

THE APPLICATION OF NUCLEAR MAGNETIC RESONANCE TO PROBE THE
STRUCTURE AND FUNCTION OF FUSION PEPTIDES OF THE INFLUENZA VIRUS
AND THE HUMAN IMMUNODEFICIENCY VIRUS

By

Yijin Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2024

ABSTRACT

Membrane-enveloped viruses have protein spikes that include a “fusion peptide” (Fp) segment that binds the target membrane of a host cell and plays a critical role in fusion (joining) of viral and target membranes. For influenza virus, this is subunit 2 of hemagglutinin which has a ~20-residue N-terminal fusion peptide region that binds target membrane. The Fp of human immunodeficiency virus (HIV) is the ~23 N-terminal residues of the glycoprotein 41 kD (gp41) subunit of the gp160 spike complex. Although the fusion mechanism of class I enveloped virus is relatively well-understood, researchers continue making efforts to reveal the full detailed picture of the fusion.

My studies related to influenza fusion peptide aims to provide answer to an outstanding question which is whether there are associated membrane changes important for fusion. Several computational studies have found increased “protrusion” of lipid acyl chains near Fp, i.e. one or more chain carbons are closer to the aqueous region than the headgroup phosphorus. Protrusion may accelerate initial joining of outer leaflets of the two membranes into a stalk intermediate. In this study, higher protrusion probability in membrane with vs. without Fp is convincingly detected by larger Mn^{2+} -associated increases in chain ^{13}C NMR transverse relaxation rates (Γ_2 's). Data analysis provides a ratio $\Gamma_{2,neighbor}/\Gamma_{2,distant}$ for lipids neighboring vs. more distant from the Fp. The calculated ratio depends on the number of Fp-neighboring lipids and the experimentally-derived range of 4 to 24 matches the range of increased protrusion probabilities from different simulations. For samples either with or without Fp, the Γ_2 values are well-fitted by an exponential decay as the ^{13}C site moves closer to the chain terminus. The decays correlate with free-energy of protrusion proportional to the number of protruded $-CH_2$ groups, with free energy per $-CH_2$ of ~ 0.25 $k_B T$. The NMR data support one major fusion role of the Fp to be much greater protrusion of lipid chains, with highest protrusion probability for chain regions closest to the headgroups.

Unlike Fp of influenza virus adopting α helical structure, the Fp of HIV adopts predominant intermolecular antiparallel β sheet structure when mole fraction cholesterol $\gg 0.3$ which is comparable to host cell fractions. The V2E engineered mutation near the N-terminus of the Fp greatly reduces gp160-mediated cell-cell fusion and gp41-induced vesicle fusion. To explore the broad population distribution of HIV Fp, REDOR NMR was applied to determine the registries (alignments) of adjacent Fp molecules in membrane-bound Fp. REDOR dephasing probed proximity between a backbone ^{13}CO label at a specific residue in one Fp molecule and backbone ^{15}N labels in adjacent Fp's at a different residue. For both WT and V2E, REDOR was measured

for 17 differently-labeled Fp's by Dr. Scott Schmick and Dr. Li Xie and the data then analyzed by me to quantitatively-determine the fractional populations, $f(t)$'s, of individual antiparallel registries indexed by t , the number of Fp residues in the sheet starting from the N-terminus. Both the WT and V2E sheets contained broad distributions of populated registries that included $t=11-20,22,23$ for WT and $t=15-21,23$ for V2E, with $\langle t \rangle^{\text{WT}} = 16.2$ and $\langle t \rangle^{\text{V2E}} = 18.5$. The $f(t)^{\text{WT}}$ values were well-fitted to free energies, $G(t)^{\text{WT}}$, that were sums of favorable contributions including one proportional to sheet length, another for registries in which Leu's were aligned in adjacent Fp's, and a third proportional to free energy of sidechain membrane insertion. The $f(t)^{\text{V2E}}$'s were similarly well-fitted except there wasn't the insertion contribution. Non-inserted V2E Fp is one basis for reduced fusion, and another is that longer V2E sheets result in shorter C-terminal hairpins, with consequent larger distances between initial apposed membranes.

The structural information of inclusion bodies (IBs) of recombinant protein (Rp) grown in bacterial host is another interest of my research. IBs is intracellular solid aggregates, a byproduct of growing cells in bacterial systems. The IBs fraction is often discarded because solubilization and subsequent refolding is difficult. There is little information about the structure of any Rp in IBs and such information may be useful for developing better solubilization and refolding. Because of this gap in knowledge, solid-state NMR was used to obtain structural information about a 109-residue "HM" Rp produced in bacteria. Several 2D and 3D ^{15}N - ^{13}C NMR correlation spectra of ^{13}C and ^{15}N labeled HM have been recorded and the assignment of the spectral crosspeaks based on amino acid type supported that there exist major α helical and minor β sheet structure in HM IBs.

Copyright by
YIJIN ZHANG
2024

Dedicated to mom and dad

ACKNOWLEDGEMENTS

I would like to seize this opportunity to thank my advisor, Dr. David Weliky, for giving me a chance to join his research group as a Ph.D student to explore NMR-related projects. I thank him for his countless guidance and support through my Ph.D career. He taught me the NMR theory and the most important thing, how to solve the scientific problem step by step. He has been very supportive for all the experiments I wanted to do. I really appreciate all the instruction he gave to me to make me have a better understanding of NMR theory.

I want to thank my guidance committee: Prof. John McCracken, Prof. James Geiger and Prof. Xiangshu Jin for their valuable inputs, guidance, and time for the committee meetings.

I am grateful for the great help I received from Dr. Dan Holmes and Dr. Li Xie from Max T Rogers NMR facility at Michigan state university. They have always been very helpful for my research. Whenever I had NMR issues, they were always willing to spent hours in subbasement to help me out. Without them, I was not able to acquire NMR data to finish my Ph.D.

I appreciate the chance to be a member of Weliky group. I would never forget the support and help from former and current group members, Dr. Ujjayini Ghosh, Dr. Ahinsa Ranaweera, Dr. Robert Wolfe, Noel Chau, Md Rokonujjaman, Forkan Saroar and Tahmina Khatun. I want to specially thank Dr. Ujjayini Ghosh for her great impact to my research career. I learned a lot of useful NMR techniques when I worked with her, and she has been very patient and insightful for my NMR theoretical and operational questions. She was generous for giving me suggestions about my career. I really appreciated the two years we worked together which inspired me to keep pursuing my Ph.D degree.

I would also want to thank Prof. Katherine Severin for her permission and instruction to use atomic absorption spectrum and Prof. Heedok Hong for his permission to use ultra-centrifuge. I thank Prof. Tuo Wang for his generosity to lend me a tripeptide standard, which plays a critical role in my research.

At the end, I want to express my appreciation to my family. I thank my parents for their endless support throughout my life. They have been very supportive for every choice I have made. I thank my brother and sister-in-law for their support and love when I was upset. I would also like to thank all the friends for their support. Their company made my Ph.D life much easier.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	viii
Chapter 1 Introduction	1
REFERENCES.....	47
Chapter 2 Materials and methods.....	53
REFERENCES.....	62
Chapter 3 Multidimensional ssNMR correlation experiment optimization	63
REFERENCES.....	83
Chapter 4 Lipid acyl chain protrusion induced by the influenza virus hemagglutinin fusion peptide detected by NMR paramagnetic relaxation enhancement.....	84
REFERENCES.....	106
Chapter 5 NMR assignment and structural probing of a small protein HM in bacterial inclusion bodies	113
REFERENCES.....	157
Chapter 6 Very broad and different distributions of antiparallel β sheet registries of the wild-type and fusion-defective V2E mutant of the membrane-bound HIV fusion peptide and roles of these distributions in: (1) mutational robustness of fusion for HIV under constant immune pressure; and (2) loss of fusion and infection with V2E mutation.....	161
REFERENCES.....	198
Chapter 7 Summary and future work.....	205
REFERENCES.....	209
APPENDIX A NMR FILE LOCATION.....	210
APPENDIX B SUPPORTING INFORMATION FOR CHAPTER 4	213
APPENDIX C SSNMR SPECTRA AND THS FLUORESCENCE DATA FOR HM.....	220
APPENDIX D SUPPORTING INFORMATION FOR REDOR DATA ANALYSIS	244

LIST OF ABBREVIATIONS

AIDS	acquired immunodeficiency syndrome
APS	ammonium persulfate
CAS	chemical shift anisotropy
CHR	C-heptad repeat
CP	cross polarization
DARR	dipolar assisted rotational resonance
DSS	trimethylsilylpropanesulfonate
Ed	ectodomain
Env	HIV enveloped protein
EPR	electron paramagnetic resonance
FID	free induction decay
Fp	Influenza fusion peptide in chapter 4 and HIV fusion peptide in chapter 6
FuSu	fusion subunit
FWHM	full-width at half-maximum
gHa2	influenza hemagglutinin protein subunit 2
gp120	HIV glycoprotein 120 kD
gp160	HIV glycoprotein 160 kD
gp41	HIV glycoprotein 41 kD
HA	hemagglutinin
HFP	HIV fusion peptide
HHCP	Hartman-Hahn cross polarization
HIV	human immunodeficiency virus
IBs	inclusion bodies
IFP	Influenza fusion peptide
IPTG	isopropyl β -D- thiogalactopyranoside
M	matrix protein
MAS	magic angle spinning
MLF	Methionine – Leucine – Phenylalanine
MPER	membrane proximal extracellular region

