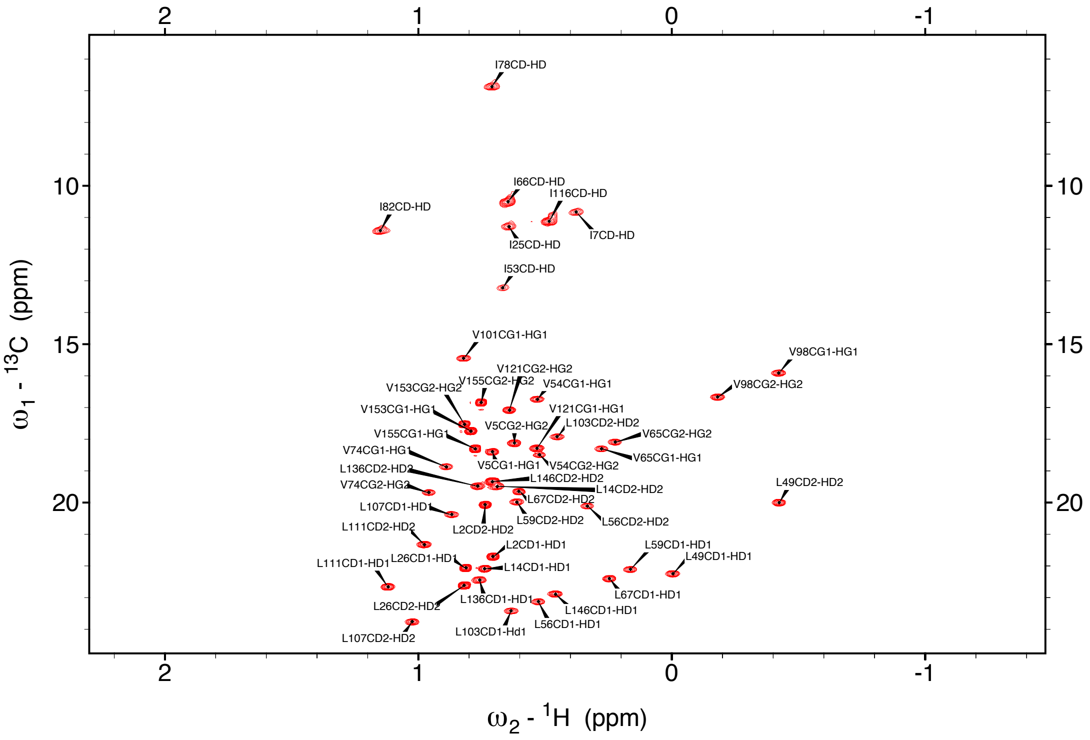


Figure 5.1: **^{1}H-^{13}C HSQC of *E. coli* RNase H [U-^{2}H, ^{15}N] ILV-[^{13}CH$_2$D].**
A constant-time HSQC spectrum allowed high resolution of ^{1}H-^{13}C correlations in the methyl region of *E. coli* RNase H. Using previous *E. coli* RNH ILV methyl assignments, proper referencing of the spectrum and transfer of chemical shifts resulted in assignment of 48 out of 49 possible peaks. L14Cδ2-Hδ2, V101Cγ2-Hγ2, and L14Cδ2-Hδ2 peaks are barely resolved in this high-resolution HSQC, and thus were not used during analysis of relaxation data.

5.4 Theory

5.4.1 Deuterium Spectral Density Mapping

The four ^{2}H spin relaxation rate constants used herein are given by:

$$R_1 = 3\xi^2[J(\omega_D) + 4J(2\omega_D)] \tag{1}$$

$$R_{1\rho} = (\frac{3\xi^2}{2})[3J(0) + 5J(\omega_D) + 2J(2\omega_D)] \tag{2}$$

$$R_Q = 9\xi^2[J(\omega_D)] \tag{3}$$

$$R_{AP} = (\frac{3\xi^2}{2})[3J(0) + J(\omega_D) + 2J(2\omega_D)] \tag{4}$$

in which the quadrupolar coupling constant is:

$$\xi = \frac{\pi e^2 qQ}{2h} \tag{5}$$

e is the charge on the electron, *eq* is the principal value of the electric field gradient tensor, *Q* is the nuclear quadrupole moment, and *h* is Planck's constant. *J(ω)* is the spectral density function at frequency *ω*, and ω_D is the ^{2}H Larmor frequency.

The following sections describe three methods of calculating the spectral density values from relaxation rate constants measured at more than one static magnetic field.

Independent Spectral Density Mapping

The relaxation rate constants, Eqs. 1-4, can be written in matrix form as:

$$\begin{bmatrix} R_1 \\ R_{1\rho} \\ R_Q \\ R_{AP} \end{bmatrix} = 3\xi^2 \begin{bmatrix} 0 & 1 & 4 \\ 3/2 & 5/2 & 1 \\ 0 & 3 & 0 \\ 3/2 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} J(0) \\ J(\omega_D) \\ J(2\omega_D) \end{bmatrix} \tag{6}$$

and constitute an overdetermined system of linear equations. In the *independent method*, Eq. 6 is solved by least-squares methods (*vide infra*) to obtain values of *J*(0), $J(\omega_D)$, and $J(2\omega_D)$ at each static magnetic field. Thus, relaxation rates collected at ^{2}H frequencies of 61.4, 76.8, 122.8, and 138.2 MHz were independently analyzed using Eq. 6 to generate the spectral densities for this method. At each of the four static magnetic fields, three spectral density values are derived from four relaxation rate constants, giving one degree of freedom per field or a total of four degrees of freedom. The four values of *J*(0) obtained for each methyl group were averaged, as were the values of $J(2\ \omega_D)$ = *J*(122.8) obtained from data acquired at 9.4 T (400 MHz for 1H spins) and $J(\omega_D)$ = *J*(122.8) obtained from data acquired at 18.8 T (800 MHz). Thus, this analysis yields values for

J(ω_D) at eight frequencies: $J(0)$, $J(61.4)$, $J(76.8)$, $J(122.8)$, $J(138.2)$, $J(153.6)$, $J(245.6)$, and $J(276.4)$, in which the frequencies are given as $\omega/(2\pi)$ in units of MHz.

Joint Spectral Density Mapping

In the second approach, relaxation data acquired at multiple static magnetic fields are analyzed jointly to consider any spectral density values that contribute to the relaxation rate constants at more than one static field. In the present case, all data sets share a common value of $J(0)$ and in addition, $J(2\omega_D)$ for data acquired at 9.4 T is identical to $J(\omega_D)$ for data acquired at 18.8 T. The resulting matrix equation is:

$$\begin{bmatrix} \widehat{\boldsymbol{R}}_{61.4} \\ \widehat{\boldsymbol{R}}_{76.8} \\ \widehat{\boldsymbol{R}}_{122.8} \\ \widehat{\boldsymbol{R}}_{138.2} \end{bmatrix} = 3\xi^2 \begin{bmatrix} \boldsymbol{V} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{X} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{V} & \boldsymbol{0} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{V} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{V} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X} \end{bmatrix} \begin{bmatrix} J(0) \\ J(61.4) \\ J(76.8) \\ J(122.8) \\ J(138.2) \\ J(153.6) \\ J(245.6) \\ J(276.4) \end{bmatrix} \qquad \textbf{(7)}$$

in which $\hat{\mathbf{R}}_i = \left[R_{1,i}, R_{1\rho,i}, R_{Q,i}, R_{AP,i}\right]^T$, $\mathbf{V} = \left[0,3/2,0,3/2\right]^T$, $\mathbf{W} = \left[1,5/2,3,1/2\right]^T$, $\mathbf{X} = \left[4,1,0,1\right]^T$, and $\mathbf{0}$ is a 4 × 1 zero-matrix. The least squares solution of this equation yields values for $J(\omega_D)$ at eight frequencies and no post-analysis averaging of common values is needed. Eight spectral density values are derived from 16 relaxation rate constants, with consequently eight degrees of freedom.

Joint-Interpolated Spectral Density Mapping

The *joint method* shown in Eq. 7 links relaxation data acquired at 9.4 T (400 MHz) and 18.8 T (800 MHz), fields that differ by a factor of two. The *joint-interpolated method* uses the data acquired at 9.4 T (400 MHz) and 11.7 T (500 MHz) to interpolate values of the relaxation rate constants at 10.55 T (450 MHz). The interpolated relaxation rate constants are then paired with the relaxation data acquired at 21.1 T (900 MHz).

The normalized correlation function for reorientation of the C-D bond vector is described as [111, 112]:

$$C(t) = e^{-\frac{t}{\tau_m}}(S^2 + \textstyle\sum_i a_i e^{-\frac{t}{\tau_{e,i}}}) \tag{8}$$

in which τ_m is the overall rotational correlation time, S^2 is the generalized order parameter, and a_i and $\tau_{e,i}$ are the amplitude and time constant for the ith term in the internal correlation function. Taking the real part of the Fourier transform of $C(t)$ yields the spectral density function:

$$J(\omega) = \frac{S^2\tau_m}{1+\omega^2\tau_m^2} + \sum \frac{a_i\tilde{\tau}_i}{1+\omega^2\tilde{\tau}_i{}^2} \tag{9}$$

in which:

$$\tilde{\tau}_i = (\frac{1}{\tau_m} + \frac{1}{\tau_{e,i}})^{-1} \tag{10}$$

which is a linear function of ω^{-2}. Thus, a first-order Taylor series approximation of a given relaxation rate constant, R_m, with respect to ω^{-2} yields:

$$R_{m,\omega_{D2}} \approx R_{m,\omega_{D1}} + (R_{m,\omega_{D2}} - R_{m,\omega_{D1}})(\frac{1}{\omega_{D2}^2} - \frac{1}{\omega_{D1}^2}) \tag{12}$$

$$R_{m,\omega_{D2}} \approx R_{m,\omega_{D3}} + (R_{m,\omega_{D2}} - R_{m,\omega_{D3}})(\frac{1}{\omega_{D2}^2} - \frac{1}{\omega_{D3}^2}) \tag{13}$$

Adding these together and dividing by two gives:

$$R_{m,\omega_{D2}} \approx \frac{R_{m,\omega_{D1}} + R_{m,\omega_{D3}}}{2} + \left(R_{m,\omega_{D1}} - R_{m,\omega_{D3}}\right)\frac{2\omega_{D2}^{-2} - \omega_{D1}^{-2} - \omega_{D3}^{-2}}{2(\omega_{D1}^{-2} - \omega_{D3}^{-2})} \tag{14}$$

Based on this derivation, the interpolation of relaxation rate constants to 10.55 T ($\omega_D/(2\pi)$ = 69.1 MHz) is:

$$R_{m,69.1} = \frac{R_{m,61.4} + R_{m,76.8}}{2} + \left(R_{m,61.4} - R_{m,76.8}\right)\frac{2\times 69.1^{-2} - 61.4^{-2} - 76.8^{-2}}{2\times(61.4^{-2} - 76.8^{-2})} \tag{15}$$

which can be expressed as:

$$R_{m,69.1} = R_{m,61.4}\left(\frac{1}{2} + \kappa\right) + R_{m,76.8}\left(\frac{1}{2} - \kappa\right) \tag{16}$$

in which:

$$\kappa = \frac{2\times 69.1^{-2} - 61.4^{-2} - 76.8^{-2}}{2\times(61.4^{-2} - 76.8^{-2})} \tag{17}$$

In matrix form, the relationship between the relaxation rate constants at the three static magnetic fields becomes:

$$\begin{bmatrix} R_{m,61.4} \\ R_{m,76.8} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1+2\kappa}{1-2\kappa} & \frac{2}{1-2\kappa} \end{bmatrix} \begin{bmatrix} R_{m,61.4} \\ R_{m,69.1} \end{bmatrix} \tag{18}$$

Consequently, for relaxation rates collected at deuterium frequencies of 61.4, 76.8, 122.8, and 138.2 MHz:

$$\begin{bmatrix} \widehat{\boldsymbol{R}}_{61.4} \\ \widehat{\boldsymbol{R}}_{76.8} \\ \widehat{\boldsymbol{R}}_{122.8} \\ \widehat{\boldsymbol{R}}_{138.2} \end{bmatrix} = 3\xi^2 \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ -\frac{1+2\kappa}{1-2\kappa}\boldsymbol{I} & \boldsymbol{0} & \frac{2}{1-2\kappa}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{Z} & \boldsymbol{0} \\ \boldsymbol{Y} & \boldsymbol{0} & \boldsymbol{Z} \end{bmatrix} \begin{bmatrix} J(0) \\ J(61.4) \\ J(122.8) \\ J(245.6) \\ J(69.1) \\ J(138.2) \\ J(276.4) \end{bmatrix} \tag{19}$$

in which:

$$\boldsymbol{Y} = \begin{bmatrix} 0 \\ 3/2 \\ 0 \\ 3/2 \\ 0 \\ 3/2 \\ 0 \\ 3/2 \end{bmatrix} \qquad \boldsymbol{Z} = \begin{bmatrix} 1 & 4 & 0 \\ 5/2 & 1 & 0 \\ 3 & 0 & 0 \\ 1/2 & 1 & 0 \\ 0 & 1 & 4 \\ 0 & 5/2 & 1 \\ 0 & 3 & 0 \\ 0 & 1/2 & 1 \end{bmatrix} \tag{20}$$

and **0** and **I** are 4-dimensional zero and identity matrices, respectively. Equation 19 is the main theoretical result of the present work. The solution of Eq. 19 yields values of the spectral density function at seven frequencies, with consequently nine degrees of freedom.

For each of the three spectral density mapping methods, spectral density values were obtained from the relaxation rates R_1, $R_{1\rho}$, R_Q, and R_{AP} measured at deuterium frequencies 61.4, 76.8, 122.8, and 138.2 MHz by solving Eqs. 6, 7, or 19 by singular value decomposition. Uncertainties in the spectral density values were obtained from the covariance matrix.

5.4.2 Model-Free Analysis

Spectral density values obtained by the *independent*, *joint*, and *joint-interpolated* methods were fit with the Lipari-Szabo model-free spectral density function [111]:

$$J(\omega) = \frac{2}{5}\left(\frac{\frac{1}{9}S_{axis}^2\tau_m}{1+\omega^2\tau_m^2} + \frac{\left(1-\frac{1}{9}S_{axis}^2\right)\tau}{1+\omega^2\tau^2}\right) \tag{21}$$

in which S^2_{axis} is the generalized order parameter for the C-D bond vector, the factor of 1/9 accounts for (rapid) rotation of the methyl group, and $\tau = (1/\tau_m + 1/\tau_e)^{-1}$. Fitting was performed by minimizing:

$$\chi^2 = \sum_{k=1}^{K}\left[\frac{J(\omega_k)-\hat{J}(\omega_k)}{\sigma_k}\right]^2 \tag{22}$$

in which K is the number of spectral density values being fit, σ_k are the uncertainties in spectral density values determined as described above, and $\hat{J}(\omega_k)$ is the fitted value of the spectral density function. Minimization was performed using weighted non-linear least-squares in MATLAB. Uncertainties in fitted model-free parameters were obtained from Monte Carlo simulations. Values of S^2, τ_m, and τ_e were restricted to the ranges [0, 1], [0 ns, 20 ns], and [0 ns, 20 ns], respectively. Goodness-of-fit to Eq. 21 was tested by comparing the χ^2 residual from the curve-fitting procedure for each methyl group to critical values obtained for a Bonferroni-corrected confidence level p = α/N, where $\alpha = 0.05$ and $N = 48$ is the number of methyl groups analyzed. The critical χ^2 values for the *independent*, *joint*, and *joint-interpolated* methods are 27.8, 20.4, and 18.4, respectively.

5.5 Results

5.5.1 Simulation of 69.1 MHz Deuterium Frequency Relaxation Rate Interpolation

The accuracy of the proposed method for interpolation of relaxation rate constants necessary for the *joint-interpolated* approach was evaluated by theoretical calculations. Relaxation rates at 9.4 T, 10.55 T, and 11.7 T were calculation using Eqs. 1-4 and the Lipari-Szabo model-free formalism, Eq. 21. The percentage differences between the interpolated and exact values of the relaxation rate constants at 10.55 T are shown in Figure 5.2. The predicted rates show errors of ≤ 1% up to a τ_c of 1 ns, consistent with the experimental uncertainties in measured ^{2}H relaxation rates (~1-2%).