NA	neuraminidase
NHR	N-heptad repeat
NMR	nuclear magnetic resonance
NP	nucleoprotein
Ntr	N-terminal region
PAF	principle axis frame
PCR	polymerase chain reaction
PRE	paramagnetic relaxation enhancement
RbSu	receptor-binding subunit
REDOR	rotational-echo double resonance
RF	radiofrequency
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPECIFIC-CP	spectrally induced filtering in combination with cross polarization
ssNMR	solid-state nuclear magnetic resonance
TEMED	N, N, N', N'-tetramethylenediamine
TMD	transmembrane domain
TMS	tetramethyl silane
vRNP	viral ribonucleoprotein

Chapter 1 Introduction

1.1 NMR introduction¹

1.1.1 Description of NMR from classic mechanics

When a magnetic dipole moment $\vec{\mu}$ is positioned in a magnetic field \vec{B} , a torque $\vec{\tau}$ is exerted on the magnetic dipole moment.

$$\vec{\tau} = \vec{\mu} \times \vec{B} \quad (1.1)$$

The torque is strongest when $\vec{\mu}$ is perpendicular to \vec{B} and vanishes when $\vec{\mu}$ is aligned with \vec{B} . For a static magnetic moment in a magnetic field, the torque tends to line up the magnetic moment with the magnetic field. However, for nucleus whose magnetic moment generated by the spin of protons and neutrons, the torque exerts then creates a change in angular momentum (magnetic moment is proportional to angular momentum $\vec{\mu} = \gamma \vec{L}$), causing the magnetic moment rotating in a fixed angle with its axis along \vec{B} . Such movement is Larmor precession in NMR and the angular frequency is given by ω ($\text{rad} \cdot \text{s}^{-1}$) = $-\gamma B_0$. The minus sign here indicates the angular momentum is precessing by right hand rule when $\gamma > 0$. The half-angle of Larmor precession is $\cos \theta = \frac{m_l}{\sqrt{l(l+1)}}$ (l is spin quantum number and m_l is spin magnetic quantum number).

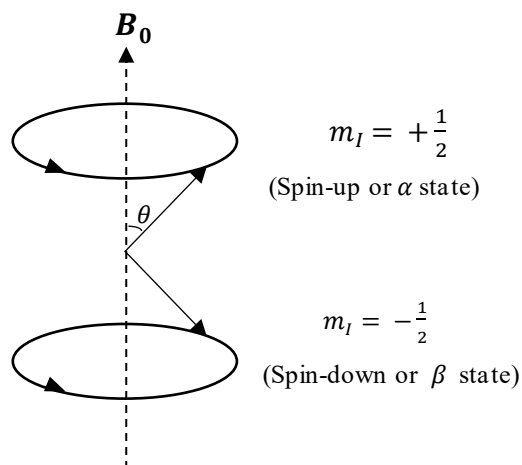


Figure 1.1. Larmor precession of spin-1/2 with Zeeman splitting.

The basic equation of torque $\vec{\tau}$ and angular momentum \vec{L} in physics is:

$$\vec{\tau} = \frac{d}{dt} \vec{L} \quad (1.2)$$

The motion of bulk nuclear magnetic moment \vec{M} magnetic field is:

$$\vec{M} = \sum_i \vec{\mu}_i \quad (1.3)$$

$$\frac{d}{dt} \vec{M} = \gamma \vec{M} \times \vec{B}$$

Combining the fact that \vec{B} is at Z-axis leading to Bloch equation without relaxation:

$$\frac{dM_x(t)}{dt} = M_y(t)\gamma B_0$$

$$\frac{dM_y(t)}{dt} = -M_x(t)\gamma B_0 \quad (1.4)$$

$$\frac{dM_z(t)}{dt} = 0$$

The solution of above equation is:

$$M_x(t) = M_x^0 \cos(\omega_0 t) - M_y^0 \sin(\omega_0 t)$$

$$M_y(t) = M_y^0 \cos(\omega_0 t) + M_x^0 \sin(\omega_0 t) \quad (1.5)$$

$$M_z(t) = M_z^0$$

where ω_0 is the Larmor frequency. The solution describes the magnetization without relaxation.

Consider the existence of transversal and longitudinal relaxation, the Bloch equation becomes:

$$M_x(t) = [M_x^0 \cos(\omega_0 t) - M_y^0 \sin(\omega_0 t)] \cdot e^{-\frac{t}{T_2}}$$

$$M_y(t) = [M_y^0 \cos(\omega_0 t) + M_x^0 \sin(\omega_0 t)] \cdot e^{-\frac{t}{T_2}} \quad (1.6)$$

$$M_z(t) = M_z^0 + \left(1 - e^{-\frac{t}{T_1}}\right) (M_z^\infty - M_z^0)$$

1.1.1.1 Detection in rotating frame of reference

What is detected in NMR is the precession of magnetization at by-plane in modern pulsed NMR experiment. The detector of NMR is a small coil of wire round the sample with its axis aligned in the my-plane. When magnetization is precessing, the bulk magnetization vector cuts the coil and induce a current. The induced current can be amplified and then recorded as NMR signal, which is called free induction decay (FID).

The applied on-resonance pulse is an oscillating magnetic field, $2B_1 \cos(\omega_{R.F.,on-res} t)$, along x direction. The magnetic field can be treated as the summation of two counter-rotating fields with angular frequency $\omega_{R.F.,on-res}$. One component is rotating at the direction of Larmor precession and the other is at the opposite direction, which can be neglected. The filed acting on the samples in

the laboratory frame can be expressed as:

$$\vec{B} = B_1 \cos(\omega_{R.F.,on-res} t) \vec{x} - B_1 \sin(\omega_{R.F.,on-res} t) \vec{y} + B_0 \vec{z} \quad (1.7)$$

For an angular momentum vector precessing at Larmor frequency, if the nucleus is observed in a frame rotating at the same frequency, the nucleus looks like static, and the precessing caused by the external magnetic field is disappeared in this reference frame. In rotating frame, if the rotating frame frequency chosen to be close to Larmor frequency, the z component of effective field is relatively small to the \vec{B}_1 field generated by radiofrequency (R.F.) pulse. The Rabi frequency of R.F. pulse, whose order of magnitude is kHz, can cause spin transitions. Ideally the rotating frame should exactly be equal to Larmor frequency. However, the real sample is always more complicated than a single spin system and there always a distribution of Larmor frequency. This could be brought by a different chemical environment from bonding electrons or a different molecular orientation with respect to the external magnetic field. Thus, the different between Larmor frequency and rotating frequency (the rotating frequency is usually the same as the transmitter frequency) is the offset Ω :

$$\Omega = \omega_0 - \omega_{rot.frame} \quad (1.8)$$

The effective field that nucleus is experiencing under radiofrequency irradiation is:

$$\vec{B}_{rot.frame} = -\frac{\vec{\omega}_{transmitter}}{\gamma}$$

$$\vec{B}_{eff} = \vec{B}_1 + (\vec{B}_0 - \vec{B}_{rot.frame}) \quad (1.9)$$

where \vec{B}_1 field is generated by a radiofrequency pulse. The \vec{B}_1 field for a 90° pulse can be described as:

$$\vec{B}_1 = B_1 \cos(\omega_{R.F.} t) \vec{x} \quad (1.10)$$

where $\omega_{R.F.} = 2\pi\nu$ and ν is the frequency of the 90° pulse.

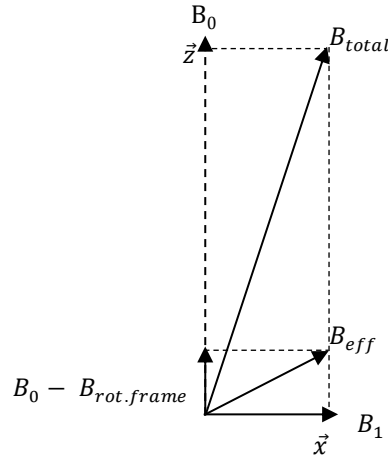


Figure 1.2. Magnetic field in laboratory frame vs. rotating frame for $\omega_{\text{rot,frame}} < \omega_0$. The magnitude of \vec{B}_1 is exaggerated for clarity.

It is clear to be seen in Figure 1.2 that a very weak magnetic field \vec{B}_1 compared with \vec{B}_0 will not alter the total magnetic field deviated from z-axis a lot in laboratory frame. But in rotating frame, the z component of \vec{B}_{eff} field is greatly attenuated so that it is possible that B_{eff} is close to the direction of radiofrequency pulse. Once a sample has been inserted into the magnetic field, the equilibrium magnetization is built up at z-axis. In order to detect signal at xy-plane, a resonance pulse which frequency is equal to transmitter frequency will be applied at x-axis, and the pulse can rotate the magnetization from z-axis to desired direction to be detectable.

1.1.1.2 Relaxation

Whenever there is a deviation to the thermal equilibrium state of the spin system placed in \vec{B}_0 field, the magnetization \mathbf{M} will decay to its thermal equilibrium \mathbf{M}_0 along z axis. \mathbf{M}_0 is defined as:

$$M_0 = \frac{1}{4} \frac{N(\gamma\hbar)^2 B_0}{k_B T} \quad (1.11)$$

where N is the total number of nuclei in the given sample.

The approach that the magnetization \mathbf{M} decays to its thermal equilibrium state is known as relaxation which can be categorized by two kinds, longitudinal and transversal relaxation.

The Bloch equations with relaxation in laboratory frame are:

$$\begin{aligned}
\frac{dM_x}{dt} &= \gamma [M_y B_0 + M_z B_1 \sin(\omega t)] - \frac{M_x}{T_2} \\
\frac{dM_y}{dt} &= -\gamma [M_x B_0 - M_z B_1 \cos(\omega t)] - \frac{M_y}{T_2} \\
\frac{dM_z}{dt} &= -\gamma [M_x B_1 \sin(\omega t) + M_y B_1 \cos(\omega t)] - \frac{M_z - M_z^0}{T_1}
\end{aligned} \tag{1.12}$$

The transverse component magnetization \vec{M} in laboratory frame observed in a rotating frame (represented with primes) with a rotating frequency ω about z-axis is given by:

$$\begin{aligned}
M'_x &= M_x \cos(\omega t) - M_y \sin(\omega t) \\
M'_y &= M_x \sin(\omega t) + M_y \cos(\omega t)
\end{aligned} \tag{1.13}$$

The Bloch equation in the rotating frame is:

$$\begin{aligned}
\frac{dM'_z}{dt} &= -\gamma B_1 M'_x - \frac{M_z - M_z^0}{T_1} \\
\frac{dM'_x}{dt} &= (\omega_0 - \omega) M'_y - \frac{M'_x}{T_2} \\
\frac{dM'_y}{dt} &= -(\omega_0 - \omega) M'_x + \gamma B_1 M'_z - \frac{M'_y}{T_2}
\end{aligned} \tag{1.14}$$

and its solutions are:

$$\begin{aligned}
M'_z &= \frac{M_z^0 [1 + T_2^2 (\omega_0 - \omega)^2]}{T_2^2 (\omega_0 - \omega)^2 + 1 + T_1 T_2 \gamma^2 B_1^2} \\
M'_x &= \frac{M_z^0 \gamma B_1 T_2^2 (\omega_0 - \omega)^2}{T_2^2 (\omega_0 - \omega)^2 + 1 + T_1 T_2 \gamma^2 B_1^2} \\
M'_y &= \frac{M_z^0 \gamma B_1 T_2}{T_2^2 (\omega_0 - \omega)^2 + 1 + T_1 T_2 \gamma^2 B_1^2}
\end{aligned} \tag{1.15}$$

Unlike rotational and vibrational relaxation times are more often of the order of 10^{-9} s and 10^{-8} s respectively, nuclear relaxation times are usually of order from milliseconds to seconds. The relaxation is the process that magnetization return to thermal equilibrium state. The magnetic field at the appropriate frequency will influence the relaxation process. For solution NMR, the fluctuating magnetic field of magnetic dipole moment of a nuclear spin caused by Brownian motion has a contribution to relaxation. Brownian motion is driven by thermal energy, causing particles to undergo random motion and collisions with surrounding molecules. This thermal

motion is responsible for the Brownian motion of particles with magnetic moments. As the particles move randomly due to Brownian motion, their magnetic moments experience fluctuations in orientation. The fluctuating field can be decomposed into the component which is parallel and perpendicular to the B_0 field. The perpendicular component oscillating at Larmor frequency contributes to both longitudinal and transverse relaxation.

The relaxation rate for spin- $\frac{1}{2}$ can be written as:

$$\begin{aligned} T_1^{-1} &= \gamma^2 [B_{xLocal}^0]^2 J(\omega_0) \\ T_2^{-1} &= (2T_1)^{-1} + \frac{1}{2} \gamma^2 [B_{zLocal}^0]^2 J(0) \end{aligned} \quad (1.16)$$

where $J(\omega)$ is spectral density and defined as:

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \quad (1.17)$$

τ_c is the correlation time (defined as the time that the whole-molecule take to rotate 1 radian) and can be calculated as $(\tau_r^{-1} + \tau_s^{-1})^{-1}$ where τ_r and τ_s are the rotational correlation time of the macromolecule and the effective electron relaxation time respectively. The spectral density describes how much fluctuation occurs at different frequencies. Longitudinal relaxation involves energy exchange at Larmor frequency. For transverse relaxation, there is another contribution which is time-independent and does not lead to oscillations between different energy levels. The contribution is $J(0)$ referring to the zero-frequency component of the spectral density. B_{xLocal}^0 and B_{zLocal}^0 are the root-mean-square average amplitude of the isotropic random magnetic field fluctuation due to various factors, such as molecular motion, local magnetic susceptibility variations due to rotational and vibrational motion of molecules, and other environmental influences including changes of temperature and pressure acting on it in a time-dependent fashion. Usually the x- and y-component of magnetization in the rotating frame is detected as NMR signal of absorption vs. frequency. The predicted lineshape of absorption mode would be Lorentzian as relaxation processes are often modeled as exponential decays in time and the exponential decay of the NMR signal, influenced by T_1 and T_2 relaxation, leads to a Lorentzian lineshape in the frequency domain. The linewidth at half-height is $(\pi T_2)^{-1}$ in frequency units in the scenario that the sample has a relative long transverse relaxation time compared to the duration of signal without any other broadening mechanisms, such as inhomogeneities in the magnetic field or chemical shift

dispersion. For ideal experiments, the linewidth is inversely proportional to T_2 . In practice, NMR peaks are often not Lorentzian and sometimes can be unsymmetrical. The reason for this is that T_2 is usually so small that the observed linewidth has a large contribution from the inhomogeneity caused by instrument. Even though the inhomogeneity can be reduced by shimming, there is always a few hertz broadening from instrument.²

1.1.2 Description of NMR from quantum mechanics

1.1.2.1 NMR interactions

Hamiltonian (\hat{H}) is an operator to describe the total energy of the system, including kinetic and potential energy in quantum mechanics.

Observables are Hermitian operators in Hilbert space and the outcome of the measurement is equal to the expectation value of the corresponding operator Ω acting on the wavefunction describing the system.

$$\langle \Omega \rangle = \frac{\int \psi^* \Omega \psi \, d\tau}{\int \psi^* \psi \, d\tau} = \frac{\langle \psi | \Omega | \psi \rangle}{\langle \psi | \psi \rangle} = \langle \psi | \Omega | \psi \rangle \quad (1.18)$$

From Equation 1.17, the expectation values of the system described by normalized wavefunction ψ is simplified to $\langle \psi | \Omega | \psi \rangle$.

If ψ is the eigenfunction of Ω with eigenvalue A , then

$$\langle \Omega \rangle = \langle \psi | \Omega | \psi \rangle = A \quad (1.19)$$

If the system is an eigenstate of Hamiltonian operator but not eigenstate of Ω , the wavefunctions can be expressed as a linear combination of eigenfunctions of Ω :

$$\psi = \sum_n c_n \psi_n \text{ where } \Omega | \psi_n \rangle = \alpha_n \psi_n \quad (1.20)$$

Then, the expectation value becomes:

$$\begin{aligned}
\langle \Omega \rangle &= \int \left(\sum_m c_m \psi_m \right)^* \Omega \left(\sum_n c_n \psi_n \right) d\tau \\
&= \sum_{m,n} c_m^* c_n \int \psi_m^* \Omega \psi_n d\tau \\
&= \sum_{m,n} c_m^* c_n \alpha_n \int \psi_m^* \psi_n d\tau \\
&= \sum_n c_n^* c_n \alpha_n \\
&= \sum_n |c_n|^2 \alpha_n
\end{aligned} \tag{1.21}$$

The wavefunctions are orthonormal set, so the final result is a weighted sum of eigenvalues of Ω . Whenever there is a pulse applied to an NMR sample, it is a perturbation to wavefunctions of sample and the evolution of wavefunction is predicted by time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\Psi\rangle = \hat{H} |\Psi\rangle \tag{1.22}$$

The interaction between any NMR active nuclei with nuclear spin quantum number $I \neq 0$ and magnetic field can be categorized into external and local interactions. The interaction of an NMR sample is expressed as:

$$\begin{aligned}
\hat{H}_{\text{tot}} &= \hat{H}_{\text{external}} + \hat{H}_{\text{local}} \\
&= \hat{H}_z + \hat{H}_{\text{rf}} + \hat{H}_{\text{local}} \\
&= \hat{H}_z + \hat{H}_{\text{rf}} + \hat{H}_{\text{cs}} + \hat{H}_D + \hat{H}_J + \hat{H}_Q
\end{aligned} \tag{1.23}$$

where \hat{H}_z is the Hamiltonian for the Zeeman interaction;

\hat{H}_{rf} is the Hamiltonian for the radiofrequency pulses;

\hat{H}_{cs} is the Hamiltonian for the chemical shift;

\hat{H}_D is the Hamiltonian for the dipolar interaction;

\hat{H}_J is the Hamiltonian for the J-coupling;

\hat{H}_Q is the Hamiltonian for the quadrupolar interaction.