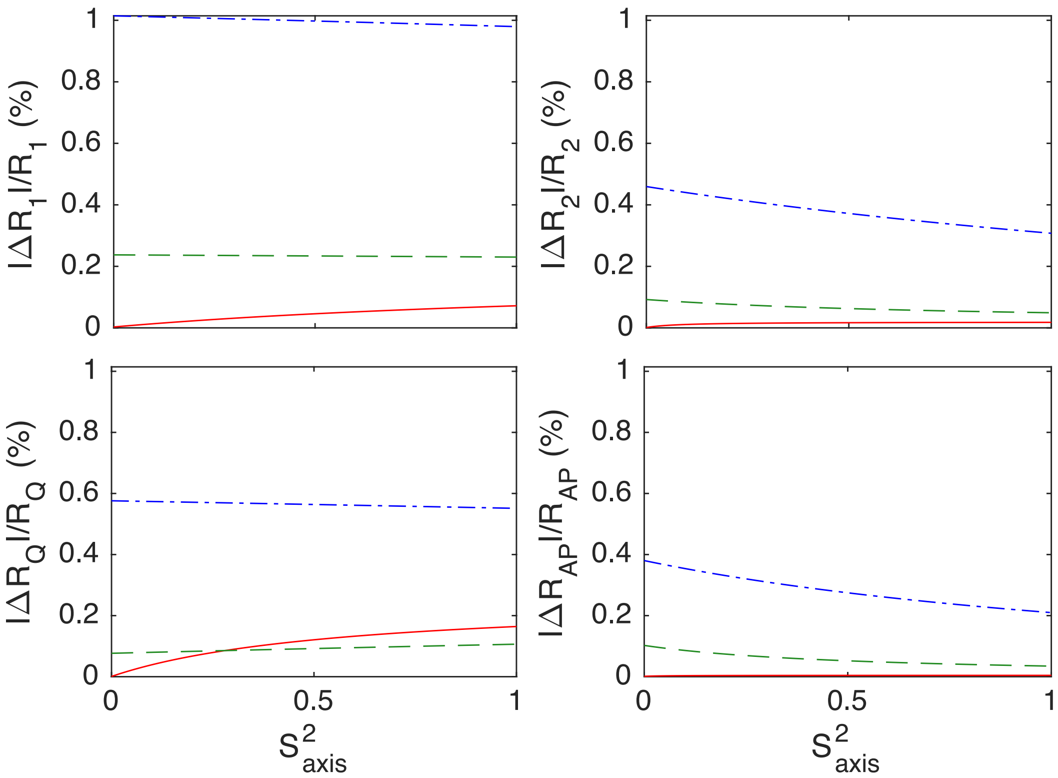


Figure 5.2: **Simulation of ^{2}H Relaxation Rate Interpolation.**
Predicted errors for Eq. 16 over a range of methyl axis order parameters S_{axis}^2 from 0 to 1, given an overall tumbling time τ_m of 10 nanoseconds (ns), and effective correlation times τ_c of (solid, red) 0.03 ns, (dashed, green) 0.3 ns, and (dashed-dotted, blue) 1 ns. [77]

5.5.2 Model Selection

Spectral density values for each methyl resonance determined by each spectral density mapping method were fit with Eq. 21. If any of the three fits had a $\chi2$ value that was less than or equal to the respective critical $\chi2$ value for a particular methyl group, then that methyl group was included in subsequent analysis. Methyl groups that did not meet this criterion were not considered further; these methyl groups require fitting with more complex motional models than Eq. 21 [90]. Overall, 33 of the 48 ILV methyl resonances satisfied this criterion. Fits to data for Leu 103 δ1 converged poorly and yielded large errors for τ_m; consequently, this methyl group was not considered further. Figure 5.3 shows the locations of ILV amino acids in RNase H. Most of the methyl groups chosen for analysis are in residues located either in α-helices or β-sheets, while the other methyl groups are typically located in dynamic loops or termini.

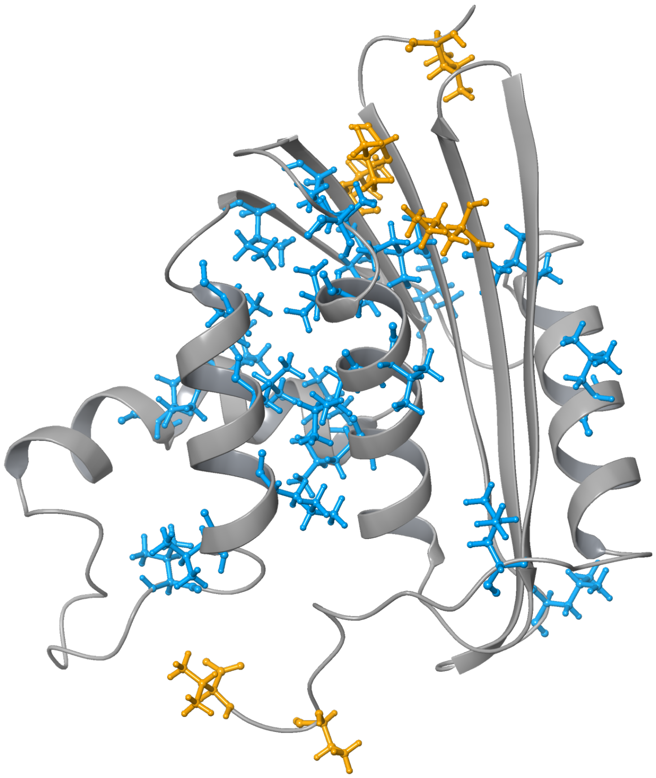


Figure 5.3: ***E. coli*** **RNase H with ^{13}CDH$_2$ Probes Highlighted.**
Ribbon diagram of RNase H with the observable ILV residues in stick representations. The ILV residues fit with the Lipari-Szabo model-free formalism, Eq. 21, are colored blue; other ILV residues that were excluded are colored orange. [77]

5.5.3 Representative Spectral Density Functions

Examples of the Lipari-Szabo model-free fits to the spectral density values obtained from the three spectral density mapping methods are shown in Figure 5.4. Values of τ_m were optimized independently for each methyl group and consequently represents an effective local correlation time and could be smaller than the global rotational correlation time of RNase H, as determined from ^{15}N spin relaxation [108]. Results are shown for Val 121 γ2, with a local correlation time of ~7 ns and Leucine 56 δ1 with a local correlation time of ~14 ns.

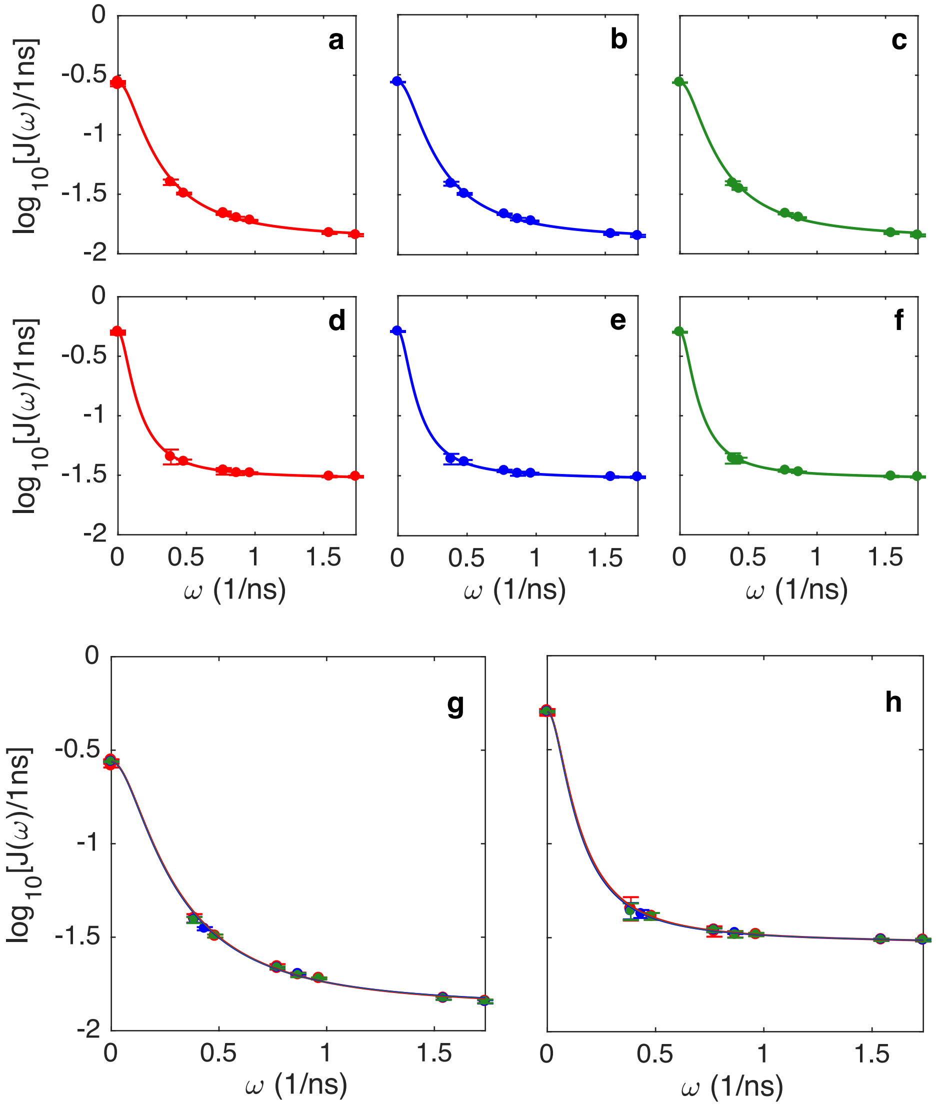


Figure 5.4: **Lipari-Szabo Model-Free Fits to the Spectral Density Values Obtained from Three SDM Methods.**

Subfigures (*a-f*) represent the individual fits for (red) *independent*, (blue) *joint*, and (green) *joint-interpolated*. The first row (*a-c*) shows plots for Val 121 γ2, with each method, plotted individually. The second row (*d-f*) shows plots for Leu 56 δ1, with each method also plotted individually. (*g*) Superposition of data and fits for the three methods are shown for (*g*) Val 121 γ2 and (*h*) Leu 56 δ1. [77]

5.5.4 Accuracy and Precision of Model Free Parameters

Figure 5.5 summarizes the results of fitting the spectral densities to the Lipari-Szabo model-free formalism to obtain S^2_{axis}, τ_m, and τ_e for 32 methyl groups. Each graph in the figure shows a one-on-one comparison between model-free parameters determined by two of the three spectral density mapping procedures. Coefficients of determination, R^2, were calculated for each of the plots based on the y = x line in Figure 5.5. The R^2 values were 0.99, 0.94, and 0.97, respectively, for S^2_{axis} comparisons, Figures 5.5(*a-c)*; 0.98, 0.93, and 0.97, respectively, for τ_m comparisons, Figures 5.5(*d-f)*; and 1.00 for all three τ_m comparisons, Figures 5.5(*d-f*). Slopes of least-squares fitted lines were 0.99, 0.97, and 0.98, respectively, for S^2_{axis} comparisons, Figures 5.5(*a-c)*; 1.01, 1.02, and 1.02, respectively, for τ_m comparisons, Figures 5.5*(d-f)*; and 1.00 for all three τ_e comparisons, Figures 5.5(*g-i)*. The average uncertainties in the fitted Lipari-Szabo parameters, S^2_{axis}, τ_m, and τ_e, differed between the three methods of spectral density mapping. The *joint-interpolated* method yielded the smallest parameter uncertainties and Table 5.1 presents the percentage improvement in average uncertainties for this method compared with the *independent* and *joint* approaches.

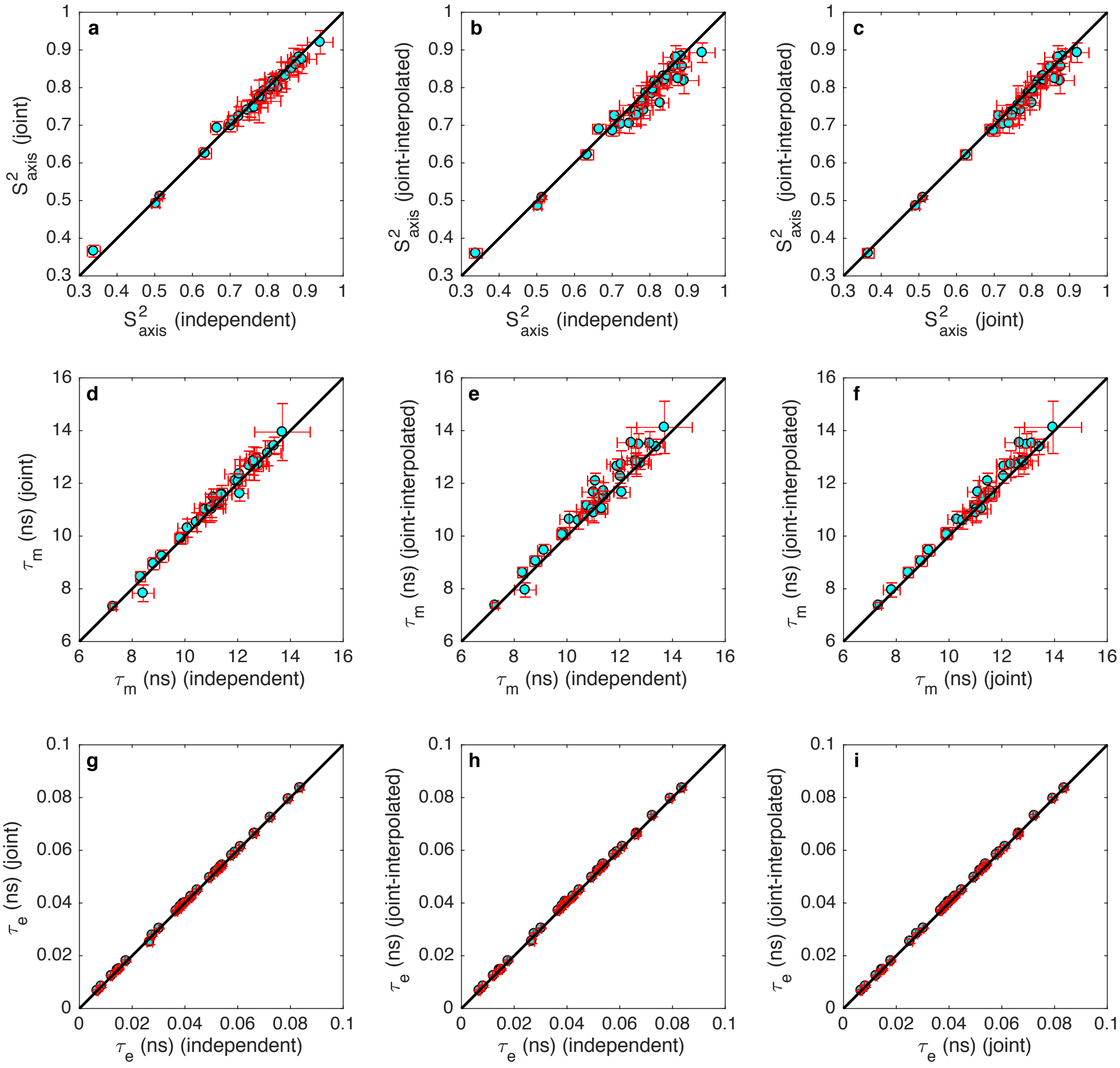


Figure 5.5: **Comparison between Model-Free Parameters Determined by Each of the Three Different SDM Methods.**

Correlation plots for S^2_{axis} (top row A-C), τ_m (middle row D-F), and τ_e (bottom row G-I). Each row compares the model free parameters generated from the independent, joint, and joint-interpolated methods. The parameters of the 32 residues are shown in cyan, while the error bars are depicted in red. The black line represents *y*=*x*. [77]

% More Precise	*Joint-Interpolated* vs. *Independent*	*Joint-Interpolated* vs. *Joint*
S^2	14.6%	10.8%
τ_m	10.5%	7.8%
τ_e	12.8%	10.1%

Table 5.1: **Improvement in Precision of Fitted Lipari-Szabo Model-Free Parameters.**

5.5.5 Cross Validation with 700 MHz Experimental Data

The fitted model-free parameters were cross-validated by back-calculating relaxation rate constants expected for data acquired at 16.4 T (700 MHz, $\omega_D/(2\pi)$ = 107.5 MHz) using Eqs. 1-4 and Eq. 21 and compared to experimental data acquired at 16.4 T. Uncertainties in back-calculated relaxation rate constants were generated by Monte Carlo simulations. The cross-validated results are shown in Figure 5.6. The three methods of spectral density mapping gave equally accurate back predictions of the relaxation rate constants compared with experimental values. R_1 comparisons had R^2 values of 1.00, $R_{1\rho}$ comparisons had R^2 values of 0.99, R_Q comparisons had R^2 values of 0.99, and R_{AP} comparisons had R^2 values of 0.98. R_1 and $R_{1\rho}$ comparisons had slopes of 1.01, R_Q comparisons for *independent* and *joint methods* had slopes of 0.99, R_Q comparison for the *joint-interpolated method* had a slope of 0.98, and R_{AP} comparisons had slopes of 1.02.

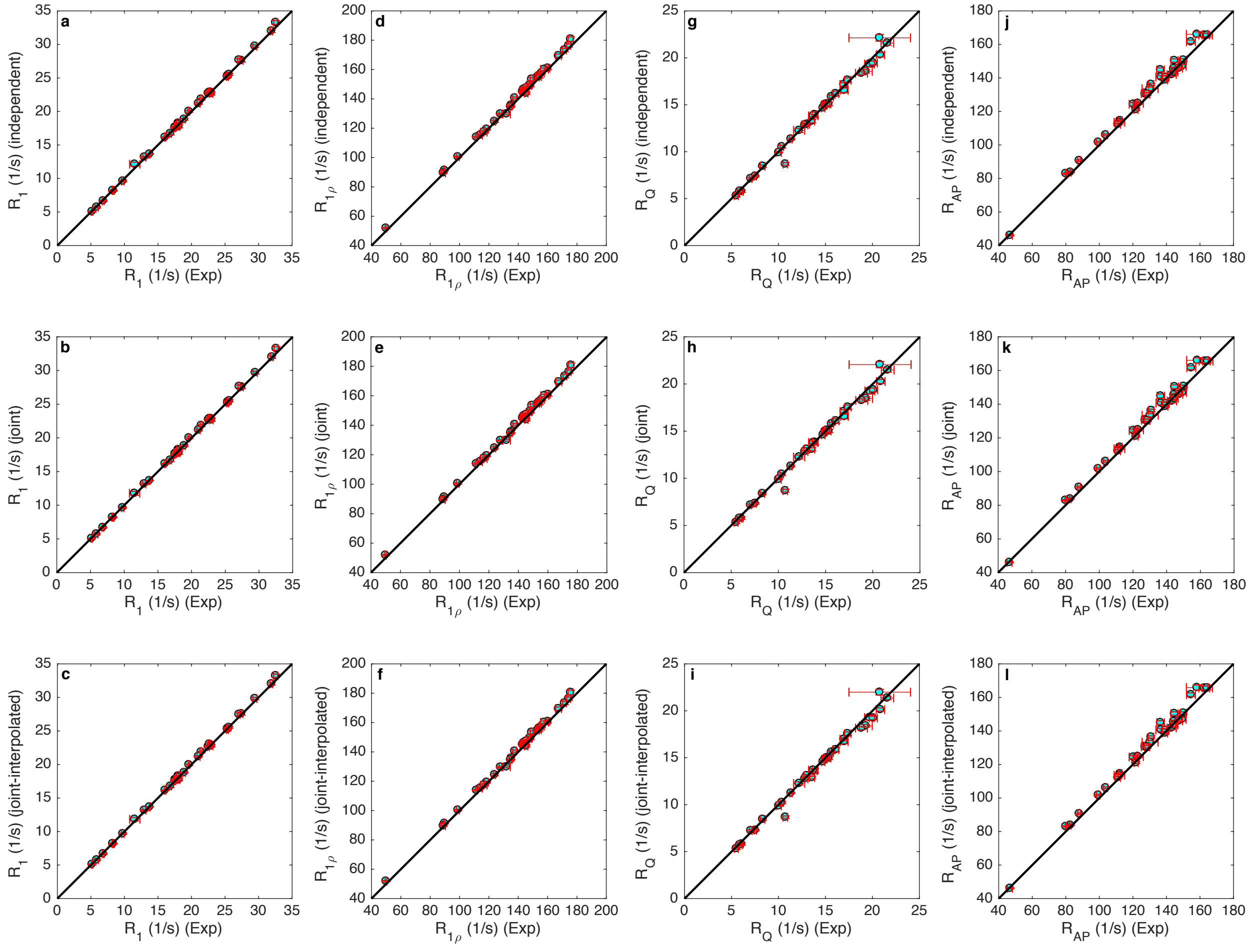


Figure 5.6: **Comparison of Back-Calculated 700 MHz ^{2}H Relaxation Rates to Experimentally Measured ^{2}H Relaxation Rates.**

Back-calculated 700 MHz relaxation rate constants (y-axis) are compared to the experimental results (x-axis). Each row represents a method of spectral density mapping, while each column represents a relaxation rate constant. The relaxation rates of the 32 residues are shown in cyan, while the error bars are depicted in red. The black line represents y=x. [77]

5.5.6 Local Diffusion Times

The values of τ_m obtained for the most rigid methyl-bearing side chains in RNase H are expected to report on the overall rotational diffusion tensor, which has been determined previously by ^{15}N spin relaxation [108]. As shown in Figure 6, the values of the local diffusion constants, D, for the 12 methyl groups with the largest values of τ_m (smallest values of D; shown in red) agree well with the predictions derived from the diffusion tensor obtained from ^{15}N spin relaxation. This agreement provides additional evidence for the accuracy of the proposed analysis for determination of rotational diffusion tensors of proteins, particularly in cases for which orientations of N-H amide bond vectors are not uniformly distributed [113].