$\hat{H}_{\text{cs}}, \hat{H}_D, \hat{H}_J$ and \hat{H}_Q is considered as local field interactions.

Zeeman interaction

Spin is intrinsic angular momentum associated with elementary particles. It is a quantum-mechanical phenomenon which has no analogy in classical physics. Spin quantum number is a

quantum number describing spin angular momentum of particles. The only possible values of magnetic spin quantum number of electrons, protons and neutrons are $\pm \frac{1}{2}$. The nucleus which has non-zero net spin values will interact with external magnetic field and be nuclear magnetic resonance (NMR) active.

If a nucleus with a spin quantum number I is placed in a magnetic field \mathbf{B}_0 , the nuclear spin energy level will split into $(2I + 1)$ energy states, which is referred as Zeeman splitting. The Zeeman interaction can be described as:

$$\hat{H}_Z = -\hat{\boldsymbol{\mu}} \cdot \mathbf{B}_0 \quad (1.24)$$

where $\hat{\boldsymbol{\mu}}$ is the magnetic dipole moment of the nucleus and is given by

$$\hat{\boldsymbol{\mu}} = \gamma\hbar\hat{\mathbf{I}} = \gamma\hbar[\hat{\mathbf{i}}\hat{I}_x + \hat{\mathbf{j}}\hat{I}_y + \hat{\mathbf{k}}\hat{I}_z] \quad (1.25)$$

where γ is gyromagnetic ratio and defined as,

$$\vec{\boldsymbol{\mu}} = \gamma \vec{\mathbf{L}} \quad (1.26)$$

\hbar is the reduced Planck constant, $\hat{\mathbf{I}}$ is the nuclear spin operator. \hat{I}_x , \hat{I}_y , and \hat{I}_z are nuclear spin operators for x, y, and z components. They are related to $\hat{\mathbf{I}}$ by

$$\hat{\mathbf{I}}^2 = \hat{I}_x^2 + \hat{I}_y^2 + \hat{I}_z^2 \quad (1.27)$$

\mathbf{i} , \mathbf{j} , and \mathbf{k} are the unit vectors in x, y, and z axes respectively.

\mathbf{B}_0 is applied magnetic field, usually aligned at Z-axis. Substituting Equations 1.8 and 1.10 to 1.7 gives

$$\hat{H}_Z = -\gamma\hbar\hat{I}_z B_0 \quad (1.28)$$

Since \hat{H}_Z is proportional to \hat{I}_z , they commute with each other and then share a complete set of eigenstates. Eigenfunction of \hat{I}_z is written as $|I, m\rangle$ where m is spin magnetic quantum number.

The eigenfunction of angular momentum $\hat{\mathbf{I}}$ represented with Dirac expressions are:

$$\begin{aligned} \hat{I}_z |I, m\rangle &= m |I, m\rangle \\ \hat{\mathbf{I}}^2 |I, m\rangle &= I(I+1) |I, m\rangle \end{aligned} \quad (1.29)$$

Substituting Equation 1.12 to 1.11:

$$\hat{H}_Z |I, m\rangle = E_{I,m} |I, m\rangle = -\gamma\hbar\hat{I}_z B_0 |I, m\rangle = -\gamma\hbar m B_0 |I, m\rangle \quad (1.30)$$

Therefore, Zeeman energy of the eigenstates are:

$$E_{I,m} = -\gamma\hbar m B_0 \quad (1.31)$$

Spin angular momentum is quantized and its projection at i -th axis (either x, y, or z) is:

$$I_i = \hbar m_i, \quad (1.32)$$

$$m_i \in \{-I, -I+1, \dots, I-1, I\}$$

The $(2I+1)$ possible values of I_z causes $(2I+1)$ multiplicity of the nucleus energy in external magnetic field. The splitting is referred as Zeeman splitting, which is always the strongest interaction between nucleus and applied field.

For a general and simple case, a spin-1/2 nucleus,

$$E_{\frac{1}{2}, +\frac{1}{2}} = -\frac{1}{2} \gamma \hbar B_0$$

$$E_{\frac{1}{2}, -\frac{1}{2}} = \frac{1}{2} \gamma \hbar B_0 \quad (1.33)$$

$$\Delta E_{\frac{1}{2}, -\frac{1}{2}} = \gamma \hbar B_0$$

The two sub-energy state is commonly referred as spin-up & spin-down or α & β state. The energy difference is transition energy of the two states.

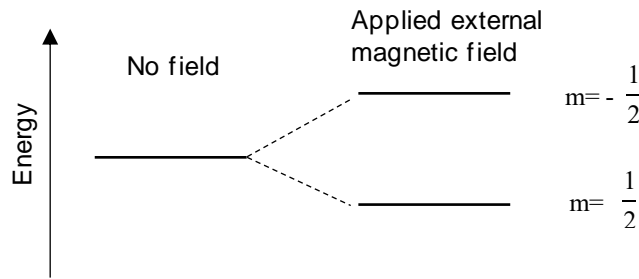


Figure 1.3. Energy splitting caused by Zeeman effect for spin-1/2 nucleus.

The local interactions are relative small compared to Zeeman energy therefore can be treated in first-order perturbation theory. Briefly, the energy shifted by local interaction is given by the diagonal elements of the local interaction Hamiltonian expressed in a basis of Zeeman eigenfunctions, $E_m^{(1)} = \langle m | \hat{H}_1 | m \rangle$, and the off-diagonal elements are negligible. (operators that are diagonal in the same basis have this basis as their common eigenfunctions and commute. So the local interaction commute with Zeeman should be retained after truncation.)

Effect of radiofrequency (R.F.) pulses

Whenever there is a pulse applied to an NMR sample, it is a perturbation to wavefunctions of sample and the evolution of wavefunction is predicted by time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\Psi\rangle = \widehat{H}|\Psi\rangle \quad (1.34)$$

A R.F. pulse introduce an oscillating magnetic field into the spin system. When there is a R.F. pulse at x -axis along with the \vec{B}_0 field, the total magnetic field and its associated Hamiltonian becomes:

$$\mathbf{B}_{\text{total}}(t) = B_1 \cos(\omega_{R.F.} t) \vec{x} + B_0 \vec{z} \quad (1.35)$$

$$\widehat{H} = -\gamma\hbar [\hat{I}_z B_0 + \hat{I}_x B_1 \cos(\omega_{R.F.} t)] \quad (1.36)$$

The oscillating field can be rewritten as sum of the components:

$$B_1^{\text{on-res}} = \frac{1}{2} B_1 [\cos(\omega_{R.F.} t) + \sin(\omega_{R.F.} t)] \quad (1.37)$$

$$B_1^{\text{off-res}} = \frac{1}{2} B_1 [\cos(\omega_{R.F.} t) - \sin(\omega_{R.F.} t)]$$

Only the on-resonance component has a great effect on spins while off-resonance component can be neglected. The on-resonance field is oscillating at the same rate as the sample spin doing Larmor precession if $\omega_{R.F.} = \omega_0$. That means the B_1 field looks static in the rotating frame for on-resonance pulse. If the pulse is turned on for the pulse duration t_p , the magnetization will rotate by angle of:

$$\theta = \omega_{R.F.} \cdot t_p \quad (1.38)$$

Thus, the effect of a R.F. pulse is to rotate the magnetization. A more accurate description about the effect of R.F. pulse is given by:

$$e^{-i\phi\widehat{A}} \widehat{B} e^{i\phi\widehat{A}} = \widehat{B} \cos \phi + \widehat{C} \sin \phi \quad (1.39)$$

if the two operators have the commutation relation (the commutator is calculated as $[\widehat{A}, \widehat{B}] = \widehat{A}\widehat{B} - \widehat{B}\widehat{A}$),

$$[\widehat{A}, \widehat{B}] = i\widehat{C} \quad (1.40)$$

As for spin angular momentum, the relation is extended to:

$$[\hat{I}_m, \hat{I}_j] = i\epsilon_{mjk} \hat{I}_k$$

$$\epsilon_{mjk} = \begin{cases} 0, & \text{if any of } m, j, k \text{ are identical} \\ 1, & \text{if } m, j, k \text{ are in cyclic order} \\ -1, & \text{if } m, j, k \text{ are in anticyclic order} \end{cases} \quad (1.41)$$

Isotropic and anisotropic chemical shift interactions

Even though Zeeman interaction is usually dominated, it does not reveal any structural information. When placed in an external magnetic field, the electrons moving around the nucleus create a secondary magnetic field opposing to the original external magnetic field, so the nucleus

experiences a weaker magnetic field. This shielding is referred as shielding field or chemical shift field which is determined by chemical shift tensor and its associated Hamiltonian is:

$$\hat{H}_{cs} = \gamma \hbar \hat{\mathbf{I}} \boldsymbol{\sigma} \mathbf{B}_0 = \gamma \hbar (\hat{I}_x \sigma_{xz}^{LF} + \hat{I}_y \sigma_{yz}^{LF} + \hat{I}_z \sigma_{zz}^{LF}) \mathbf{B}_0 \quad (1.42)$$

where $\boldsymbol{\sigma}$ is the shielding tensor observed at laboratory-frame. $\boldsymbol{\sigma}$ is second rank tensor, represented by:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \quad (1.43)$$

Generally, the electron density distribution around the nucleus is not spherically symmetric. The shielding tensor can be decomposed into a symmetric $\boldsymbol{\sigma}^s$ and antisymmetric $\boldsymbol{\sigma}^{as}$ component:

$$\boldsymbol{\sigma}^s = \begin{pmatrix} \sigma_{xx} & \frac{1}{2}(\sigma_{xy} + \sigma_{yx}) & \frac{1}{2}(\sigma_{xz} + \sigma_{zx}) \\ \frac{1}{2}(\sigma_{xy} + \sigma_{yx}) & \sigma_{yy} & \frac{1}{2}(\sigma_{yz} + \sigma_{zy}) \\ \frac{1}{2}(\sigma_{xz} + \sigma_{zx}) & \frac{1}{2}(\sigma_{yz} + \sigma_{zy}) & \sigma_{zz} \end{pmatrix} \quad (1.44)$$

$$\boldsymbol{\sigma}^{as} = \begin{pmatrix} 0 & \frac{1}{2}(\sigma_{xy} - \sigma_{yx}) & \frac{1}{2}(\sigma_{xz} - \sigma_{zx}) \\ \frac{1}{2}(\sigma_{yx} - \sigma_{xy}) & 0 & \frac{1}{2}(\sigma_{yz} - \sigma_{zy}) \\ \frac{1}{2}(\sigma_{zx} - \sigma_{xz}) & \frac{1}{2}(\sigma_{zy} - \sigma_{yz}) & 0 \end{pmatrix}$$

The $\boldsymbol{\sigma}^{as}$ has limited effect to chemical shift when Zeeman truncation applies. This is because the anisotropic contributions involve higher-order terms that are neglected in Zeeman truncation.

It is possible to find a new axis frame that $\boldsymbol{\sigma}^s$ is diagonal. The frame is the principal axis frame (PAF) and the diagonals are principle values. The shielding tensor associated with the PAF has the three principal values σ_{xx} , σ_{yy} , and σ_{zz} . The values depend on the molecular orientation of the molecule relative to the \vec{B}_0 field.(shown in Figure1.4) Conventionally,

$$\sigma_{iso} = \frac{1}{3} (\sigma_{xx}^{PAF} + \sigma_{yy}^{PAF} + \sigma_{zz}^{PAF})$$

$$\delta = \sigma_{zz}^{PAF} - \sigma_{iso} \quad (1.45)$$

$$\eta = \frac{\sigma_{xx}^{PAF} - \sigma_{yy}^{PAF}}{\sigma_{zz}^{PAF}}$$

σ_{iso} is isotropic chemical shift and is δ chemical shift anisotropy. In solid state NMR, chemical shift anisotropy is very important and reveals some structural information. Chemical shift

expressed in PAF is:

$$\hat{H}_{cs} = \hbar\gamma B_0 \left\{ \sigma_{iso} + \frac{1}{2} \delta_{cs} [3\cos^2\theta - 1 - \eta_{cs} \sin^2\theta \cos(2\phi)] \right\} \hat{I}_Z \quad (1.46)$$

where θ, ϕ are the polar angles of the field B_0 in the PAF.

The corresponding frequency is:

$$\omega_{cs}(\theta, \phi) = -\omega_0 \hbar \left\{ \sigma_{iso} + \frac{1}{2} \delta_{cs} [3\cos^2\theta - 1 - \eta_{cs} \sin^2\theta \cos(2\phi)] \right\} \quad (1.47)$$

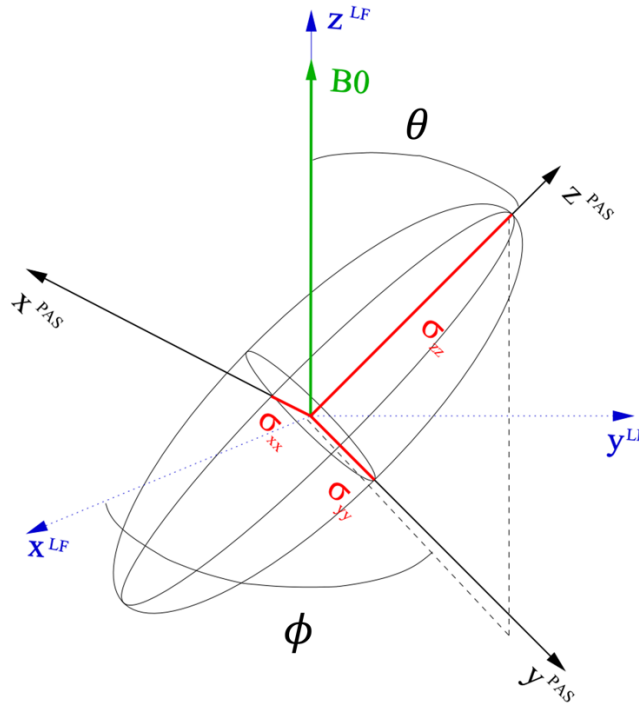


Figure 1.4. The principal axes frame and its principal values.³

For powder samples lacking motion averaging, all molecular orientations are present and make its associate contribution to chemical shift. The spectrum is a powder pattern with lines from different molecular orientations. In a powder pattern, the relative intensity is proportional to the number of the molecules present at the orientation corresponding to a particular frequency.

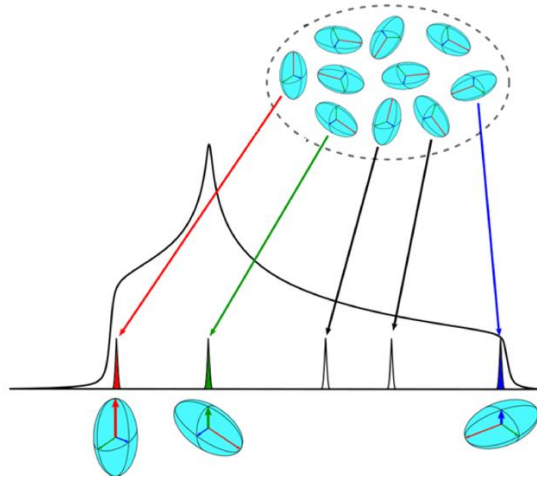


Figure 1.5. Illustration of a powder pattern that each different molecular orientation showing signals at its corresponding frequency.⁴

Dipolar interaction

Another important interaction is dipolar interaction, which is the direct interaction between two dipoles. Different than J-coupling which is an indirect interaction mediated by bonding electrons, dipolar coupling is through space. In solution, dipolar coupling is averaged to zero by molecular tumbling. On the contrary, it is a major reason for line broadening of solid sample.

The complete expression of dipolar Hamiltonian is:

$$\hat{H}_D = \gamma_1 \gamma_2 \frac{\mu_0}{4\pi} \hbar \left\{ \frac{\hat{\mathbf{I}}_1 \cdot \hat{\mathbf{I}}_2}{r^3} - 3 \frac{(\hat{\mathbf{I}}_1 \cdot \mathbf{r})(\hat{\mathbf{I}}_2 \cdot \mathbf{r})}{r^5} \right\} \quad (1.48)$$