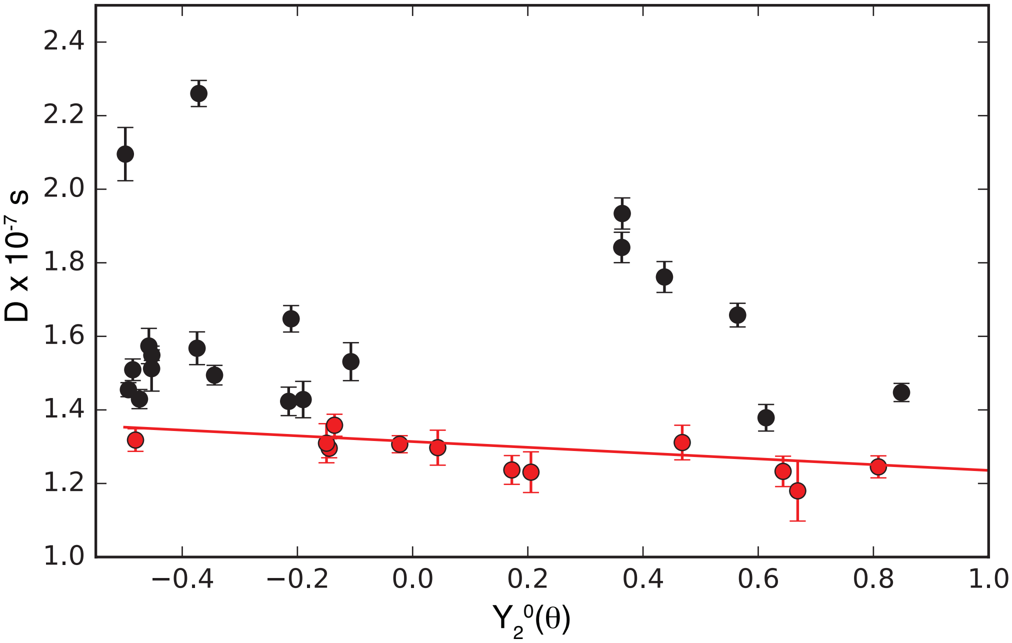


Figure 5.7: **Comparison of Local Diffusion Times to ^{15}N-Derived Diffusion Tensors.** Local diffusion times ($D = 1/(6\tau_m)$) plotted versus $Y_2^0(\theta)$, in which θ is the orientation of the C-D bond in the structure of RNase H oriented in the principal axis system of the diffusion tensor determined from ^{15}N spin relaxation. The values of D for the 12 slowest tumbling residues (for which internal motions are well-separated from overall rotation) are shown in red and agree quantitatively with predictions from the overall rotational diffusion tensor determined from ^{15}N relaxation data (red line), when adjusted for the sample temperature and D_2O viscosity. [77]

5.6 Discussion

NMR spin relaxation methods are powerful approaches for characterizing conformational dynamics of biological macromolecules. Weaknesses of these methods are that the target spectral density function can only be sampled at the discrete eigenfrequencies of the spin system being studied and that model spectral density functions can have many optimizable parameters. Both of these weaknesses are alleviated by acquiring additional relaxation data at multiple static magnetic fields.

The present work examined three methods of performing spectral density mapping of ^{2}H relaxation in ^{13}CH$_2$D methyl group isotopomers, which have been termed independent, joint, and joint-interpolated. The latter two approaches make use of a unique feature of ^{2}H quadrupolar relaxation: the number of spectral density values entering into the relaxation equations is reduced if data are acquired at pairs of static magnetic fields that differ by a factor of two. As relatively few such paired fields are available at present (in the current work, only 400 MHz/800 MHz), the joint-interpolation method approximates such pairings, in the current work by interpolating between 400 MHz and 500 MHz to generate relaxation data at 450 MHz for pairing with 900 MHz data. The major advantages of the joint and joint-interpolation methods are the consequent additional increases in the statistical degrees of freedom compared with the independent method for subsequent analyses.

Comparisons between the three spectral density mapping methods and cross validation with an independent data set indicated that the three approaches were equally accurate; however, the joint-interpolation method yielded an improvement in precision for the final fitted model-free parameters of between 10-15%, compared to the independent method and of between 7-11% compared with the joint analysis. These results would make the joint-interpolation method preferable over the other two methods. Notably, no additional data need to be acquired to implement the joint-interpolated method and as more commercial NMR spectrometers with frequencies of 1000 MHz and 1200 MHz become available, additional pairs of fields become accessible, further increasing the power of these methods.

Chapter 6

NMR Studies of Localized Protein Dynamics in a Psychrophilic Cytochrome c_{552}

This final thesis chapter consists of work initiated in parallel and that is currently ongoing. Herein, the use of NMR spectroscopy to understand protein thermostability in psychrophilic (cold-loving) cytochrome c_{552} is described. Isolated from the bacterium *Colwellia psychrerythraea (Cp)*, biochemical and structural studies of *Cp*cyt c_{552} implicate three conserved methionine residues in heme-ligation and propose that elevated backbone flexibility in these residues accounts for an uncharacteristically high melting temperature (T_m) compared to meso- and thermophile *c*-type cytochromes [114, 115]. Progress was made in NMR sample preparation, backbone protein NMR experiments, and insight from 1D ^{1}H NMR experiments focused on the heme group bound to *Cp* cytochrome c_{552} is discussed. Additionally, chemical shifts are used to predict protein dynamics as a first test of a multiple methionine axial ligand hypothesis. Overall, findings from this last chapter bring centerstage a defining aspect of NMR spectroscopy: even as pioneering biophysical methods trend towards structural elucidation of larger proteins and macromolecular complexes, some of the most important biology occurs so localized in proteins that the specificity and resolution of NMR experiments remain unmatched in the information they can provide.

6.1 *Cp*cyt c_{552} Crystal Structure and Protein Folding Measurements

Colwellia psychrerythraea ("*Cp*") is a bacterium found in cold, marine environments. Specifically, the bacterium has been found to grow at temperatures below 10°C. In addition to being a psychrophile, *Cp* is also classified as a hydrocarbonclastic bacterium. This implies that *Cp* is able to efficiently metabolize crude or refined petroleum products. As an example, species of *Colwellia* were identified in marine samples from the Gulf of Mexico following the blowout of BP's *Deepwater Horizon* well in 2010. The persistence of these bacteria at temperatures approaching 2-5°C poses a simple yet poorly understood issue: given the dependence of catalytic rate constants on temperature (Arrhenius equation), how do proteins and enzymes carry out biochemical reactions at appreciable rates under these conditions? With only 20% of Earth's ecosystems permanently at temperatures over 5°C, this issue is particularly salient. These lines of inquiry animated the current study of electron transfer metalloprotein cytochrome c_{552} from *Cp* as a way of gaining initial insight into psychrophilicity. A rich literature of mesophilic and thermophilic c-type cytochrome data exists, thus facilitating comparison with psychrophilic cytochrome. The following takeaways come from *Cp* cytochrome c_{552} biochemical and biophysical experiments performed in the Magyar lab at Columbia University [114, 115]:

i) A 2 Å resolution crystal structure of *Cp* cytochrome c_{552} (*Cp*cyt c_{552}) superimposes well with *c*-type cytochrome structure from mesophilic marine bacterium *Marinobacter hydrocarbonoclasticus* (*Mh*) despite only 57% sequence similarity (Figure 6.1.1).

ii) *Cp*cyt c_{552} crystallizes as a dimer and size-exclusion chromatography as well as analytical ultracentrifugation confirm dimers persist in solution as well. The crystal structure shows that hydrogen bonds between protein residues (Y39, K42, N43, and R52), heme propionates and water comprise the interface between *Cp*cyt c_{552} monomers.

iii) The crystal structure also shows that heme is covalently bound to C14 and C17, while H18 and M57 axially ligate the heme iron. Distances between the iron and ligating residue are consistent with the presence of low spin Fe(III).

iv) The presence of six methionine residues in *Cp*cyt c_{552} is noteworthy. Specifically, M57, M60, and M62 are shown to be conserved throughout the primary sequences of psychrophiles and psychrotolerant bacteria. While M57 is conserved in mesophiles, which is understandable given its position as the native axial methionine for heme Fe, sequence positions corresponding to M60 and M62 show no conservation for methionine or even a consensus amino acid type (Figure 6.1.2). Superimposing structures of *Cp*cyt c_{552} and *Mh*cyt c_{552} with residue-averaged B-factors leads to a corresponding observation: as expected, overall flexibility (B-factors) is greater for the psychrophile compared to the mesophile, but interestingly enough, two regions of high B-factor measurement correspond to the M60 and M62 (Figure 6.1.3).

v) Compared to mesophilic proteins, *Cp*cyt c_{552} exhibits a similar melting temperature (T_m) but a lower van’t Hoff enthalpy (ΔH_{vH}). This is surprising, since measurement of a lower T_m and increased enthalpy would be expected for a psychrophilic protein. Another peculiar observation: *Cp*cyt c_{552} consists of 16 alanine residues (~20% of the protein). The preponderance of these hydrophobic yet flexible residues in *Cp*cyt c_{552} agrees well with observations of decreased enthalpy and increased protein flexibility. However, these appear to be incongruous with the observed unchanged T_m (compared to the mesophilic eukaryotic yeast iso-1-cytochrome *c*).

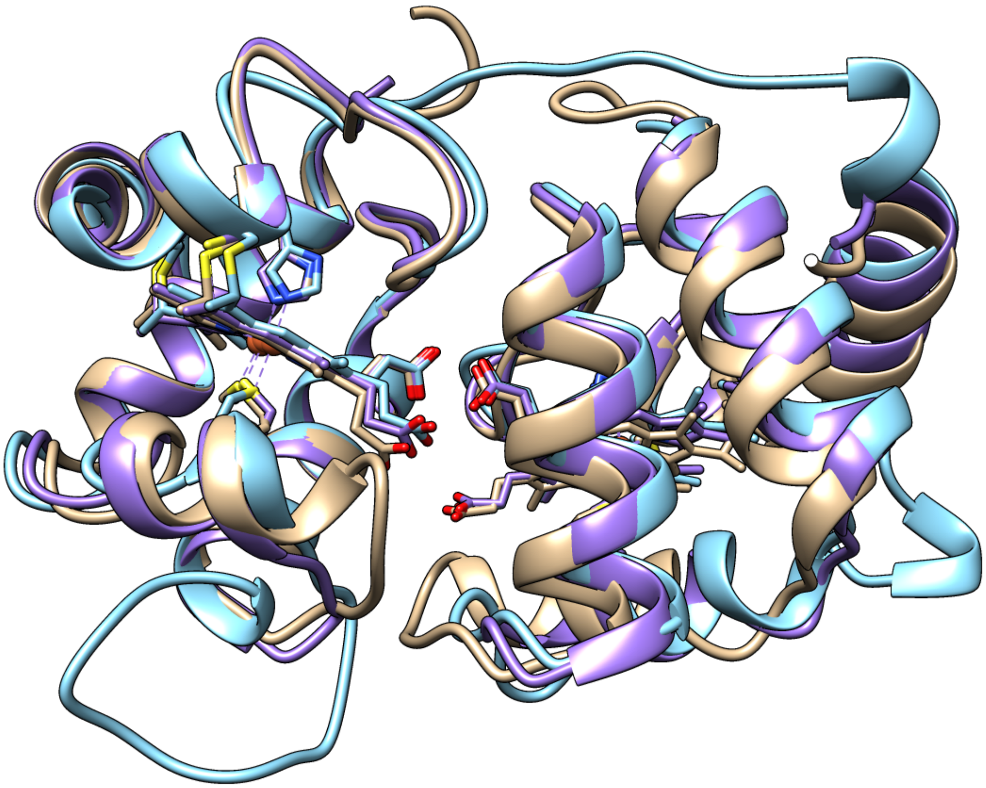
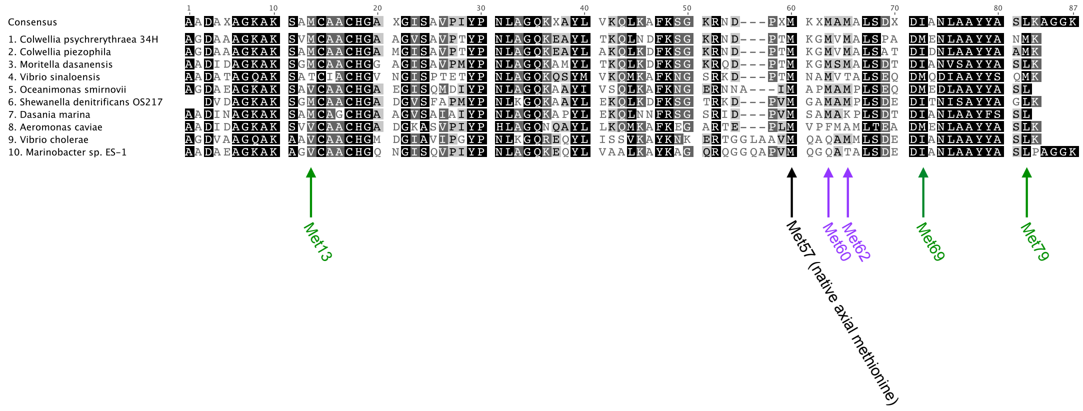


Figure 6.1.1: **Superposition of Psychrophilic, Mesophilic, and Two-Heme *c*-type Cytochromes.**

The structures of dimeric *Cp*cyt c_{552} (PDB:4O1W, purple), *Mh*cyt c_{552} (PDB:1CNO, tan), and *P. stutzeri* two-heme *c*-type cytochrome (PDB:1ETP, blue). The individual structures superimpose well, showing the ubiquitous nature of the cytochrome fold. Reprinted by permission from RSC Pub: Metallomics: integrated biometal science, 6, Harvilla, P.B., Wolcott, H.N., Magyar, J.S., The structure of ferricytochrome c_{552} from the psychrophilic marine bacterium *Colwellia psychrerythraea*. 1126-1130, Copyright 2014. [115]



```
Consensus                           AADAXAGKAK SAMCAACHGA XGISAVPIYP NLAGQKXAYL VKQLKAFKSG KRND---PXM KXMAMALSDX DIANLAAYYA SLKAGGK

1. Colwellia psychrerythraea 34H    AGDAAAGKAK SVMCAACHGA AGVSAVPTYP NLAGQKEAYL TKQLNDFKSG KRND---PTM KGMVMALSPA DMENLAAYYA NMK
2. Colwellia piezophila             AGDAAAGKAK SAMCAACHGA MGISAVPTYP NLAGQKEAYL AKQLKDFKSG KRND---PTM KGMVMALSAT DIDNLAAYYA AMK
3. Moritella dasanensis             AADIDAGKAK SGMCAACHGG AGISAVPMYP NLAGQKAMYL TKQLKDFKSG KRQD---PTM KGMSMALSDT DIANVSAYYA SLK
4. Vibrio sinaloensis               AADATAGQAK SATCIACHGV NGISPTETYP NLAGQKQSYM VKQMKAFKNG SRKD---PTM NAMVTALSEQ DMQDIAAYYS QMK
5. Oceanimonas smirnovii            AGDAEAGKAK SAVCAACHGA EGISQMDIYP NLAGQKAAYI VSQLKAFKNG ERNNA---IM APMAMPLSEQ DMEDLAAYYA SL
6. Shewanella denitrificans OS217     DVDAGKAK SGMCAACHGA DGVSFAPMYP NLKGQKAAYL EKQLKDFKSG TRKD---PVM GAMAMPLSDE DITNISAYYA GLK
7. Dasania marina                   AADINAGKAK SAMCAGCHGA AGVSAIAIYP NLAGQKAPYL EKQLNNFRSG SRID---PVM SAMAKPLSDT DIANLAAYFS SL
8. Aeromonas caviae                 AADIDAGKAK SVVCAACHGA DGKASVPIYP HLAGQNQAYL LKQMKAFKEG ARTE---PLM VPFMAMLTEA DMENLAAYYA SLK
9. Vibrio cholerae                  AGDVAAGQAK AAVCAACHGM DGIAVIPGYP NLKGQREQYL ISSVKAYKNK ERTGGLAAVM QAQAMMLSDE DIANLAAYYA SLK
10. Marinobacter sp. ES-1           AADAEAGKAK AGVCAACHGQ NGISQVPIYP NLAGQKEQYL VAALKAYKAG QRQGGQAPVM QGQATALSDE DIANLAAYYA SLPAGGK
```