With \mathbf{r} is distance vector of the two involved spins and the scalar products expressed in polar coordinates (shown in Figure 1.6), the Hamiltonian comes to:

$$\hat{H}_D = \frac{\mu_0}{4\pi} \hbar \frac{\gamma_1 \gamma_2}{r^3} (A+B+C+D+E+F) \quad (1.49)$$

$$A = -\hat{I}_{1Z} \hat{I}_{2Z} (3\cos^2\theta - 1)$$

$$B = \frac{1}{4} [\hat{I}_{1+} \hat{I}_{2-} + \hat{I}_{1-} \hat{I}_{2+}] (3\cos^2\theta - 1)$$

$$C = -\frac{3}{2} [\hat{I}_{1Z} \hat{I}_{2+} + \hat{I}_{1+} \hat{I}_{2Z}] (\sin\theta \cos\theta e^{-i\phi})$$

$$D = -\frac{3}{2} [\hat{I}_{1Z} \hat{I}_{2-} + \hat{I}_{1-} \hat{I}_{2Z}] (\sin\theta \cos\theta e^{i\phi})$$

$$E = -\frac{3}{4} \hat{I}_{1+} \hat{I}_{2+} \sin^2\theta e^{-2i\phi}$$

$$F = -\frac{3}{4} \hat{I}_1 \cdot \hat{I}_2 \sin^2 \theta e^{2i\phi}$$

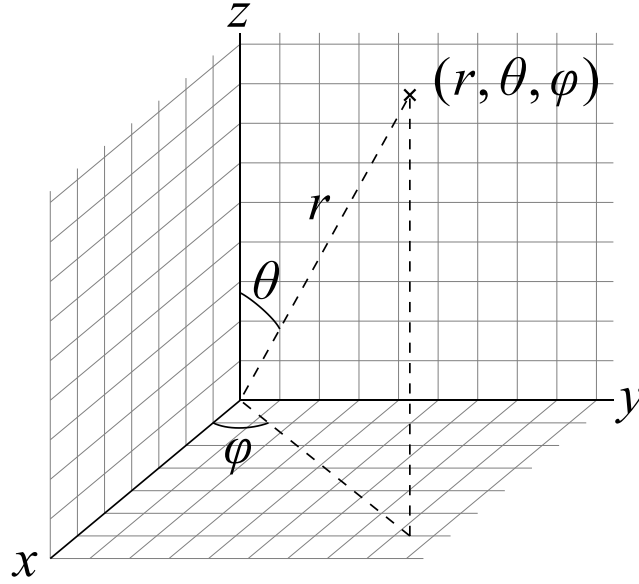


Figure 1.6. The physics convention of spherical polar coordinates.

The truncated heteronuclear dipolar coupling is:

$$\hat{H}_D^{S,I} = -\frac{\mu_0}{4\pi} \hbar \sum_j \sum_k \frac{\gamma_I \gamma_S}{r_{jk}^3} \left\{ \frac{1}{2} (3 \cos^2(\theta_{jk}) - 1) \right\} \cdot 2\hat{I}_z \hat{S}_z \quad (1.50)$$

The truncated homonuclear dipolar coupling is:

$$\begin{aligned} \hat{H}_D^{I,I} &= \sum_j \sum_{k < j} \frac{1}{2} (\mathbf{D}_{LF}^{jk})_{zz} (3\hat{I}_z^j \hat{I}_z^k - \hat{\mathbf{I}}^j \cdot \hat{\mathbf{I}}^k) \\ &= -\frac{\mu_0}{4\pi} \hbar \sum_j \sum_{k < j} \frac{\gamma^2}{r_{jk}^3} \left\{ \frac{1}{2} (3 \cos^2(\theta_{jk}) - 1) \right\} (3\hat{I}_z^j \hat{I}_z^k - \hat{\mathbf{I}}^j \cdot \hat{\mathbf{I}}^k) \end{aligned} \quad (1.51)$$

where θ_{jk} is the angle between B_0 and the vector connecting nuclei j and k . Dipolar coupling constant is given by:

$$\omega_d = \frac{\mu_0}{4\pi} \hbar \frac{\gamma_1 \gamma_2}{r_{1,2}^3} = 2\pi \cdot 122 \text{ kHz} \frac{\gamma_1}{\gamma^{1H}} \frac{\gamma_2}{\gamma^{1H}} \frac{1}{(r_{1,2}/1 \text{ \AA})^3} \quad (1.52)$$

For example, for the ^{13}C - ^1H spin pair, the dipolar coupling constant is 23kHz for a distance of 1.1 Å.

Quadrupolar interaction

Quadrupole moment Q presents for spin $> \frac{1}{2}$ nuclei. The interaction between electric quadrupole

and electric field gradients (the electric field gradient tensor is denoted as \mathbf{V}) is:

$$\begin{aligned}
\hat{H}_Q &= \frac{eQ}{2I(2I-1)} \hat{\mathbf{I}} \mathbf{V} \hat{\mathbf{I}} \\
&= \frac{eQ}{2I(2I-1)} V_{zz}^{LF} \frac{1}{2} (3\hat{I}_z \hat{I}_z - \hat{\mathbf{I}} \cdot \hat{\mathbf{I}}) \\
&= \frac{eQ}{2I(2I-1)} V_{zz}^{PAF} \frac{1}{2} [3\cos^2\theta - 1 - \eta_Q \sin^2\theta \cos(2\phi)] \frac{1}{2} (3\hat{I}_z \hat{I}_z - \hat{\mathbf{I}} \cdot \hat{\mathbf{I}}) \\
&= \frac{eQeq}{2I(2I-1)} \frac{1}{2} [3\cos^2\theta - 1 - \eta_Q \sin^2\theta \cos(2\phi)] \frac{1}{2} (3\hat{I}_z^2 - \hat{\mathbf{I}} \cdot \hat{\mathbf{I}})
\end{aligned} \tag{1.53}$$

$V_{zz}^{PAF} = e \cdot q$ is the customary expression for the bond-structure-dependent principal value of the field-gradient tensor for spin-1 nuclei.

η_Q is the asymmetry parameter of the electric field gradient tensor.

All the interaction described above is the minor energy adjustment to Larmor frequency.

1.1.2.2 Density operators and its application

Theoretically, Hamiltonian describes the system's total energy, and it can be used to predict how the system evolve under certain interactions. So once the wavefunction has been determined, the observable expectation values should be predicable. However, the real system contains a large number of nuclear spins, and it is not possible to find the wavefunction of each spin. Spin system is not completely polarized so that it cannot be described by simple solutions of the Schrödinger equation. Complete polarization would imply that all nuclear spins are perfectly aligned with the external magnetic field. In many realistic scenarios, the nuclear spin system is not fully aligned or polarized in a particular direction. As a result, the behavior of the spin system cannot be adequately described by simple or idealized solutions of the Schrödinger equation. Density operator, on the other hand, combining quantum mechanics and statistical mechanics, is appropriate to represent spin system with any polarization.

Density operator $\hat{\rho}$ is defined as:

$$\hat{\rho} = \overline{|\psi\rangle\langle\psi|} \tag{1.54}$$

The overbar indicates taking an ensemble average (taking average of the whole system). For example, the superposition state of spin- $\frac{1}{2}$, the matrix representation of $\hat{\rho}$ will be:

$$|\psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle$$

$$\rho = \begin{pmatrix} \langle\alpha|\hat{\rho}|\alpha\rangle & \langle\alpha|\hat{\rho}|\beta\rangle \\ \langle\beta|\hat{\rho}|\alpha\rangle & \langle\beta|\hat{\rho}|\beta\rangle \end{pmatrix} \equiv \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} = \begin{pmatrix} \overline{c_\alpha c_\alpha^*} & \overline{c_\alpha c_\beta^*} \\ c_\beta c_\alpha^* & \overline{c_\beta c_\beta^*} \end{pmatrix} \quad (1.55)$$

The equilibrium density operator obeys Boltzmann distribution:

$$\hat{\rho}_{\text{eq}} = \frac{e^{-\hat{H}/kT}}{\text{Tr}(e^{-\hat{H}/kT})} = \frac{1}{Z} e^{-\hat{H}/kT} \quad (1.56)$$

For Zeeman interaction with temperature above 1K, $|\hbar\gamma B_0/kT| \ll 1$, approximation for the expansion of exponential operator is valid then density operator become:

$$\hat{\rho}_{\text{eq}} \approx \frac{1}{Z} \left(\hat{1} + \frac{\hbar\omega_0}{kT} \hat{I}_z \right) \quad (1.57)$$

Any existing interaction will have impact to the density operator describing the spin system. The density operator reacts according to:

$$\frac{d\hat{\rho}}{dt} = -i[\hat{H}, \hat{\rho}] \quad ([\hat{H}, \hat{\rho}] \text{ is commutator}) \quad (1.58)$$

which represents a generalization of the Schrödinger equation. We are more familiar with the form:

$$\frac{d}{dt} \psi = -i\hat{H}\psi \quad (1.59)$$

It is easy to solve when the spin system is in an eigenstate of the Hamiltonian. There is ensemble of nuclear spins in an NMR sample and the density operator is used to represent to quantum state of the system.

If the Hamiltonian is time-independent, solution is:

$$\hat{\rho}(t) = e^{-i\hat{H}t} \hat{\rho}(0) e^{i\hat{H}t} \quad (1.60)$$

The solution for the time-dependent Hamiltonian is given by:

$$\hat{\rho}(t) = \hat{T} e^{-i\int_0^t \hat{H}(t) dt} \hat{\rho}(0) e^{i\int_0^t \hat{H}(t) dt} \quad (1.61)$$

where \hat{T} is the Dyson time-ordering operator. This operator is required when Hamiltonian does not commute with itself at different times t .