Figure 6.1.2: **Sequence Alignment of Homologous *c*-type Cytochrome Sequences.** Ten different cytochrome c_{552} protein sequences are aligned. Psychrophilic sequences are 1-3, while sequences 4-7 comes from psychrotolerant organisms. The remainder are mesophilic cytochromes. Importantly, position 57 is conserved for methionine across all sequences. The suspected axial methionine substitutes (M60 and M62) are only conserved in psychrophilic and psychrotolerant species. Reprinted by permission from RSC Pub: Metallomics: integrated biometal science, 6, Harvilla, P.B., Wolcott, H.N., Magyar, J.S., The structure of ferricytochrome c_{552} from the psychrophilic marine bacterium *Colwellia psychrerythraea*. 1126-1130, Copyright 2014. [115]

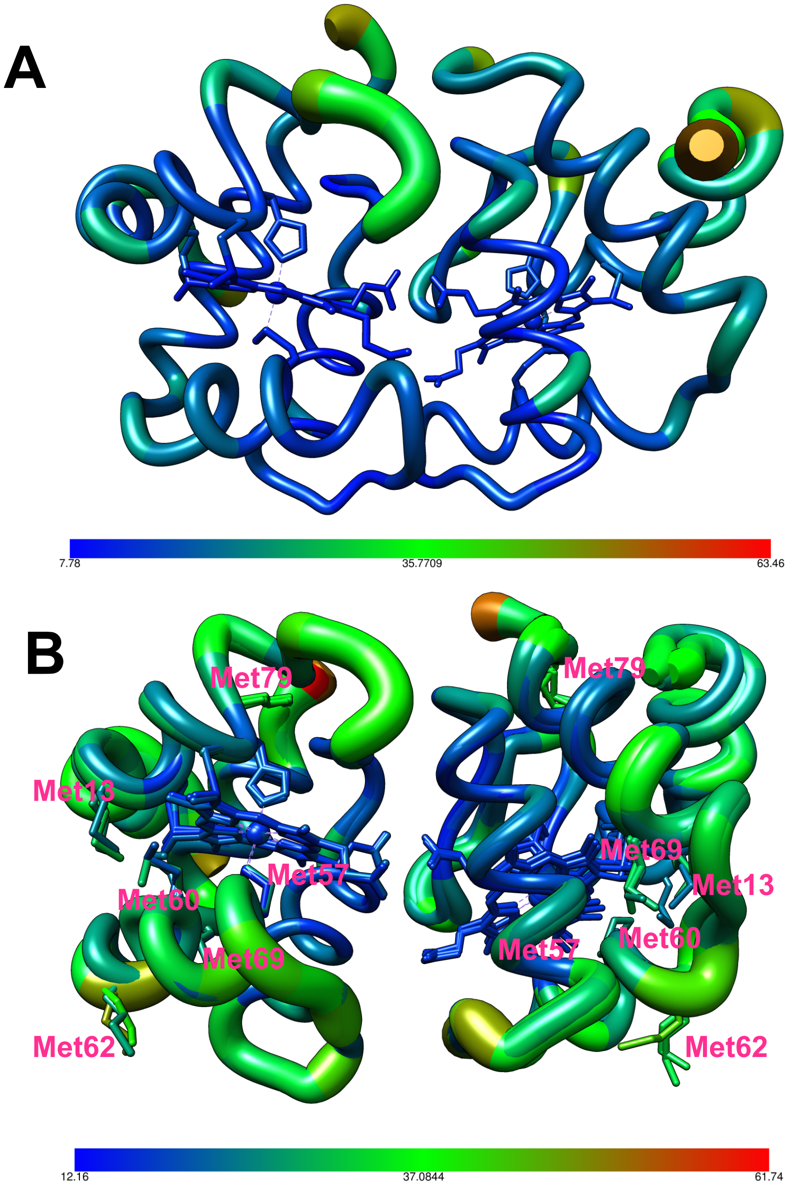


Figure 6.1.3: **B-factor overlay for Psychrophilic and Mesophilic cytochrome c_{552}.** The structures of *Mh*cyt c_{552} (PDB:1CNO, **A**) and *Cp*cyt c_{552} (PDB:4O1W, **B**) both colored and portrayed with thicker worm-like structure according to B-factor measurement. The *Cp*cyt c_{552} shows both overall high B-factors, which indicates greater flexibility, but also localized, elevated B-factors near the region (T55-A63) of 1interest. Reprinted by permission from RSC Pub: Metallomics: integrated biometal science, 6, Harvilla, P.B., Wolcott, H.N., Magyar, J.S., The structure of ferricytochrome c_{552} from the psychrophilic marine bacterium *Colwellia psychrerythraea*. 1126-1130, Copyright 2014. [115]

These takeaways have motivated the current study of *Cp*cyt c_{552} by NMR spectroscopy. In this system, localized protein flexibility, as opposed to global protein dynamics, is believed to affect the *Cp*cyt c_{552} unfolding pathway. Specifically, M60 and M62 are suspected of acting as substitute-ligands for axial heme ligation, thus capturing partially unfolded protein in local minima as the reaction coordinate moves from folded to unfolded protein. In the crystal structure of *Cp*cyt c_{552} (PDB: 4O1W), M60 and M62 show elevated B-factors (15 Å^2 for M57 vs 19 Å^2 for M60 and 39 Å^2 for M62), and superposition of all three dimers from the asymmetric unit show differences in the orientations of M60 and M62, but not of M57, M69, or M79. Measurement of NMR chemical shifts and spin relaxation experiments is a way to confirm differences among *Cp*cyt c_{552} methionine residues under solution conditions and quantitate the differences in backbone and sidechain flexibility that contribute to the observed protein stability (T_m) of *Cp*cyt c_{552}.

6.2 Protein Expression and Purification of Isotopically-Labeled *Cp*cyt c_{552}

To study using NMR spectroscopy, *Cp*cyt c_{552} produced with nitrogen (15N) and carbon (13C) isotopes incorporated was necessary. Traditional schemes for recombinant protein expression in *E. coli* with isotopically labeled precursors, where transformed bacteria first grown in rich media is adapted to H_2O-based M9 Minimal media before inducing with IPTG, did not lead to expression of *Cp*cyt c_{552}. Expression can be assessed in two ways: i) cell lysis followed by an SDS-PAGE gel or ii) a straightforward visual observation of the pellet, which appear dark red when *Cp*cyt c_{552} is overexpressed. Cell pellets using the method of above-described method did showed no red color.

After a comprehensive search of the literature, an expression protocol was obtained from a publication concerned with study of multi-heme proteins [116]. The publication identified that for cytochromes and heme-containing proteins in general, addition of 1mM aminolevulinic acid (ALA; Sigma-Aldrich) to the expression media was sufficient to see significant gains in expression yield of the target protein. ALA, which interestingly is a natural non-protein amino acid, is a precursor for porphyrin synthesis. Therefore, it follows naturally that its inclusion in the expression media would allow both cytochrome protein and heme synthesis to proceed at rates conducive to expression and folding of *Cp*cyt c_{552}.

The following is a quick summary of the successful expression and purification of *Cp*cyt c_{552}. *E. coli* BL21(DE3) cells were transformed with two plasmids: one plasmid contained the gene for *Cp*cyt c_{552} (Amp^{R}) and the other coded for a heme cassette (Cam^{R}). Colonies were used to inoculate overnight cultures (LB media plus 100 μg/mL Amp and 25 μg/mL). These small volume cultures were then used to inoculate one liter of LB media (plus antibiotics). Using the Marley method (Marley and Bracken citation), inoculated rich media was shaken at 37°C at 225 rpm until OD$_{600}$ read 0.7. At this optical density, bacterial cultures were pelleted in a centrifuge at 4K rpm, rich media decanted off the bacterial pellet, the pellets were resuspended in a M9 salt-only media then pelleted once more, and finally the pellets were resuspend in M9 Minimal expression media containing ^{15}N labeled ammonium chloride and ^{13}C labeled glucose. Note that the Marley method advises that the final volume of expression media be 4x smaller than the volume of rich media used for bacterial growth, e.g. if one liter of LB media was used to for bacterial growth, the bacterial pellet is resuspended in only 250 milliliters of M9 Minimal media. Finally, after allowing the resuspended BL21 cells in M9 Minimal media shake at 225 rpm for one hour at 30°C, protein expression was induced by adding 500 μM IPTG. Expression was allowed to proceed for 36-48 hours. The expression media was visibly red by this time, and centrifugation yields an unequivocally dark red pellet.

Purification of *Cp*cyt c_{552} began with cell lysis in a 50 mM ammonium acetate (pH 5.0) buffer, with DNase, RNase, and protease inhibitor, and sonication. After five minutes of sonication (15 seconds on: 30 seconds off), the lysate was cleared by centrifugation at 14K rpm for 30 minutes. Supernatant was applied to a cation exchange column (5 mL HiTrap SP HP; GE Life Sciences) and eluted with a linear gradient from zero to 500 mM sodium chloride (NaCl). Elution volumes with a dark red appearance were combined and oxidized with cobalt (III) ethylenediaminetetraacetate (Co^{III}EDTA) (citation in supplement of the first Magyar paper). Further *Cp*cyt c_{552} purification involved a high-resolution cation exchange column (Mono S 10/100; GE Life Sciences), with elution again over a linear gradient from zero to 500 mM sodium chloride (NaCl). Elution volumes with a red hue were combined and buffer exchanged from 50 mM acetate (pH 5.0) to 50 mM sodium phosphate buffer (pH 7.0). As a final step to ensure purity, the protein sample was applied to a size-exclusion column (Superdex 200 10/300 GL; GE Life Sciences). A monodisperse peak was observed, and the corresponding elution fractions were oxidized with potassium ferricyanide (Sigma-Aldrich) to form Fe^{III} *Cp*cyt c_{552}. MALDI-TOF mass spectrometry and UV-Vis spectrophotometry (Soret band at 411 nm and Fe-methionine(S) coordination at 698 nm) were used to confirm presence at Fe^{III} *Cp*cyt c_{552}.

6.3 NMR spectroscopy of *Cp*cyt c_{552}

*Cp*cyt c_{552} NMR samples contained 10% D_2O in 50 mM sodium phosphate buffer (pH 7.0) with 0.1% sodium azide and 100 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Additionally, samples were either reduced using excess sodium dithionite (to Fe^{II}) or oxidized using excess potassium ferricyanide (to Fe^{III}). Wilmad tubes of 500 μL protein sample or 5 mm Shigemi tubes of 350 μL protein sample were used for multidimensional NMR experiments.

NMR data were collected on reduced and oxidized uniformly ^{13}C, ^{15}N-labeled *Cp*cyt c_{552} at 14.1 T on a Bruker DRX600 console equipped with a triple-resonance z-axis gradient cryogenic probe. Sample temperature was calibrated to 293 K using 98% $^{2}H_{4}$-methanol [56]. In addition to 1D proton spectra and 2D ^{15}N sensitivity enhanced HSQC spectra, a comprehensive suite of three-dimensional backbone chemical shift assignment spectra was recorded. Table 6.3.1 lists the completed experiments for both reduced and oxidized *Cp*cyt c_{552}, as well as experimental details.

Oxidized Cpcyt c_{552}	**Spectral Width (kHz)**	**TD ($t_3 \times t_2 \times t_1$)**	**Scans (NS)**
HNCA	7.5 × 2 × 4.4	1024 × 80 × 64	16
HN(CO)CA	8 × 2 × 5	1024 × 80 × 128	8
HNCACB	7 × 2 × 9.4	1024 × 80 × 128	16
CBCA(CO)NH	6 × 1.5 × 9.4	1024 × 64 × 124	16
HNCO	7.8 × 2 × 2.2	1024 × 80 × 128	4
HN(CA)CO	7 × 1.8 × 3.1	1024 × 64 × 124	16
^{15}N HSQC	7 × 1.6	1024 × 256	8
Reduced Cpcyt c_{552}			
HN(CO)CA	6 × 1.8 × 5	1024 × 64 × 128	16
HNCACB	6 × 1.8 × 9.4	1024 × 64 × 128	16
CBCA(CO)NH	6 × 1.8 × 9.4	1024 × 64 × 124	16
HNCO	6.5 × 1.8 × 3.8	1024 × 64 × 128	8
^{15}N HSQC	7 × 1.9	1024 × 256	16

Table 6.3.1: ***Cp*cyt c_{552} Backbone Triple-Resonance Assignment Experiments.**
NMR backbone assignment experiments were used to assign the oxidized and reduced forms of *Cp*cyt c_{552}. Three dimensional spectra showed good S/N and resolution, which was especially valuable for assignment of the repetitive alanine fragments found in the protein construct. For oxidized ***Cp*cyt c_{552}**, peaks that demonstrate paramagnetic quenching in the ^{15}N HSQC also show reduced peak intensities in the carbon dimension of 3D spectra.

Chemical shifts were analyzed using the Chemical Shift Index 3.0 (CSI 3.0) of the Wishart group [117]. Backbone chemical shifts analyzed by the CSI 3.0 webserver included C_α, C_β, C_o, H_N, and N. In addition to returning secondary structure predictions, CSI 3.0 also outputs an estimate of the random coil index for each residue. For *Cp*cyt c_{552}, this serves as an initial test for differential dynamics among methionine residues and against the remaining backbone residues. Furthermore, the computed RCI can be converted to other measures of protein mobility, including order parameters (S^2), B-factors, and NMR RMSD.

6.4 Results

Initial attempts at isotope labeling of *Cp*cyt c_{552} failed to yield protein. Including ALA in the expression media and restricting bacteria growth to rich media for as long as possible proved vital for successful *Cp*cyt c_{552} expression. Porphyrin synthesis is a biochemically intensive process, and matching protein synthesis with its needed heme cofactor might be too burdensome in minimal nutrient media without spiking the solution with porphyrin precursor. The successful *Cp*cyt c_{552} expression scheme is depicted in Figure 6.4.1.

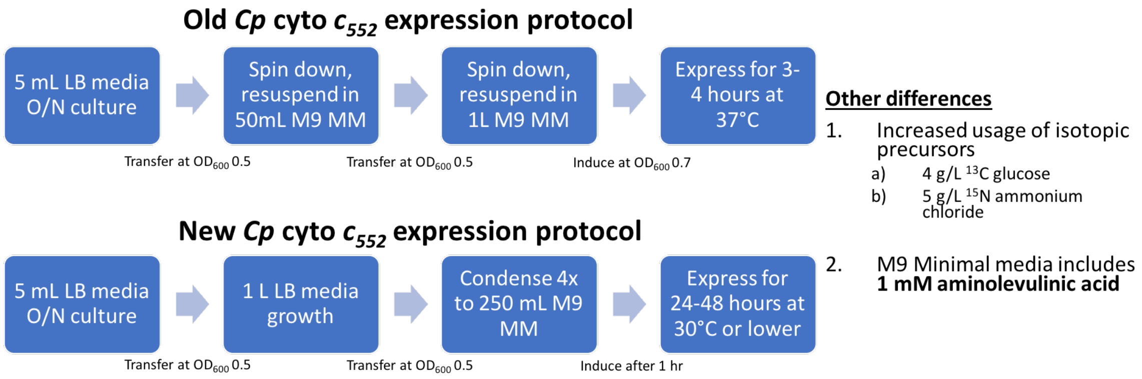


Figure 6.4.1: **Protocol for Isotopic Expression of *Cp*cyt c_{552}.**
Successful ^{15}N and ^{13}C isotope incorporation into *Cp*cyt c_{552} recombinantly expressed protein required both increased levels of isotopic precursor (^{13}C glucose and ^{15}N ammonium chloride) as well as inclusion of aminolevulinic acid (ALA), which is a precursor in the porphoryin synthesis pathway.

6.4.1 Comparison of Oxidized and Reduced *Cp*cyt c_{552} Spectral Properties

The *Cp*cyt c_{552} protein construct used for NMR experiments comprised 79 amino acids, with four prolines (Figure 6.4.2). Figures 6.4.3 and 6.4.4 show ^{15}N HSQC spectra corresponding to oxidized and reduced *Cp*cyt c_{552} samples, respectively. Both spectra show exceptional signal-to-noise and dispersion in both the proton and nitrogen dimension. For the oxidized version of *Cp*cyt c_{552}, all proton-nitrogen correlations, except for H17-A19 and T27-Y28, were assigned. The ^{15}N HSQC spectrum corresponding to the reduced form of *Cp*cyt c_{552} is completely assigned. Table 6.5.1 shows a table of backbone chemical shift assignments for both forms of *Cp*cyt c_{552}. For the majority of residues, if a proton-nitrogen correlation could be assigned, backbone C_α, C_β, and CO chemical shifts were typically assigned as well. Overall, approximately 90.8% of backbone chemical shifts were assigned for oxidized *Cp*cyt c_{552} and 97.4% of reduced *Cp*cyt c_{552}.backbone shifts were assigned.