For modern NMR experiments, pulse sequences are widely used. In this case, Hamiltonian is piece-wise constant, which means Hamiltonian is constant for a period of time. The density operator at the end of the pulse sequence is:

$$\hat{\rho}(t_1+t_2+\dots+t_n) = e^{-i\hat{H}_n t_n} \dots e^{-i\hat{H}_2 t_2} e^{-i\hat{H}_1 t_1} \hat{\rho}(0) e^{i\hat{H}_1 t_1} e^{i\hat{H}_2 t_2} \dots e^{i\hat{H}_n t_n} \quad (1.62)$$

The physical observables or the thermodynamic averages of the expectation value of operator is by taking the following trace:

$$\langle \hat{A} \rangle = \text{Tr}(\rho \mathbf{A}) := \sum_i (\rho \mathbf{A})_{ii} = \sum_i \sum_j \rho_{ij} A_{ji} \quad (1.63)$$

Once the density operator of the spin system is available, NMR signal $S(t)$ can be calculated as:

$$S(t) \sim \text{Tr}\{\rho(t) \cdot (\mathbf{I}_x + i\mathbf{I}_y)\} \quad (1.64)$$

where \mathbf{I}_x and \mathbf{I}_y are corresponding to real and imaginary signal respectively.

NMR signals are typically acquired in the time domain and then transformed into the frequency domain using Fourier transformation. The NMR signal is the magnetization vector rotating in the rotating frame and is conveniently represented by the complex number. It is detected and stored as real and imaginary component and can be manipulated mathematically using the complex plane. After Fourier transformation, the complex time-domain signal is converted into a complex frequency-domain spectrum. The real part corresponds to the absorption peaks associated with chemical shifts, while the imaginary part corresponds to the dispersion peaks. Detecting x and y components of the NMR signal is known as quadrature detection. The y component is useful to distinguish between frequencies that are offset from the carrier frequency with different signs. The presence of imaginary signal is crucial for phase adjustment to maximize the signal and minimize baseline distortions.

The effect of R.F. pulses examined by density operator

In the rotating frame, the Hamiltonian of R.F. pulse at x-axis with associated magnetic field B_1 is

$$\hat{H}_{\text{R.F.}} = -\hbar\omega_1 \hat{\mathbf{I}}_x \quad (1.65)$$

The corresponding density operator is:

$$\begin{aligned} \hat{\rho}(t) &= e^{i\omega_1 \hat{\mathbf{I}}_x t} \hat{\rho}(0) e^{-i\omega_1 \hat{\mathbf{I}}_x t} \\ &= e^{i\omega_1 \hat{\mathbf{I}}_x t} \frac{1}{Z} \left(\hat{\mathbf{1}} + \frac{\hbar\omega_0}{kT} \hat{\mathbf{I}}_z \right) e^{-i\omega_1 \hat{\mathbf{I}}_x t} \\ &= \frac{1}{Z} + \frac{1}{Z} \frac{\hbar\omega_0}{kT} e^{i\omega_1 \hat{\mathbf{I}}_x t} \hat{\mathbf{I}}_z e^{-i\omega_1 \hat{\mathbf{I}}_x t} \\ &= \frac{1}{Z} + \frac{1}{Z} \frac{\hbar\omega_0}{kT} [\hat{\mathbf{I}}_z \cos(\omega_1 t) + \hat{\mathbf{I}}_y \sin(\omega_1 t)] \end{aligned} \quad (1.66)$$

The y magnetization is:

$$\begin{aligned}
\langle M_y(t) \rangle &= \text{Tr}\{\hat{\rho}(t) \cdot \mu_y\} = \gamma \hbar \cdot \text{Tr}\{\hat{\rho}(t) \cdot \hat{I}_y\} \\
&= \gamma \hbar \frac{1}{Z} \text{Tr}(\mathbf{1} \cdot \mathbf{I}_y) + \gamma \hbar \frac{1}{Z} \frac{\hbar \omega_0}{kT} \text{Tr}\{\mathbf{I}_z \mathbf{I}_y \cos(\omega_1 t) + \mathbf{I}_y^2 \sin(\omega_1 t)\} \\
&= \gamma \hbar \frac{1}{Z} \frac{\hbar \omega_0}{kT} \cdot \frac{1}{2} \sin(\omega_1 t)
\end{aligned} \tag{1.67}$$

Coupling for spin-1 nucleus ^2H in ^{13}C - ^2H bond

Hamiltonian of quadrupolar coupling in frequency unit is:

$$\hat{H}_Q = \frac{eQeq}{2I(2I-1)\hbar} \frac{1}{2} [3\cos^2\theta - 1 - \eta_Q \sin^2\theta \cos(2\phi)] \frac{1}{2} (3\hat{I}_z^2 - \hat{\mathbf{I}} \cdot \hat{\mathbf{I}}) \tag{1.68}$$

In general, $\eta_Q \cong 0$ for aliphatic ^{13}C - ^2H bonds due to the approximate uniaxiality of the electron density in the σ bonds make the resulting field-gradient tensor close to ^2H nucleus. The Hamiltonian reduce to:

$$\hat{H}_Q = \frac{eQeq}{2I(2I-1)\hbar} \frac{1}{2} [3\cos^2\theta - 1] \frac{1}{2} (3\hat{I}_z^2 - \hat{\mathbf{I}} \cdot \hat{\mathbf{I}}) \tag{1.69}$$

The matrix representation of the \hat{I}_α operators in a basis of eigenstates $|1\rangle, |0\rangle, |-1\rangle$ are:

$$\mathbf{I}_x = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \mathbf{I}_y = i \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad \mathbf{I}_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{1.70}$$

Matrix representation for \mathbf{H}_Q is derived as:

$$\begin{aligned}
\mathbf{H}_Q &= \frac{\omega_Q}{3} (3\mathbf{I}_z \mathbf{I}_z - I(I+1) \cdot \mathbf{1}) \\
&= \frac{\omega_Q}{3} \left(3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \\
&= \frac{\omega_Q}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned} \tag{1.71}$$

where the pre-factor of the matrix is:

$$\frac{\omega_Q}{3} = \frac{eQeq}{4\hbar} \frac{1}{2} [3\cos^2\theta - 1] = \frac{1}{4} \chi \frac{1}{2} [3\cos^2\theta - 1] \tag{1.72}$$

$\chi = \frac{eQeq}{4\hbar}$ is often terms as quadrupolar dipolar coupling constant and has value of 170kHz for a typical aliphatic C- ^2H bond.

For an initial density operator \mathbf{I}_x , evolution under \mathbf{H}_Q is given by¹:

$$\rho(0) = \mathbf{I}_x \xrightarrow{\mathbf{H}_Q} \rho(t) = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 & e^{-i\omega_Q t} & 0 \\ e^{i\omega_Q t} & 0 & e^{i\omega_Q t} \\ 0 & e^{-i\omega_Q t} & 0 \end{bmatrix} \quad (1.73)$$

Element of final matrix is derived from $\rho(0)_{mn} \cdot e^{-i(\mathbf{H}_{mm} - \mathbf{H}_{nn})t}$.

Signal is given by¹:

$$S(t) \sim \text{Tr}\{\rho(t) \cdot (\mathbf{I}_x + i\mathbf{I}_y)\} = e^{i\omega_Q t} + e^{-i\omega_Q t} \xrightarrow{\text{Fourier Transform}} \delta(\omega - \omega_Q) + \delta(\omega + \omega_Q) \quad (1.74)$$

The corresponding spectrum is two lines at $\pm\omega_Q$ relative to the Larmor frequency. They are transitions from $m = 0 \rightarrow 1$ and $m = -1 \rightarrow 0$ respectively. Note only transition $\Delta m = \pm 1$ is permitted in NMR.

$$\pm\omega_Q = \pm \frac{3}{4} \chi \frac{1}{2} [3\cos^2\theta - 1] \quad (1.75)$$

Case-1: When $\theta = 0^\circ$, $\pm\omega_Q = \pm \frac{3}{4} \chi$;

Case-2: When $\theta = 54.7^\circ$, $\pm\omega_Q = 0$;

Case-3: When $\theta = 90^\circ$, $\pm\omega_Q = \mp \frac{3}{8} \chi$.

For a random orientated sample, all orientations are equally probable. The number of the nuclear spin is proportional to the surface area of the sphere of radius r ($2\pi r \sin \theta d\theta$) and the probability of finding nuclear spins oriented between θ and $\theta + d\theta$ is roughly proportional to $\sin \theta$ ($\frac{dN}{N} =$

$\frac{2\pi r \sin \theta d\theta}{4\pi r^2} = \frac{1}{2} \sin \theta d\theta$, N is the total number of spins),⁵ so the expecting spectrum is:

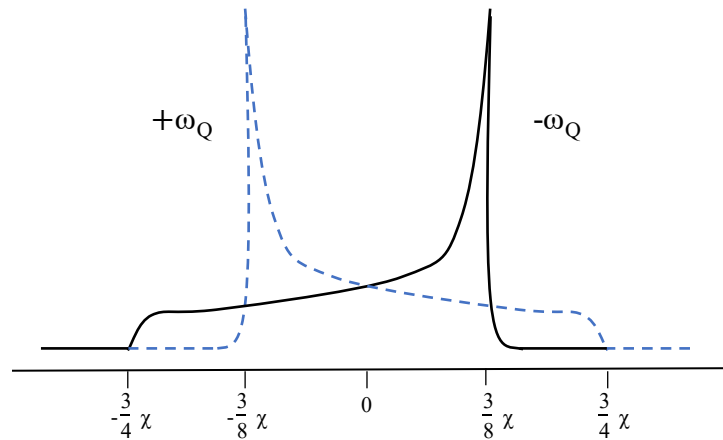


Figure 1.7. Stick spectrum of a $^{13}\text{C}-^2\text{H}$ bond.