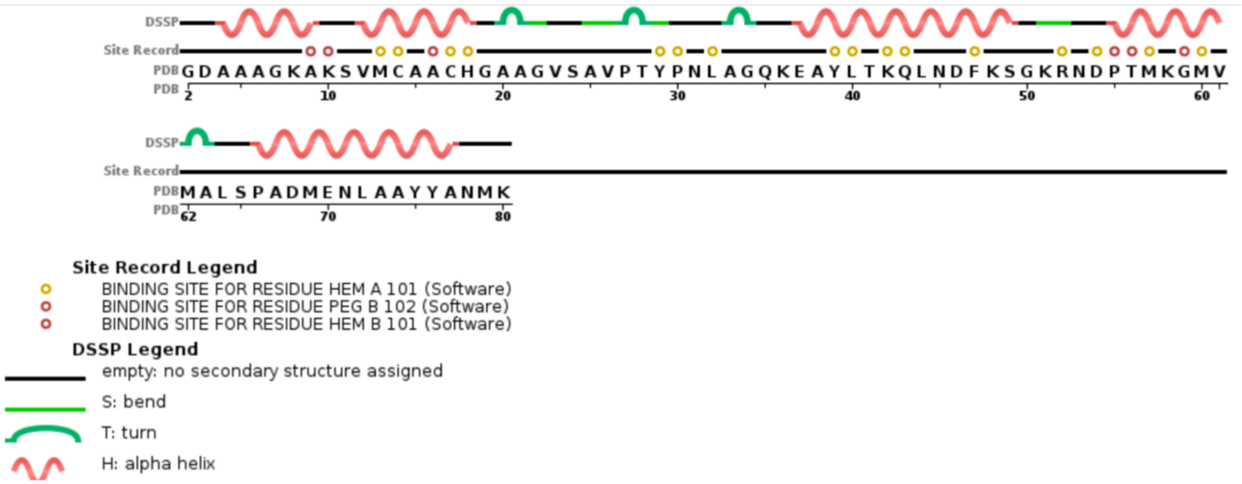


Figure 6.4.2: ***Cp*cyt *c*$_{552}$ Protein Construct Used for NMR Spectroscopy.**
The protein sequence from PDB 4O1W is the same protein construct used for NMR spectroscopy. The secondary structure found in 4O1W is visible above the protein sequence. Five α-helices are expected. Also, note the annotation of H19 and M57 as heme binding atoms (secondary structure diagram from PDF:4O1W; www.rcsb.org).

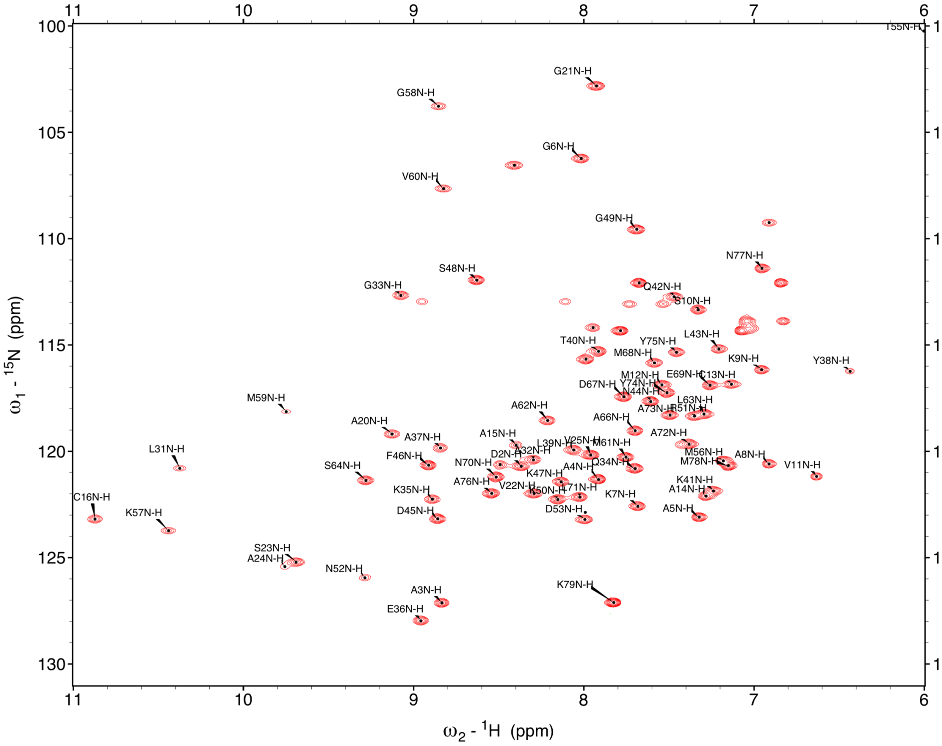


Figure 6.4.3: **^{15}N HSQC of Oxidized *Cp*cyt *c*$_{552}$.**

Approximately 91% of amide correlations were assigned for oxidized *Cp*cyt *c*$_{552}$. The residues that were unassigned clustered near the covalently-bound heme ligand and in the two-residue gap between P27 and P30. Paramagnetic relaxation effects are presumed to be responsible for quenching of the missing *Cp*cyt *c*$_{552}$ resonances.

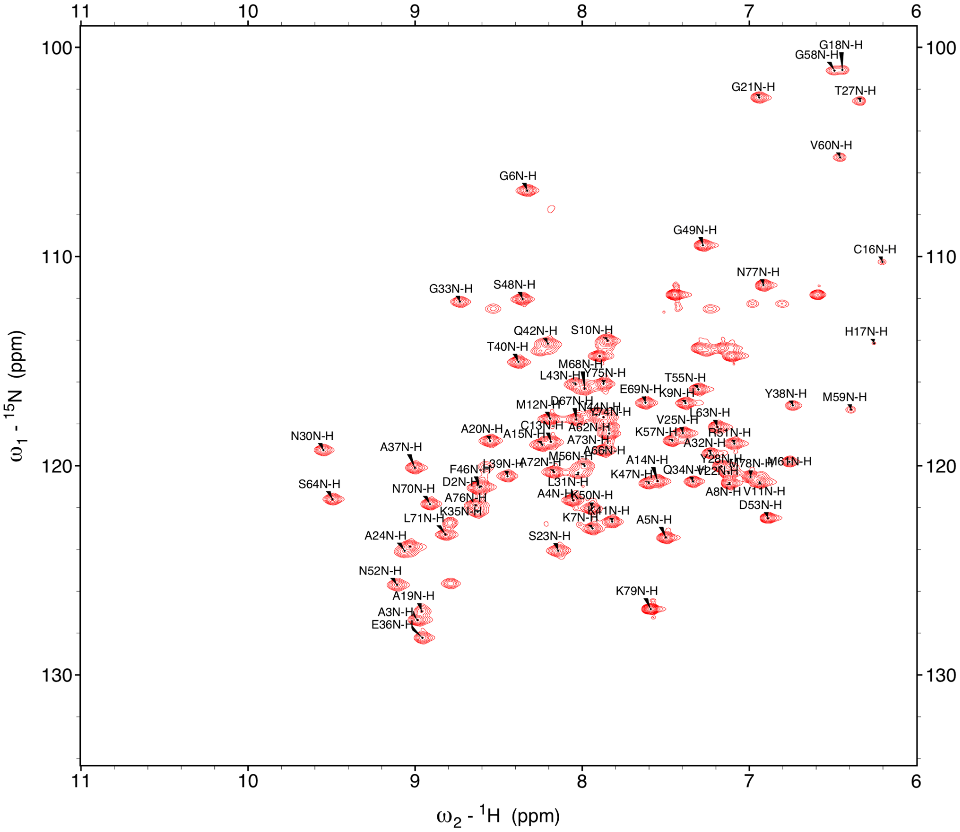


Figure 6.4.4: **^{15}N HSQC of Reduced *Cp*cyt *c*$_{552}$.**
The ^{15}N HSQC of reduced *Cp*cyt *c*$_{552}$ shows a complete assignment of amide resonances. Peak intensities are uniform across all amide correlations, and the reduced form of *Cp*cyt *c*$_{552}$ lacks the hyperfine shifts present in the oxidized spectrum.

Immediately obvious from comparison of these two spectra is the considerable peak shift between oxidized and reduced forms of the *Cp*cyt *c*$_{552}$. The ^{15}N HSQC footprint of reduced *Cp*cyt *c*$_{552}$ is highly characteristic of a globular folded protein. However, significant peak shifts are observed when compared the oxidized state of the protein. The largest chemical shift perturbations are localized to residues surrounding heme-ligating H18 and M57. For example, residue 16, which is a cysteine that is covalently bound to the heme cofactor and sits immediately adjacent to the axial ligand H17, is shifted furthest downfield of any amide correlation in the proton dimension. Likewise, K57, which borders the other axial ligand M58, is shifted almost as much downfield. The oxidized ^{1}H chemical shifts of both residues are greater than 10.5 ppm. Figure 6.4.5 gives a more global sense of chemical shift perturbations between the oxidized and reduced forms of *Cp*cyt c_{552}. Chemical shift perturbations (CSP) were calculated according to:

$$\Delta\delta_{HN} = \sqrt{(D_H)^2 + (0.15 \times D_N)^2}$$

The changes in electronic structure between Fe^{II} and Fe^{III} clearly manifest as paramagnetic-induced chemical shift perturbations, which are localized to the regions of amino acid residues surrounding both axial ligands of the heme iron.

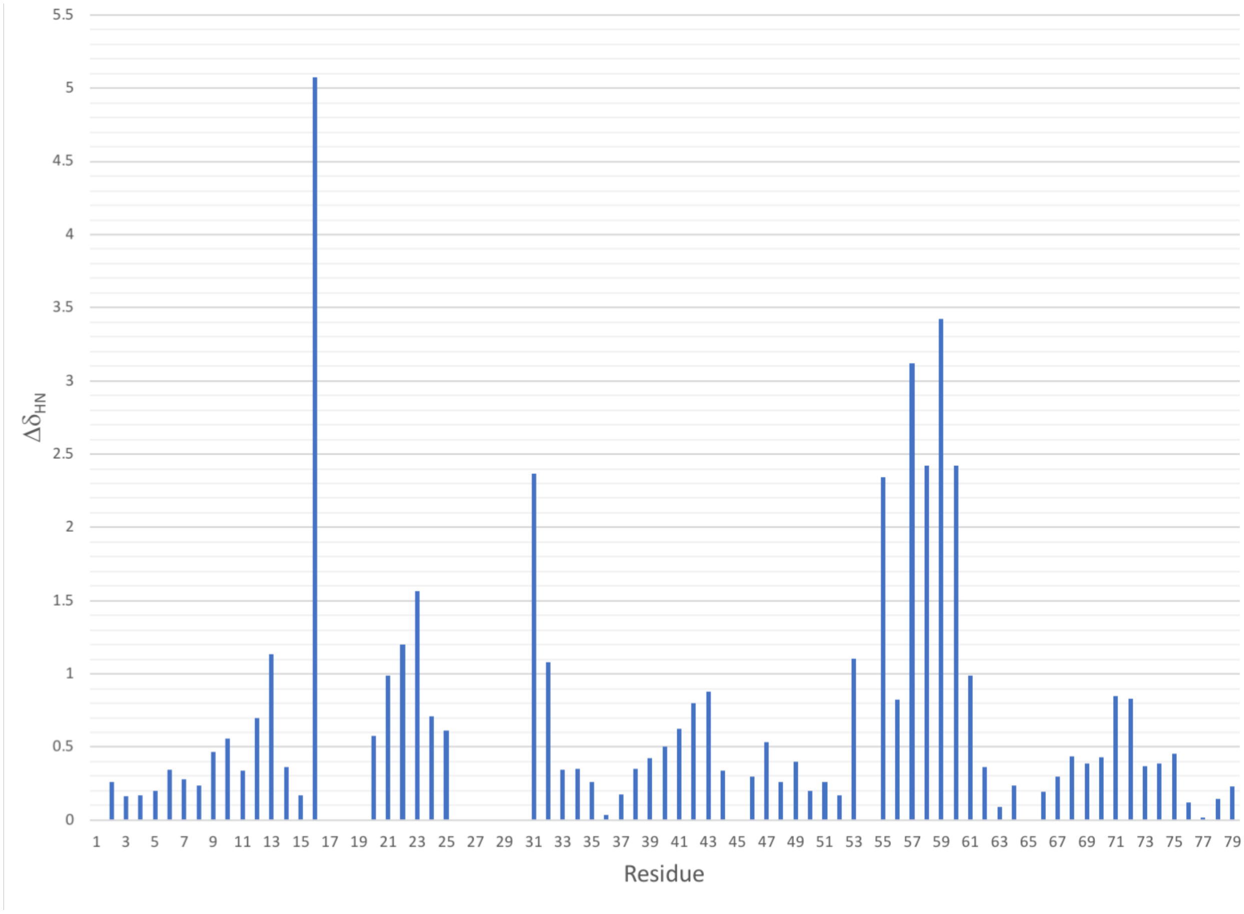


Figure 6.4.5: **Chemical Shift Perturbations between Oxidized and Reduced *Cp*cyt *c*$_{552}$.** The most significant CSPs occur for residues spatially near the heme cofactor. These include the residues near the covalently-bound heme (C14-C16) and surrounding the axial heme ligand. As opposed to the former, residues T55-M61 all show at least a 1 ppm ($\Delta\delta_{HN}$) CSP. The large perturbations in this region of *Cp*cyt *c*$_{552}$ suggest that M56, M59, and M61 are all in close proximity to the heme ligand.

6.4.2 Predictions of Protein Backbone Dynamics using RCI

Backbone resonance assignments, namely C$_\alpha$, C$_\beta$, C$_o$, H$_N$, and N backbone chemical shifts, were used to estimate values of the model-free order parameter (S^2) from the Random Coil Index (RCI). Figure 6.4.6 and 6.4.7 show the results of residue-specific RCI predictions. In these bar graphs, a larger RCI value for a residue corresponds to a more dynamic character. The result for reduced *Cp*cyt *c*$_{552}$ shows an interesting trend: the four residues immediately after M57 are predicted to have a large RCI compared the rest of the protein (Figure 6.4.6.A). Only N-terminal residues are predicted to have as large an RCI. Pivoting to Figure 6.4.6.B, prediction of S^2 clearly labels M60 and M62 as flexible residues compared to other methionine residues and the remaining *Cp*cyt c_{552} residues. Figure 6.4.7 gives the RCI and S^2 predictions for oxidized *Cp*cyt c_{552}. Here, the predictions are less clear. A stretch of amino acids between N44-S48 are predicted to exhibit larger RCI values and the lowest order parameters of all oxidized *Cp*cyt c_{552} residues.

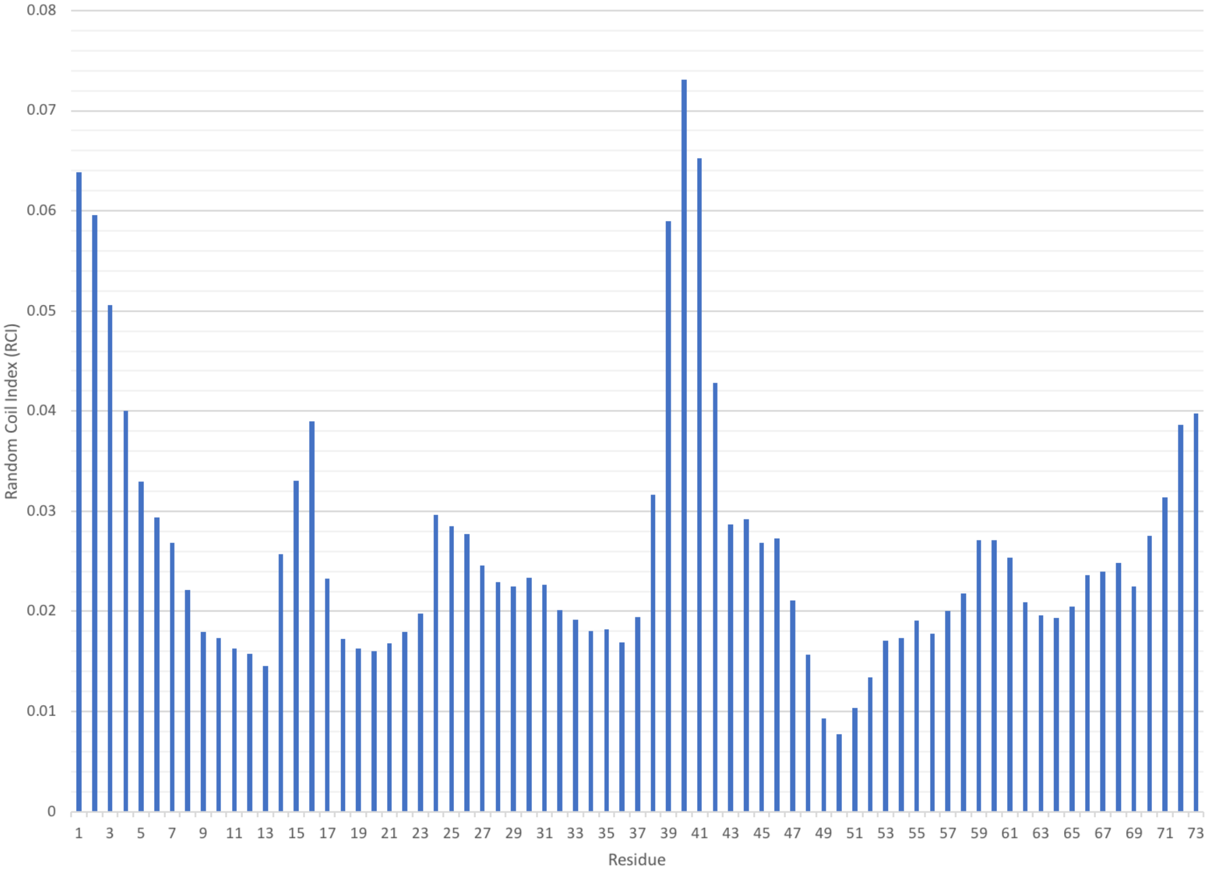


Figure 6.4.6: **Predictions of Random Coil Index for Oxidized *Cp*cyt c_{552}.** RCI predictions for oxidized *Cp*cyt c_{552} indicate only modest flexibility for M57 and M59. Instead, the N-terminal portion of α-helix 3 is predict1ed to be as dynamic as the N- and C-terminal portions of the protein. Additionally, the residues surrounding the covalently-bound heme cofactor, including C14-C17, are predicted to have elevated RCI values. As a caveat, it should be

noted that the RCI method is trained with protein chemical shifts that might not fully characterize redox proteins the way they do with non-redox globular proteins.

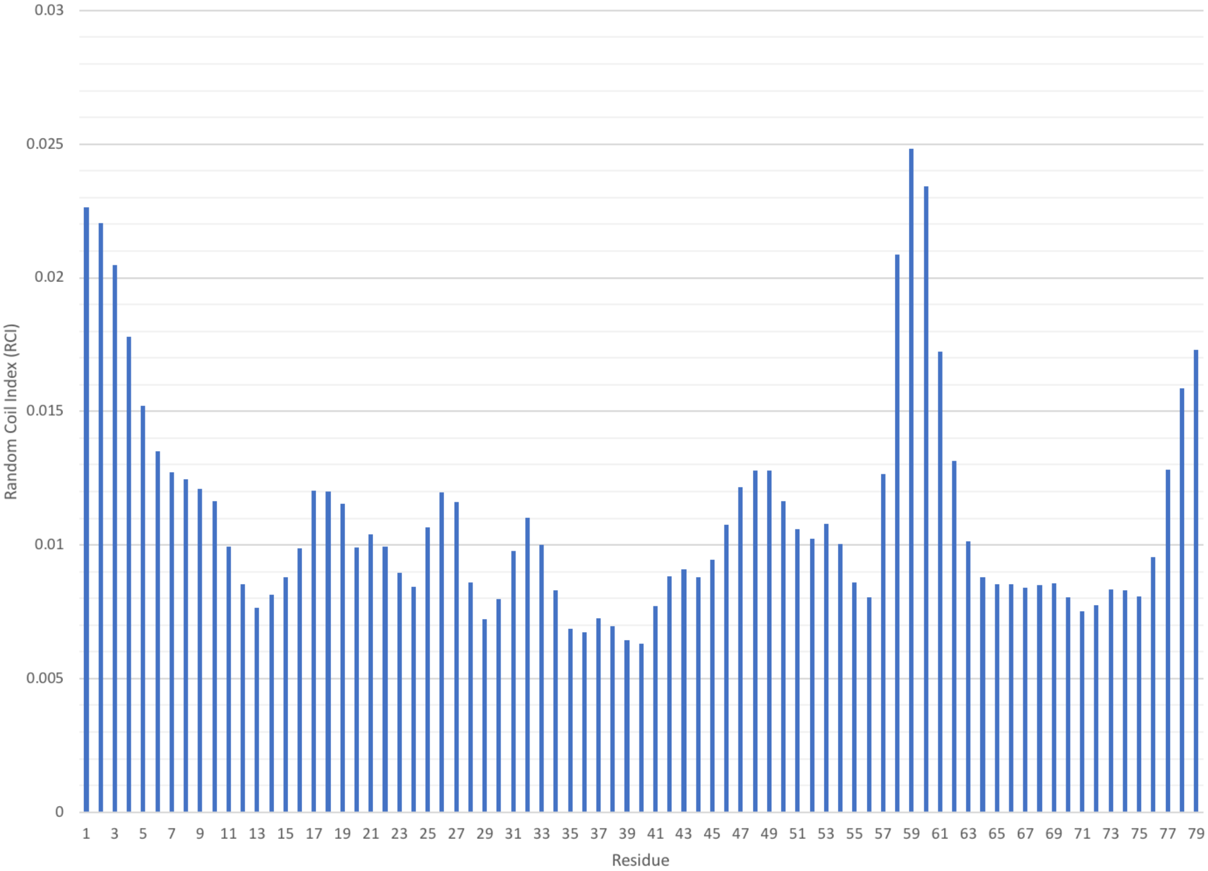


Figure 6.4.7: **Predictions of Random Coil Index for Reduced *Cp*cyt c_{552}.**
For reduced *Cp*cyt c_{552}, RCI predictions using backbone chemical shifts suggest the presence of elevated dynamics for residues K57, G58, M59, V60, M61, and A62. The predicted dynamical character is on par with the predicted N- and C-terminal residues. The larger RCI for M59 and M61 (~0.025 and 0.017) compared to M56 (0.008) adds support to the hypothesis that the dynamic character of non-axial ligand methionine residues as a potential way to substitute for the native axial ligand M56.

6.5 Discussion

Initial spectroscopy experiments with *Cp*cyt c_{552} resulted in near-complete backbone resonance assignment of both oxidized and reduced forms of protein. One important takeaway is that, despite the paramagnetic bleaching effect from the unpaired electron in the oxidized *Cp*cyt c_{552}, over 90%

of the backbone chemical shifts were assigned and only five backbone amides were not visible in a ^{15}N HSQC spectrum. Notably, three of those five amides range from the second cysteine residue covalently bound to heme to the residue immediately after the histidine that axially ligates the heme iron. It makes sense that these residues suffer paramagnetic relaxation effects and consequently have reduced peak intensities that renders them invisible to detection. An orthogonal readout of the change in protein electronic environment are CSPs. As shown in Figure 6.2.8, CSPs are exclusively localized to chemical shifts corresponding to protein residues in close proximity to heme. In aggregate, these results give confidence that the redox character of *Cp*cyt c_{552} is intact and observable, and NMR experiments that measure residue-specific (Met) kinetics are viable.

Chemical shift-based predictions also give support to the initial hypothesis. Based off B-factor analysis of the crystal structure and sequence alignment of psychrophilic/psychrotolerant cytochrome c, it was speculated that methionine sulfur substitution is responsible for the sustained protein stability (T^m) despite a decrease in enthalpy. RCI predictions for the reduced form of *Cp*cyt c_{552} demonstrate just that: M57 is predicted to be significantly ordered (RCI value of M57 equal to RCI values for α-helix residues), while the RCI value for M60 is >3x higher than M57 and the RCI value for M62 is ~2x higher than M57. Order parameters for M60 and M62 are 0.85 and 0.89, respectively, whereas S^2 for M57 is 0.95. Altogether, these predictions support the notion that the flexibility of specific methionine residues may contribute to the observed protein stability of *Cp*cyt c_{552}. Compared to both other methionine residues (M13 and M69) and the remaining backbone residues, the RCI and S^2 predictions for M60 and M62 stick out. While predictions for oxidized *Cp*cyt c_{552} don’t implicate the two methionine residues in increased flexibility the way chemical shifts from the reduced state do, one issue could be confounding our predictions. The prediction algorithm was trained on sets of globular proteins. Likely, redox proteins were not especially enriched in this test set. Furthermore, extracting information about dynamics encoded into paramagnetic-affected chemical shifts may not be viable.

To confirm the increased local protein dynamics of M60 and M62, NMR spin relaxation will be performed. First, heteronuclear steady-state {^{1}H}-^{15}N NOE experiments will be recorded on oxidized and reduced *Cp*cyt c_{552}. These 2D NMR experiments will allow quick confirmation of the relative flexibility within the suspect methionine residues (M57, M60, M62) and compared to the remaining residues. Second, the full suite of NMR spin relaxation experiments will be implemented to measure R_1 and R_2 relaxation rate constants. Together with the {^{1}H}-^{15}N NOE, model-free order parameters and rate constants will be derived from the experimental rates. From this, the localized dynamics of M57, M60, and M62 will be well described. Finally, a selectively-labeled *Cp*cyt c_{552} U-[Met-^{13}CH$_3$] NMR will be used to measure methionine side-chain dynamics. Examples exist in the literature demonstrating use of methionine ^{13}C$^{\varepsilon}$-(H$^{\varepsilon}$)$_3$ methyl groups to probe side-chain dynamics in fast (10^{-10}-10^{-9} sec) and slow (10^{-6}-10^{-3} sec) timescale regimes [118-120]. ^{1}H-^{1}H cross-related relaxation experiments measure fast timescale dynamics, while measurement with a Hahn-echo experiment of double- and zero-quantum relaxation rates (R_{ZQ} and R_{DQ}) allows characterization of dynamics in the micro- to millisecond timescale. As a first test for the viability of Met methyl experiment, *Cp*cyt c_{552} U-[Met-^{13}CH$_3$] protein in a protonated background was expressed in *E. coli* and purified to homogeneity. A ^{1}H-^{13}C SOFAST-HMQC [58] spectrum was recorded and the results are shown in Figure 6.5.1. Six methionine ^{1}H$^{\varepsilon}$-^{13}C$^{\varepsilon}$ peaks are expected from the primary sequence, but only five peaks are visible in the recorded spectrum. The missing methionine methyl correlation is most likely M57. A 1D ^{1}H spectrum of oxidized *Cp*cyt c_{552} is presented in Figure 6.2.12. The M57 ^{1}H$^{\varepsilon}$-^{13}C$^{\varepsilon}$ correlation is hyperfine shifted to -11 ppm. Thus, five Met peaks are leftover for five methionine residues (M13, M60, M62, M69, and M79). Among these four Met peaks, M60 is the most dynamic residue according to our predictions and potentially corresponds to the peak in Figure 6.2.11 marked with an “X” (^{1}H 2.95 ppm, ^{13}C 19.1 ppm). However, M79 is the penultimate residue in *Cp*cyt c_{552}, and the pronounced exchange broadening in the proton dimension for peak “X” could be indicative of its position at the protein C-terminal and the accompanying dynamics that stem from absence of secondary structure. Assignment of these Met methyl peaks will require a combination of H(CCC)(CO)NH and ^{13}C aliphatic NOESY experiments, which will confirm these initial observations and also provide more chemical shift information for use in prediction of protein dynamics using side-chain RCI.

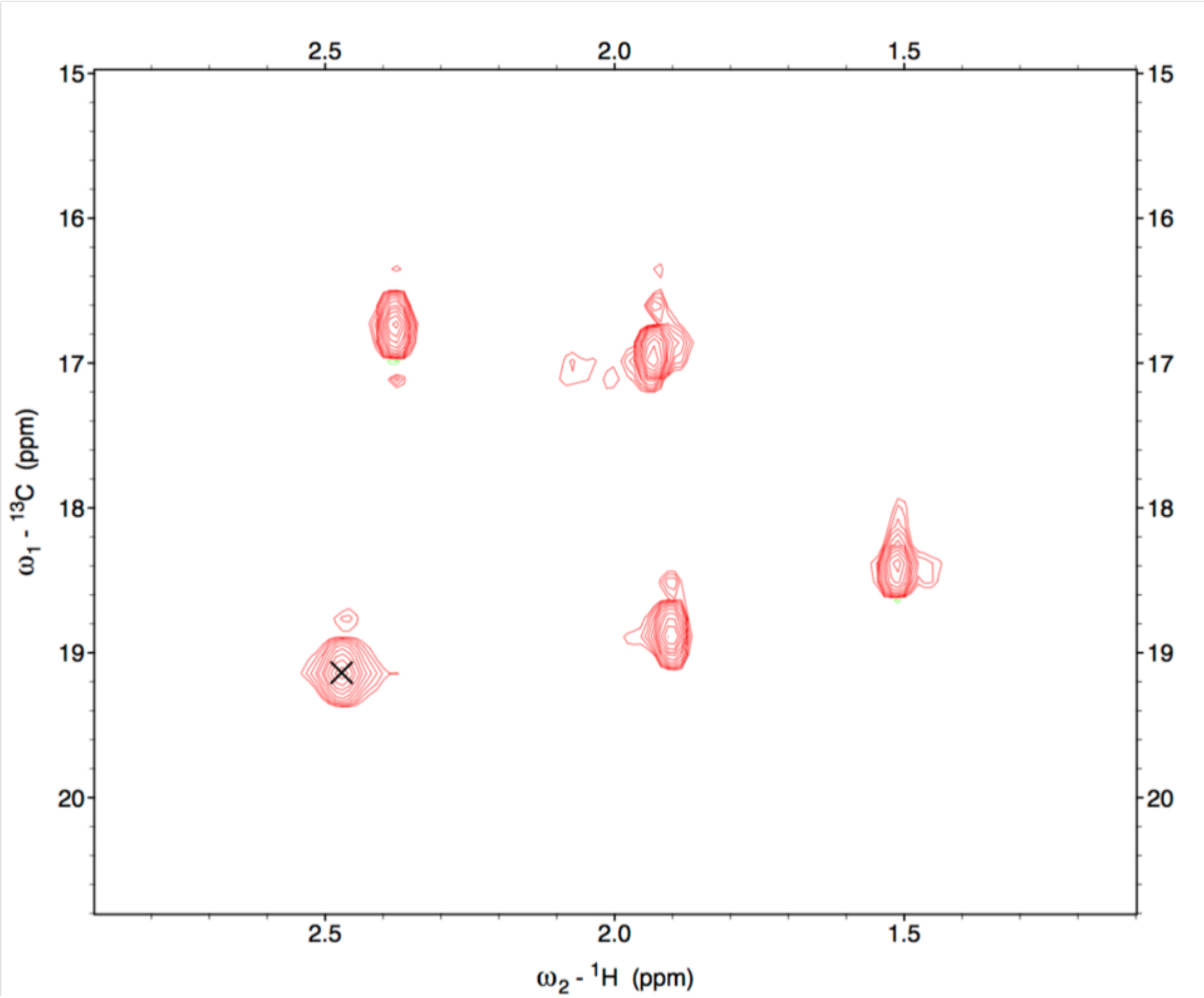


Figure 6.5.1: **Methyl Region from a ^{13}C HMQC of U-[Met-^{13}CH$_3$] Oxidized *Cp*cyt *c*$_{552}$.** ^{13}C SOFAST HMQC experiments of an oxidized Met-^{13}CH$_3$ labeled *Cp*cyt *c*$_{552}$ show five ^{1}H$^{\varepsilon}$-^{13}C$^{\varepsilon}$ correlations. The protein construct used for NMR experiments contains six methionine residues. The missing methionine residue is the axial heme ligand M57. As shown in Figure 6.2.12, the M57 ^{1}H$^{\varepsilon}$-^{13}C$^{\varepsilon}$ correlation is hyperfine shifted upfield to -11 ppm at 285 K. At 293 K, the M57 methyl should resonate at about -10 ppm; increasing the spectral width in both the proton and carbon dimension in the about spectrum still did not allow observation of this shift. The peak labeled with an "**X**" has the largest linewidth in the proton dimension; the observed linebroadening could be due to increased dynamics from shuffling of axial methionine residues (M57, M60, M62) or lack of secondary structure in the protein C-terminal (M79).

Lastly, parallel studies of unlabeled (only ^{1}H) *Cp*cyt c_{552} were conducted by Sarish Lone of the Magyar group at Barnard College. One dimensional ^{1}H NMR experiments were recorded for oxidized and reduced *Cp*cyt c_{552} samples. Past work has shown hyperfine shifts for the four heme methyl groups and the axial methionine methyl of cytochromes and heme-containing proteins [121, 122]. The ^{1}H NMR spectrum of reduced *Cp* cytochrome yielded an amide footprint typical of a folded protein. No hyperfine proton peaks were observed as hyperfine shifted upfield or downfield. However, as visible in Figure 6.5.2, a one-dimensional proton spectrum of oxidized *Cp*cyt c_{552} shows four downfield-shifted protons (two peaks near 30 ppm called H1 and H2, one peak at ~20 ppm called H3, and one peak at ~18 ppm called H4) presumably originating from the heme cofactor and one broad, upfield-shifted proton peak assumed to come from the methyl group of the *Cp*cyt c_{552} axial M57 residue (~ -11 ppm). The temperature dependence of the heme methyl and M57 methyl shifts was measured, and the resulting Curie plot (1000/T vs ^{1}H chemical shift) is viewed in Figure 6.5.3. Across a temperature range 285 K to 320 K, H1 and H2 heme protons barely change chemical shift and exhibited a non-Curie dependence. However, both H3 and H4 are shifted downfield by approximately 3 ppm, while the M57 methyl shift moves downfield by 10 ppm and the linewidth of the peak narrows noticeably. The larger temperature dependence of the M57 methyl protons is indicative of increased relative flexibility of the residue at temperatures that approach the T_m of *Cp*cyt c_{552}. Additionally, the temperature dependence of M57 is classified as Curie-type dependence and confirms the presence of high spin Fe^{II}. This serves as yet another piece of evidence supporting a unique cytochrome stabilization mechanism whereby M57, M60, and M62 share the position of axial heme ligand as *Cp*cyt c_{552} unfolds, thus introducing intermediate local minima in the protein free energy landscape and preventing immediate unfolding of *Cp*cyt c_{552}.

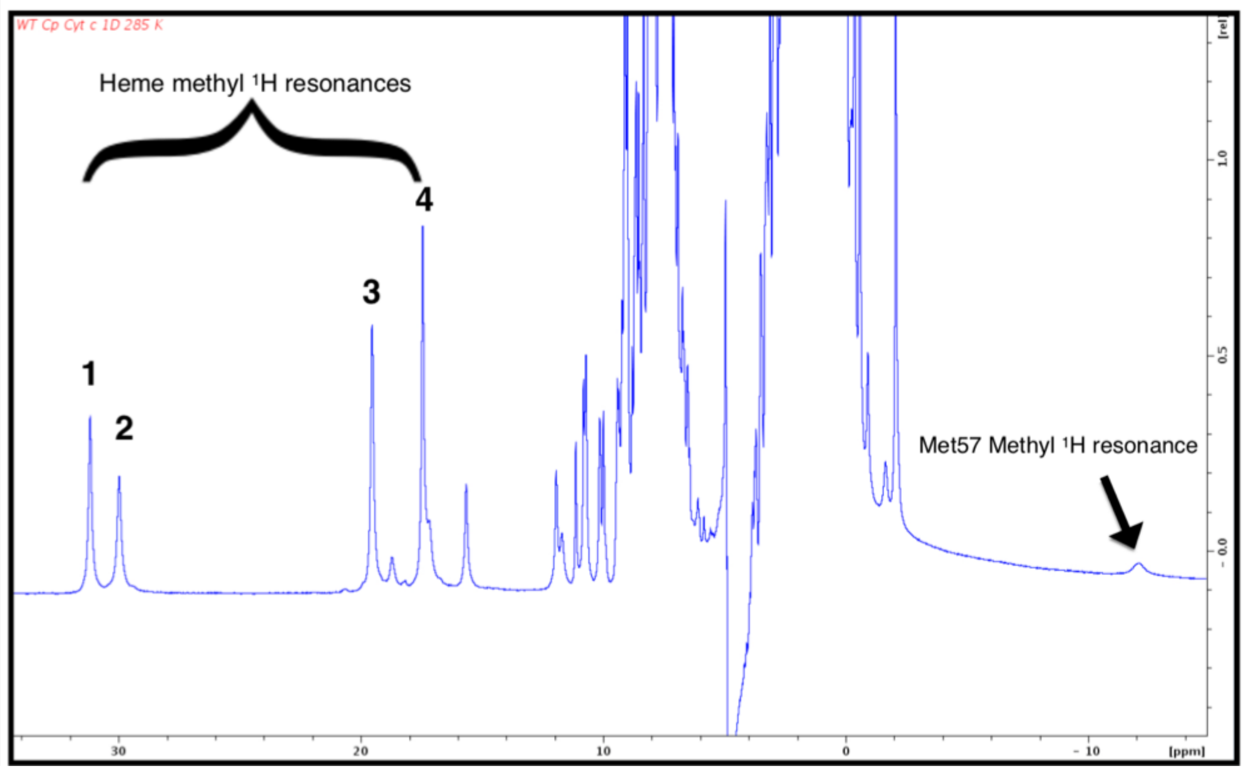


Figure 6.5.2: **1D ^{1}H NMR Spectrum of Oxidized *Cp*cyt c_{552} at 285 K.**
This 1D proton spectrum of oxidized *Cp*cyt c_{552} was acquired using an excitation sculpting pulse sequence and uncharacteristically large spectral width (~70 ppm). The four heme methyl proton resonances are annotated as H1-H4. As observed in studies of cytochrome c_{552} from other organisms, the methyl group of the axial ligand methionine resonates significantly downfield of the typical protein methyl chemical shift. M56 is observed to resonate at -11 ppm, and demonstrates significant peak broadening.

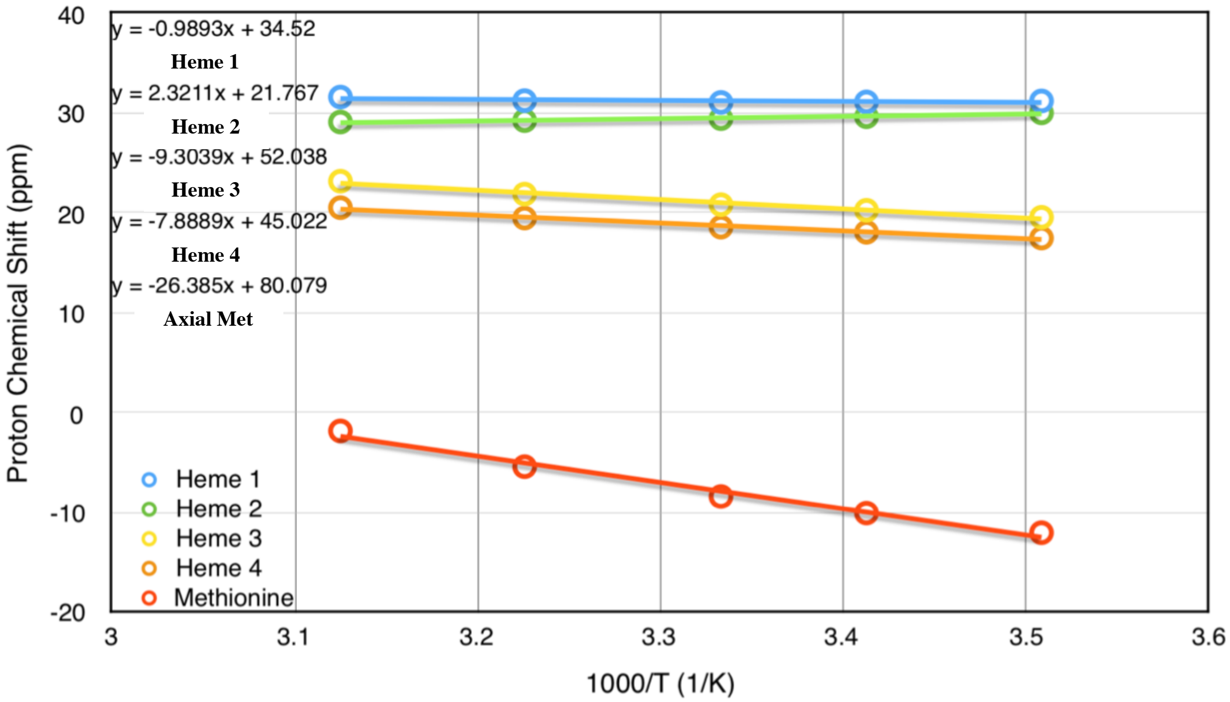


Figure 6.5.3: **Curie Plots of Temperature Dependence on Chemical Shifts for Oxidized *Cp*cyt *c*$_{552}$ between 285 K and 320 K.**

Curie plots show the reciprocal temperature dependence of ^{1}H chemical shift. For both heme methyl protons H3 and H4, the downward slope of their regression lines and positive chemical shift values indicate an anti-Curie dependence. The axial methionine proton, though, shows a Curie-type dependence. This indicates the presence of high spin Fe^{II}.

Table 6.5.1: **Oxidized and Reduced *Cp*cyt *c*$_{552}$ Backbone Triple-Resonance Assignments.**

Oxidized Backbone Assignments

Residue	C (ppm)	CA (ppm)	CB (ppm)	H (ppm)	N (ppm)
G1	165.1	39.87	-	7.978	-
D2	172.8	49.15	38.62	8.351	120.7
A3	176.4	52.42	16.17	8.825	127.1
A4	177.9	52.23	15.02	7.898	121.3
A5	178.7	51.2	15.37	7.31	123.1
G6	170.9	43.08	-	7.999	106.2
K7	177.7	56.06	29.31	7.669	122.6
A8	176.5	51.87	15.13	6.892	120.6
K9	174.3	52.58	29.09	6.934	116.2
S10	172.4	58.65	59.86	7.303	113.3

V11	175.2	62.19	28.49	6.606	121.2
M12	173	54.49	30.17	7.514	116.8
C13	174.6	50.41	31.1	7.095	116.8
A14	176.5	52.1	16.39	7.257	122.1
A15	175.4	52.48	15.86	8.376	119.7
C16	-	-	-	10.89	123.2
A19	175.3	52.58	15.8	-	-
A20	174.7	47.92	15.9	9.123	119.2
G21	171.5	45.77	-	7.928	102.9
V22	174.2	59.57	30.91	8.304	122
S23	172.6	55.62	63.71	9.704	125.2
A24	174.5	49.48	17.73	9.751	125.4
V25	173.3	56.4	32.01	7.953	120.2
N30	175.1	51.93	36.5	-	-
L31	173.7	52.87	42.96	10.39	120.8
A32	177.1	51.74	16.29	8.301	120.4
G33	171	43.29	-	9.065	112.7
Q34	170.8	51.99	28.36	7.692	120.8
K35	176.3	52.37	31.12	8.885	122.3
E36	175.8	57.86	27.52	8.941	128
A37	177.8	52.09	14.86	8.834	119.8
Y38	173.7	58.53	34.92	6.41	116.3
L39	175.1	55.65	40.2	8.033	119.9
T40	173.2	64.46	-	7.887	115.3
K41	175	56.59	29.68	7.213	121.9
Q42	174.2	54.59	24.79	7.436	112.7
L43	174.8	56	37.29	7.177	115.2
N44	175.2	54.49	36.55	7.584	117.6
D45	177.5	54.18	36.8	8.839	123.2
F46	176.3	57.6	38	8.908	120.7
K47	175.2	57.13	30.54	8.129	121.5
S48	173.5	56.22	61.29	8.617	111.9
G49	171.3	42.61	-	7.68	109.6
K50	174.2	55.89	30.23	8.139	122.3
R51	172.1	52.55	28.98	7.344	118.4
N52	171.8	50.96	37.25	9.278	125.9
D53	171.2	51.22	43.56	7.994	123.1
P54	172.4	62.46	30.05	-	-

T55	171.3	58.11	65.42	5.964	103.5
M56	168.6	46.8	33.72	7.164	120.5
K57	177	59.45	30.63	10.49	123.8
G58	174.9	45.19	-	8.881	103.8
M59	176	54.17	32.34	9.813	118.1
V60	175	58.93	30.38	8.853	107.7
M61	174.4	55.49	29.06	7.748	120.3
A62	174.6	48.73	16.29	8.213	118.5
L63	173.2	50.73	39.97	7.283	118.2
S64	170.6	53.41	60.55	9.266	121.4
P65	176.7	63.51	28.96	-	-
A66	176.1	51.93	15.21	7.685	119
D67	176.1	54.63	38.97	7.745	117.4
M68	173.8	57.54	29.48	7.56	115.8
E69	176.7	55.72	26.64	7.237	116.9
N70	174.8	52.85	33.43	8.495	121.2
L71	174.6	55.17	39.5	7.991	122.1
A72	175.7	52.5	17.76	7.349	119.6
A73	176.5	52.15	14.96	7.471	118.3
Y74	174.3	57.57	34.84	7.491	117.2
Y75	176.4	60.28	35.49	7.437	115.3
A76	176.3	52.66	17.78	8.521	122
N77	172.1	50.76	36.92	6.936	111.4
M78	172	54.27	30.4	7.138	120.7
K79	178.5	54.9	31.46	7.82	127.1

Reduced Backbone Assignments

Residue	C (ppm)	CA (ppm)	CB (ppm)	H (ppm)	N (ppm)
G1	177.2	42.81	-	-	-
D2	173	52.16	41.51	8.609	121
A3	176.6	55.3	18.84	8.987	127.3
A4	178.1	55.16	17.82	8.054	121.7
A5	179	54.14	18.25	7.501	123.4
G6	171.2	46.39	-	8.331	106.8
K7	177.9	59.12	32.26	7.941	123
A8	176.9	55.1	18.05	7.124	120.8
K9	175	55.77	32.38	7.383	117
S10	172.7	61.77	62.9	7.849	114

V11	175.6	65.16	31.33	6.94	120.8
M12	174	58.2	33.89	8.196	117.7
C13	175.1	55.16	35.45	8.188	118.8
A14	176	54.06	18.85	7.55	120.7
A15	174.6	54.84	18.8	8.24	119
C16	171.2	54.49	35.83	6.206	110.2
H17	168.3	55.62	26.98	6.254	114.1
G18	172.9	43.67	-	6.443	101.1
A19	174.9	54.9	18.3	8.961	127
A20	173.9	50.08	18.29	8.552	118.8
G21	169.9	47.2	-	6.943	102.4
V22	173	61.02	32.84	7.128	120.4
S23	172.1	57.2	64.89	8.148	124.1
A24	174.3	52.05	20.21	9.068	124.1
V25	-	58.77	34.66	7.399	118.5
P26	172.2	64.89	32.89	-	-
T27	171.5	62.15	69.11	6.338	102.6
Y28	-	50.05	36.57	7.171	120
P29	172.7	61	29.31	-	-
N30	173.4	52.71	37.99	9.553	119.2
L31	173.2	53.99	43.94	8.023	120.3
A32	176.5	53.81	18.34	7.233	119.4
G33	170.5	45.52	-	8.734	112.1
Q34	170.6	54.34	30.91	7.34	120.8
K35	176.4	54.97	33.68	8.628	122.2
E36	176.2	60.7	30.4	8.959	128.2
A37	178	55.2	17.96	9.006	120.1
Y38	174.3	62	38.16	6.743	117.1
L39	175.9	59.08	43.5	8.45	120.5
T40	173.8	67.76	68.38	8.386	115
K41	175.4	59.93	32.98	7.823	122.7
Q42	174.5	57.99	27.66	8.21	114.1
L43	175.2	59.02	41.24	8.044	116.1
N44	-	57.48	39.41	7.922	117.5
D45	165.3	56.94	39.49	-	-
F46	175.8	59.24	38.9	8.618	121.1
K47	174.9	59.05	32.8	7.606	120.8
S48	173.2	58.7	63.71	8.359	112

G49	171	45.11	-	7.28	109.5
K50	173.9	58.29	32.9	7.947	122
R51	171.8	55.27	31.44	7.093	118.9
N52	171.3	53.29	39.6	9.113	125.7
D53	-	53.16	45.79	6.891	122.5
P54	175.7	64.9	31.95	-	-
T55	173.7	65.48	66.95	7.306	116.3
M56	174.8	52.62	23.13	7.987	120
K57	175.2	59.87	31.67	7.464	118.8
G58	173.2	46.36	-	6.495	101.1
M59	175.1	54.91	30.04	6.393	117.3
V60	174.8	61.02	31.7	6.457	105.3
M61	174.1	58.05	31.34	6.763	119.8
A62	174.6	51.43	19.04	7.852	118.1
L63	173.4	53.62	42.97	7.195	118.2
S64	-	56.19	63.37	9.497	121.6
P65	176.9	66.43	31.8	-	-
A66	176.5	55.04	18.16	7.871	119.3
D67	176.5	57.88	42	8.037	117.8
M68	174.4	60.78	32.55	7.987	116.3
E69	177.2	59.01	29.61	7.625	117
N70	175.3	55.97	36.51	8.913	121.9
L71	175.4	58.91	43.13	8.82	123.3
A72	176.1	55.78	20.92	8.172	120.3
A73	176.8	55.17	17.84	7.841	118.5
Y74	174.4	60.47	37.79	7.871	117.7
Y75	176.4	63.22	38.17	7.874	116.1
A76	176.3	55.33	20.55	8.641	121.9
N77	172	53.44	39.58	6.919	111.4
M78	171.8	56.85	32.66	6.995	120.5
K79	-	57.23	33.96	7.592	126.8

Chapter 7

Conclusions and Future Directions

In Chapter 3, data was presented that indicates the first Zn finger of GTSF1 specifically binds an RNA motif of *GGUUC(G/A)*. To assess this, an RNA fragment encompassing the T-arm of tRNA was titrated with the 1-115 residue fragment of GTSF1. Our working hypothesis is that the function of GTSF1 might first require binding of RNA to the Zn finger 1, which then leaves Zn finger 2 poised for binding to another molecular species. tRNA-derived fragments, including 3' and 5' portions, have only recently been implicated in silencing of transposable elements in mammalian cells. In addition to using site-directed mutagenesis (specifically, alanine scanning) to probe the [GTSF1 Zn finger 1]:RNA interface, NMR dynamics experiments will be used to characterize the Zn finger 2 residues under conditions where RNA is bound. Under apo- and RNA-bound conditions, relaxation experiments measuring magnetization less-sensitive to signal loss due to slower molecular tumbling might allow the judgment of increased ordering of Zn finger 2. Such an observation would lend support to a folding-upon-binding mechanism for GTSF1 and might allow discovery of the true *in vivo* binding partner(s) and mechanism of GTSF1.

As covered in the end of Chapter 6, 13-carbon relaxation experiments are currently being explored for characterization of *Cp*cyt methionine methyl groups. Initial experiments will focus on assignment of methionine ^{1}H-^{13}C correlation using ^{13}C -NOESY or relayed scalar coupling experiments. With only six methyl correlations to assign, measurement of pico-nanosecond and micro-millisecond dynamics should shortly follow.

References

1. Plateau, P. and M. Gueron, *Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences*. Journal of the American Chemical Society, 1982. **104**(25): p. 7310-7311.

2. Sklenář, V. and A. Bax, *Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra*. Journal of Magnetic Resonance (1969), 1987. **74**(3): p. 469-479.

3. Szewczak, A.A., G.W. Kellogg, and P.B. Moore, *Assignment of NH resonances in nucleic acids using natural abundance 15N-1H correlation spectroscopy with spin-echo and gradient pulses*. FEBS Letters, 1993. **327**(3): p. 261-264.

4. Damblon, C., et al., *1H−15N HMQC for the Identification of Metal-Bound Histidines in 113Cd-Substituted Bacillus cereus Zinc β-Lactamase*. Journal of the American Chemical Society, 1999. **121**(49): p. 11575-11576.

5. Kay, L.E., et al., *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins*. Journal of Magnetic Resonance (1969), 1990. **89**(3): p. 496-514.
6. Bochkarev, A., et al., *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA*. Nature, 1997. **385**(6612): p. 176-81.

7. Farrow, N.A., et al., *Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer*. Biochemistry, 1995. **34**(3): p. 868-878.

8. Tollinger, M., et al., *Slow Dynamics in Folded and Unfolded States of an SH3 Domain*. Journal of the American Chemical Society, 2001. **123**(46): p. 11341-11352.

9. Palmer, A.G., *NMR Characterization of the Dynamics of Biomacromolecules*. Chemical Reviews, 2004. **104**(8): p. 3623-3640.
10. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**: p. 806.

11. Parrish, S., et al., *Functional Anatomy of a dsRNA Trigger: Differential Requirement for the Two Trigger Strands in RNA Interference*. Molecular Cell, 2000. **6**(5): p. 1077-1087.
12. Ambros, V., *The functions of animal microRNAs*. Nature, 2004. **431**: p. 350.

13. Iwasaki, Y.W., M.C. Siomi, and H. Siomi, *PIWI-Interacting RNA: Its Biogenesis and Functions*. Annual Review of Biochemistry, 2015. **84**(1): p. 405-433.

14. Ghildiyal, M. and P.D. Zamore, *Small silencing RNAs: an expanding universe*. Nature Reviews Genetics, 2009. **10**: p. 94.

15. Meister, G., *Argonaute proteins: functional insights and emerging roles*. Nature Reviews Genetics, 2013. **14**: p. 447.

16. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions*. Nature Reviews Genetics, 2009. **10**: p. 155.

17. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units*. Genome Res, 2004. **14**(10A): p. 1902-10.

18. Watanabe, T., et al., *Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes*. Nature, 2008. **453**: p. 539.

19. Mlotshwa, S., G.J. Pruss, and V. Vance, *Small RNAs in viral infection and host defense*. Trends in Plant Science, 2008. **13**(7): p. 375-382.

20. Xia, H., et al., *siRNA-mediated gene silencing in vitro and in vivo*. Nature Biotechnology, 2002. **20**: p. 1006.

21. Onodera, Y., et al., *Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation*. Cell, 2005. **120**(5): p. 613-622.
22. Aravin, A.A., G.J. Hannon, and J. Brennecke, *The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race*. Science, 2007. **318**(5851): p. 761.

23. Zeng, Y., R. Yi, and B.R. Cullen, *Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha*. The EMBO Journal, 2005. **24**(1): p. 138.

24. Lee, Y.S., et al., *Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways*. Cell, 2004. **117**(1): p. 69-81.

25. Lim, A.K. and T. Kai, *Unique germ-line organelle, nuage, functions to repress selfish genetic elements in <em>Drosophila melanogaster</em&gt*. Proceedings of the National Academy of Sciences, 2007. **104**(16): p. 6714.
26. Ipsaro, J.J., et al., *The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis*. Nature, 2012. **491**: p. 279.

27. Nishimura, T., et al., *PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development*. EMBO reports, 2018.

28. Montgomery, T.A., et al., *PIWI Associated siRNAs and piRNAs Specifically Require the Caenorhabditis elegans HEN1 Ortholog henn-1*. PLOS Genetics, 2012. **8**(4): p. e1002616.

29. Czech, B. and G.J. Hannon, *One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing*. Trends in Biochemical Sciences, 2016. **41**(4): p. 324-337.

30. Beyret, E., N. Liu, and H. Lin, *piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism*. Cell Research, 2012. **22**: p. 1429.

31. Malone, C.D., et al., *Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary*. Cell, 2009. **137**(3): p. 522-535.

32. Ipsaro, J.J. and L. Joshua-Tor, *From guide to target: molecular insights into eukaryotic RNA-interference machinery*. Nature Structural &Amp; Molecular Biology, 2015. **22**: p. 20.

33. Djuranovic, S., A. Nahvi, and R. Green, *miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay*. Science, 2012. **336**(6078): p. 237.

34. Kim, D.H., et al., *MicroRNA-directed transcriptional gene silencing in mammalian cells*. Proceedings of the National Academy of Sciences, 2008.

35. Morris, K.V., et al., *Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells*. Science, 2004. **305**(5688): p. 1289.

36. McManus, M.T. and P.A. Sharp, *Gene silencing in mammals by small interfering RNAs*. Nature Reviews Genetics, 2002. **3**: p. 737.

37. Peters, L. and G. Meister, *Argonaute Proteins: Mediators of RNA Silencing*. Molecular Cell, 2007. **26**(5): p. 611-623.

38. Nishida, K.M., et al., *Gene silencing mechanisms mediated by Aubergine–piRNA complexes in Drosophila male gonad*. RNA, 2007.

39. Le Thomas, A., et al., *Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state*. Genes & Development, 2013.

40. Houwing, S., et al., *A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish*. Cell, 2007. **129**(1): p. 69-82.

41. Sienski, G., D. Dönertas, and J. Brennecke, *Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression*. Cell, 2012. **151**(5): p. 964-980.

42. Muerdter, F., et al., *A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in Drosophila*. Molecular Cell, 2013. **50**(5): p. 736-748.

43. Li, C., et al., *Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies*. Cell, 2009. **137**(3): p. 509-521.

44. Aravin, A.A., et al., *A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice*. Molecular Cell, 2008. **31**(6): p. 785-799.

45. Watanabe, T., et al., *Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus*. Science, 2011. **332**(6031): p. 848.

46. Das, P.P., et al., *Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the Caenorhabditis elegans Germline*. Molecular Cell, 2008. **31**(1): p. 79-90.

47. Wang, G. and V. Reinke, *A C. elegans Piwi, PRG-1, Regulates 21U-RNAs during Spermatogenesis*. Current Biology, 2008. **18**(12): p. 861-867.

48. Ashe, A., et al., *piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans*. Cell, 2012. **150**(1): p. 88-99.

49. Coustham, V., et al., *The C. elegans HP1 homologue HPL-2 and the LIN-13 zinc finger protein form a complex implicated in vulval development*. Developmental Biology, 2006. **297**(2): p. 308-322.

50. Andreeva, A. and H. Tidow, *A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes*. Bioinformatics, 2008. **24**(20): p. 2277-2280.

51. Czech, B., et al., *A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway*. Molecular Cell, 2013. **50**(5): p. 749-761.

52. Dönertas, D., G. Sienski, and J. Brennecke, *Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex*. Genes & Development, 2013. **27**(15): p. 1693-1705.

53. Ohtani, H., et al., *DmGTSF1 is necessary for Piwi–piRISC-mediated transcriptional transposon silencing in the Drosophila ovary*. Genes & Development, 2013. **27**(15): p. 1656-1661.

54. Lepre, C.A., J.M. Moore, and J.W. Peng, *Theory and Applications of NMR-Based Screening in Pharmaceutical Research*. Chemical Reviews, 2004. **104**(8): p. 3641-3676.

55. Lepre, C.A. and J.M. Moore, *Microdrop screening: A rapid method to optimize solvent conditions for NMR spectroscopy of proteins*. Journal of Biomolecular NMR, 1998. **12**(4): p. 493-499.

56. Findeisen, M., T. Brand, and S. Berger, *A 1H-NMR thermometer suitable for cryoprobes*. Magnetic Resonance in Chemistry, 2006. **45**(2): p. 175-178.

57. Lakomek, N.-A., J. Ying, and A. Bax, *Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods*. Journal of Biomolecular NMR, 2012. **53**(3): p. 209-221.

58. Schanda, P., Ē. Kupče, and B. Brutscher, *SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds*. Journal of Biomolecular NMR, 2005. **33**(4): p. 199-211.

59. Van Nostrand, E.L., et al., *Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)*. Nature Methods, 2016. **13**: p. 508.

60. Jensen, M.R., R.W.H. Ruigrok, and M. Blackledge, *Describing intrinsically disordered proteins at atomic resolution by NMR*. Current Opinion in Structural Biology, 2013. **23**(3): p. 426-435.

61. Mittag, T. and J.D. Forman-Kay, *Atomic-level characterization of disordered protein ensembles*. Current Opinion in Structural Biology, 2007. **17**(1): p. 3-14.

62. Krężel, A. and W. Maret, *Dual Nanomolar and Picomolar Zn(II) Binding Properties of Metallothionein*. Journal of the American Chemical Society, 2007. **129**(35): p. 10911-10921.

63. van Dongen, E.M.W.M., et al., *Variation of Linker Length in Ratiometric Fluorescent Sensor Proteins Allows Rational Tuning of Zn(II) Affinity in the Picomolar to Femtomolar Range*. Journal of the American Chemical Society, 2007. **129**(12): p. 3494-3495.

64. Wishart, D.S. and B.D. Sykes, *The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data*. Journal of Biomolecular NMR, 1994. **4**(2): p. 171-180.

65. Barraud, P., M. Schubert, and F.H.T. Allain, *A strong 13C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins*. Journal of Biomolecular NMR, 2012. **53**(2): p. 93-101.

66. Shen, Y. and A. Bax, *Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N*, in *Artificial Neural Networks*, H. Cartwright, Editor. 2015, Springer New York: New York, NY. p. 17-32.

67. Rieping, W., et al., *ARIA2: Automated NOE assignment and data integration in NMR structure calculation*. Bioinformatics, 2007. **23**(3): p. 381-382.

68. Tidow, H., et al., *Solution Structure of the U11-48K CHHC Zinc-Finger Domain that Specifically Binds the 5′ Splice Site of U12-Type Introns*. Structure, 2009. **17**(2): p. 294-302.

69. Wilkinson, M.L., et al., *The 2′-O-methyltransferase responsible for modification of yeast tRNA at position 4*. RNA, 2007.

70. Schorn, A.J., et al., *LTR-Retrotransposon Control by tRNA-Derived Small RNAs*. Cell, 2017. **170**(1): p. 61-71.e11.

71. Chapman, K.B., A.S. Byström, and J.D. Boeke, *Initiator methionine tRNA is essential for Ty1 transposition in yeast*. Proceedings of the National Academy of Sciences, 1992. **89**(8): p. 3236.

72. Xu, H. and J.D. Boeke, *Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency*. Proceedings of the National Academy of Sciences, 1990. **87**(21): p. 8360.

73. Keam, S.P., et al., *The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells*. Nucleic Acids Research, 2014. **42**(14): p. 8984-8995.

74. Peng, J.W. and G. Wagner, *Frequency Spectrum of NH Bonds in Eglin c from Spectral Density Mapping at Multiple Fields*. Biochemistry, 1995. **34**(51): p. 16733-16752.

75. Abyzov, A., et al., *Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation*. Journal of the American Chemical Society, 2016. **138**(19): p. 6240-6251.

76. Gill, M.L., R.A. Byrd, and I.I.I.A.G. Palmer, *Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region*. Physical Chemistry Chemical Physics, 2016. **18**(8): p. 5839-5849.

77. Hsu, A., et al., *Enhanced spectral density mapping through combined multiple-field deuterium 13CH2D methyl spin relaxation NMR spectroscopy*. Methods, 2018. **138-139**: p. 76-84.

78. Zhu, G., et al., *Protein Dynamics Measurements by TROSY-Based NMR Experiments*. Journal of Magnetic Resonance, 2000. **143**(2): p. 423-426.

79. Favier, A. and B. Brutscher, *Recovering lost magnetization: polarization enhancement in biomolecular NMR*. Journal of Biomolecular NMR, 2011. **49**(1): p. 9-15.

80. Ferrage, F., D. Cowburn, and R. Ghose, *Accurate Sampling of High-Frequency Motions in Proteins by Steady-State 15N–{1H} Nuclear Overhauser Effect Measurements in the Presence of Cross-Correlated Relaxation*. Journal of the American Chemical Society, 2009. **131**(17): p. 6048-6049.

81. Gairí, M., et al., *An optimized method for 15N R1 relaxation rate measurements in non-deuterated proteins*. Journal of Biomolecular NMR, 2015. **62**(2): p. 209-220.

82. O'Brien, P.A. and A.G. Palmer, *TROSY pulse sequence for simultaneous measurement of the 15N R1 and {1H}–15N NOE in deuterated proteins*. Journal of Biomolecular NMR, 2018. **70**(4): p. 205-209.

83. Delaglio, F., et al., *NMRPipe: A multidimensional spectral processing system based on UNIX pipes*. Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.

84. Nussinov, R., et al., *Intrinsic protein disorder in oncogenic KRAS signaling*. Cellular and Molecular Life Sciences, 2017. **74**(17): p. 3245-3261.

85. Csizmok, V., et al., *Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling*. Chemical Reviews, 2016. **116**(11): p. 6424-6462.

86. Dethoff, E.A., et al., *Functional complexity and regulation through RNA dynamics*. Nature, 2012. **482**: p. 322.

87. Kay, L.E., *New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy*. Journal of Molecular Biology, 2016. **428**(2, Part A): p. 323-331.

88. Narayanan, C., et al., *Applications of NMR and computational methodologies to study protein dynamics*. Archives of Biochemistry and Biophysics, 2017. **628**: p. 71-80.

89. Millet, O., et al., *Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 1. Measurement of Five Relaxation Rates per Deuteron in 13C-Labeled and Fractionally 2H-Enriched Proteins in Solution*. Journal of the American Chemical Society, 2002. **124**(22): p. 6439-6448.

90. Skrynnikov, N.R., O. Millet, and L.E. Kay, *Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions*. Journal of the American Chemical Society, 2002. **124**(22): p. 6449-6460.

91. Sheppard, D., et al., *Deuterium Spin Probes of Backbone Order in Proteins: 2H NMR Relaxation Study of Deuterated Carbon α Sites*. Journal of the American Chemical Society, 2009. **131**(43): p. 15853-15865.

92. Tugarinov, V. and L.E. Kay, *A 2H NMR Relaxation Experiment for the Measurement of the Time Scale of Methyl Side-Chain Dynamics in Large Proteins*. Journal of the American Chemical Society, 2006. **128**(38): p. 12484-12489.

93. Vallurupalli, P. and L.E. Kay, *A Suite of 2H NMR Spin Relaxation Experiments for the Measurement of RNA Dynamics*. Journal of the American Chemical Society, 2005. **127**(18): p. 6893-6901.

94. Vallurupalli, P., et al., *New RNA Labeling Methods Offer Dramatic Sensitivity Enhancements in 2H NMR Relaxation Spectra*. Journal of the American Chemical Society, 2006. **128**(29): p. 9346-9347.

95. Ishima, R. and K. Nagayama, *Quasi-Spectral-Density Function Analysis for Nitrogen-15 Nuclei in Proteins*. Journal of Magnetic Resonance, Series B, 1995. **108**(1): p. 73-76.

96. Johnson, E., W.J. Chazin, and M. Rance, *Effects of Calcium Binding on the Side-chain Methyl Dynamics of Calbindin D9k: A 2H NMR Relaxation Study*. Journal of Molecular Biology, 2006. **357**(4): p. 1237-1252.

97. Ishima, R., et al., *Comparison of Methyl Rotation Axis Order Parameters Derived from Model-Free Analyses of 2H and 13C Longitudinal and Transverse Relaxation Rates Measured in the Same Protein Sample*. Journal of the American Chemical Society, 2001. **123**(25): p. 6164-6171.

98. Tadokoro, T. and S. Kanaya, *Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes*. The FEBS Journal, 2009. **276**(6): p. 1482-1493.

99. Nowotny, M., *Retroviral integrase superfamily: the structural perspective*. EMBO reports, 2009. **10**(2): p. 144.

100. Moelling, K. and F. Broecker, *The reverse transcriptase–RNase H: from viruses to antiviral defense*. Annals of the New York Academy of Sciences, 2015. **1341**(1): p. 126-135.

101. Hollien, J. and S. Marqusee, *Comparison of the folding processes of T. thermophilus and E. coli Ribonucleases H11Edited by C. R. Matthews*. Journal of Molecular Biology, 2002. **316**(2): p. 327-340.

102. Butterwick, J.A., et al., *Multiple Time Scale Backbone Dynamics of Homologous Thermophilic and Mesophilic Ribonuclease HI Enzymes*. Journal of Molecular Biology, 2004. **339**(4): p. 855-871.

103. Katayanagi, K., et al., *Three-dimensional structure of ribonuclease H from E. coli*. Nature, 1990. **347**: p. 306.

104. Tugarinov, V., V. Kanelis, and L.E. Kay, *Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy*. Nature Protocols, 2006. **1**: p. 749.

105. Tugarinov, V., J.E. Ollerenshaw, and L.E. Kay, *Probing Side-Chain Dynamics in High Molecular Weight Proteins by Deuterium NMR Spin Relaxation: An Application to an 82-kDa Enzyme*. Journal of the American Chemical Society, 2005. **127**(22): p. 8214-8225.

106. Yang, W., et al., *Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from Escherichia coli*. Journal of Biological Chemistry, 1990. **265**(23): p. 13553-13559.

107. Mandel, A.M., M. Akke, and I.I.I.A.G. Palmer, *Backbone Dynamics ofEscherichia coliRibonuclease HI: Correlations with Structure and Function in an Active Enzyme*. Journal of Molecular Biology, 1995. **246**(1): p. 144-163.

108. Kroenke, C.D., et al., *Longitudinal and Transverse 1H−15N Dipolar/15N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules*. Journal of the American Chemical Society, 1998. **120**(31): p. 7905-7915.

109. Yamazaki, T., M. Yoshida, and K. Nagayama, *Complete assignments of magnetic resonances of ribonuclease H from Escherichia coli by double- and triple-resonance 2D and 3D NMR spectroscopies*. Biochemistry, 1993. **32**(21): p. 5656-5669.

110. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy*. Bioinformatics, 2015. **31**(8): p. 1325-1327.

111. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity*. Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.

112. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results*. Journal of the American Chemical Society, 1982. **104**(17): p. 4559-4570.

113. Lee, L.K., et al., *Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation*. Journal of Biomolecular NMR, 1997. **9**(3): p. 287-298.

114. Oswald, V.F., et al., *Overexpression, purification, and enthalpy of unfolding of ferricytochrome c552 from a psychrophilic microorganism*. Journal of Inorganic Biochemistry, 2014. **131**: p. 76-78.

115. Harvilla, P.B., H.N. Wolcott, and J.S. Magyar, *The structure of ferricytochrome c552 from the psychrophilic marine bacterium Colwellia psychrerythraea 34H*. Metallomics, 2014. **6**(6): p. 1126-1130.

116. Fernandes, A.P., et al., *Isotopic labeling of c-type multiheme cytochromes overexpressed in E. coli*. Protein Expression and Purification, 2008. **59**(1): p. 182-188.

117. Berjanskii, M.V. and D.S. Wishart, *A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts*. Journal of the American Chemical Society, 2005. **127**(43): p. 14970-14971.

118. Ergel, B., et al., *Protein Dynamics Control the Progression and Efficiency of the Catalytic Reaction Cycle of the Escherichia coli DNA-Repair Enzyme AlkB*. Journal of Biological Chemistry, 2014.

119. Skrynnikov, N.R., et al., *Probing Slow Time Scale Dynamics at Methyl-Containing Side Chains in Proteins by Relaxation Dispersion NMR Measurements: Application to Methionine Residues in a Cavity Mutant of T4 Lysozyme*. Journal of the American Chemical Society, 2001. **123**(19): p. 4556-4566.

120. London, R.E. and J. Avitabile, *Calculated carbon-13 NMR relaxation parameters for a restricted internal diffusion model. Application to methionine relaxation in dihydrofolate reductase*. Journal of the American Chemical Society, 1978. **100**(23): p. 7159-7165.

121. Keller, R.M., K. Wüthrich, and P.G. Debrunner, *Proton magnetic resonance reveals high-spin iron (II) in ferrous cytochrome P450 cam from Pseudomonas putida*. Proceedings of the National Academy of Sciences of the United States of America, 1972. **69**(8): p. 2073-2075.

122. Banci, L., et al., *Analysis of the Temperature Dependence of the 1H and 13C Isotropic Shifts of Horse Heart Ferricytochrome c: Explanation of Curie and Anti-Curie Temperature Dependence and Nonlinear Pseudocontact Shifts in a Common Two-Level Framework*. Journal of the American Chemical Society, 1998. **120**(33): p. 8472-8479.