1.1.3 ssNMR techniques

Magical angle spinning (MAS)

From Equation 1.46, the chemical shift is dependent on the orientation of nuclear spin chemical shielding tensor with respect to the external magnetic field. It's common that solid samples have many crystallites randomly oriented with respect to B_0 , each possible orientation has a contribution to chemical shift anisotropy resulting in a broad peak. The NMR peak that a unoriented sample produced is called powder lineshapes. Due to the dependence of the NMR frequency and the nuclear spins orientation, it is reasonable that rapid rotational or translational motion of molecules averages the powder pattern on some extent. If the sample motion falls in fast-motion region, which means the motional rate exceeds internal couplings. The fast molecular motion averages the orientations of neighboring nuclei leads to the averaging of internal couplings and only the isotropic peak will be left (e.g. the orientations of the dipolar coupling vectors rapidly change due to molecular rotation).

If motion is not fast enough, it still has an influence on the NMR frequency. The consequence of motional averaged is that the NMR frequency now has a segmental orientation dependence which can be summarized as:¹

$$\bar{\omega}(\theta_a, \phi_a) = \bar{\delta} \frac{1}{2} [3\cos^2\theta_a - 1 - \bar{\eta}\sin^2\theta_a \cos(2\phi_a)] \quad (1.76)$$

where (θ_a, ϕ_a) denote the polar coordinates of the B_0 field in the PAF of the averaged tensor $\bar{\sigma}$. The illustration of (θ_a, ϕ_a) can be found in Figure 1.4 but the chemical shift tensor of a single molecule should be instead with an average tensor which represents the effective interaction tensor that results from the random or isotropic motion of molecules. Macroscopic sample rotating is another method to achieve interaction tensor averaging. Introducing the rotation, frequency becomes:¹

$$\omega = \left\{ \frac{1}{2} \delta [3\cos^2\theta_p - 1 - \eta\sin^2\theta_p \cos(2\phi_p)] \right\} \cdot \left[\frac{1}{2} (3\cos^2\theta_r - 1) \right] \quad (1.77)$$

Where θ_r is the angle between rotation axis and B_0 field; (θ_p, ϕ_p) denotes the polar angles of the rotation axis in PAF (defined in Figure 1.4). The second square bracket is the impact from sample rotation. The rotation angle in the laboratory frame can be described as:

$$\langle 3\cos^2\theta_r(t) - 1 \rangle = \frac{1}{2} (3\cos^2\alpha - 1)(3\cos^2\beta - 1) \quad (1.78)$$

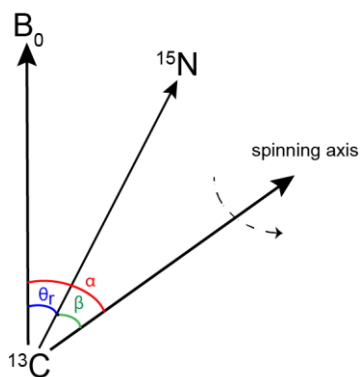


Figure 1.8. Schematic representation of the geometry of the ^{13}C - ^{15}N vector in ssNMR sample under MAS.

Figure 1.8 is the illustration of α , β , and θ_r . Clearly when $\alpha = 54.7^\circ$, the whole term reduces to zero. The angle is called magic angle. Basically, ssNMR experiments should be conducted under this condition to eliminate the line broadening caused by CSA and dipolar coupling to obtain a narrower linewidth. If the spinning frequency is faster than the largest coupling in the sample by a factor of 3 or 4, the anisotropic interaction described by second rank tensor can be averaged to zero resulting in an isotropic peak. A slower spinning speed leads to a series of spinning sidebands in addition to the isotropic line separated by spinning frequency shown in Figure 1.9.

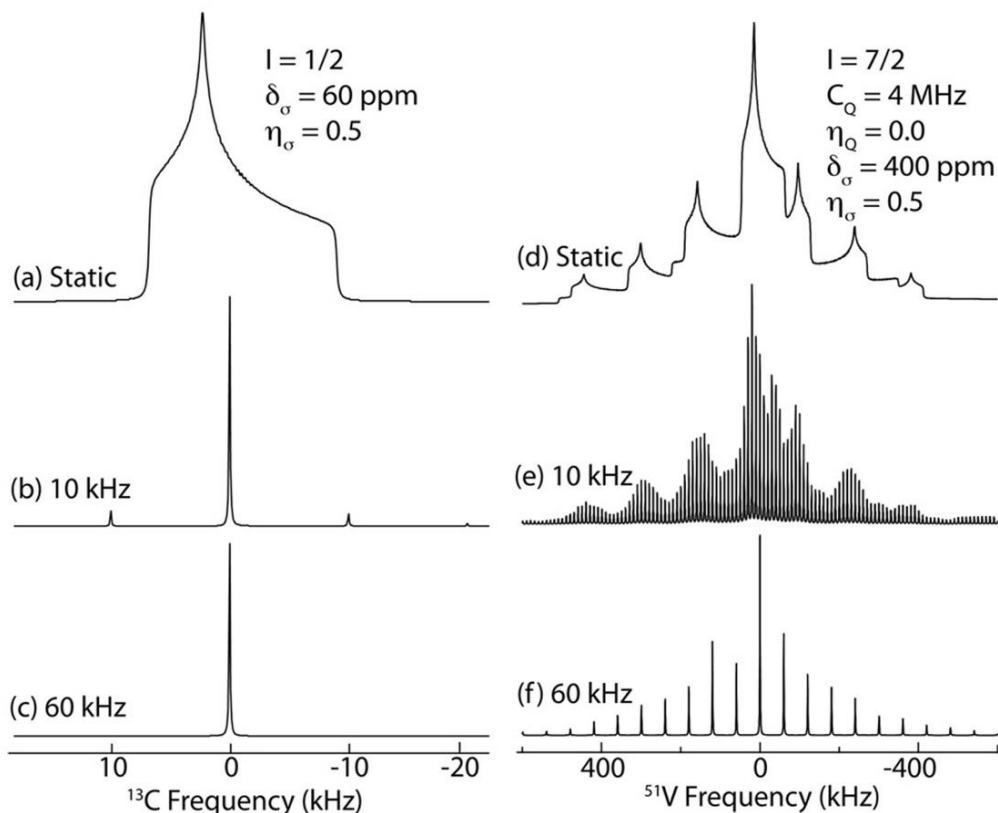


Figure 1.9. The effect of MAS on ^{13}C and ^{51}V spectra at spinning rate of 10 kHz and 60 kHz. The broad ^{13}C powder pattern resulting from chemical shift at static (a) is broken to an isotropic peak and two spinning sidebands (b). The 60 kHz spinning rate exceeds the magnitude of the anisotropic interaction and averages out the chemical shift anisotropic interaction (c). It is very common that the spinning frequency is not fast enough to average out neither the CAS nor the quadrupolar interaction for half-integer quadrupolar nuclei in low-symmetry environments. Therefore, the broad ^{51}V powder pattern caused by quadrupolar and chemical shift interactions at static (d) is broken up to a series of spinning sidebands (e and f).⁶

Cross Polarization

The low abundance of the nuclei implies a low sensitivity. Besides, the relaxation of the low abundance nuclei tends to be very long. Observing low sensitivity nuclei with direct polarization is very time-consuming. Cross polarization is a strategy which reduces the data acquisition time and increase the signal-to-noise ratio as well.

In cross polarization experiment, the magnetization aligned at a strong \vec{B}_1 field exceeding the dipolar couplings is spin-locked, which means the magnetization will not dephase neither by chemical shift nor dipolar couplings. It only relaxes with the longitudinal relaxation in the rotating frame $T_{1\rho}$. If the abundant spin I magnetization is spin-locked, and the simultaneous irradiation is applied to dilute spin S under the Hartmann-Hahn cross polarization (HHCP) condition:

$$\gamma_I B_{1,I} = \gamma_S B_{1,S} \quad (1.79)$$

the flip-flop term B (Equation 1.48) of the dipolar Hamiltonian induce the rapid magnetization transfer between spin I and S.

The match condition indicates that the two spins are having the same effective energy in their own rotating frames, and precessing at equal rates. Usually, the effective field of spin I and S are greater than their chemical shifts and shielding tensor, so the chemical shift and CAS Hamiltonian are negligible. The average heteronuclear dipolar coupling is the reason for polarization transfer from the abundant spin to the rare spin. The transferred polarization of rare spin is enhanced by ratio of γ_I/γ_S at maximum. For example, $\gamma_{1H}/\gamma_{13C} \approx 4$; $\gamma_{1H}/\gamma_{15N} \approx 10$. Additionally, the rare spins not only benefits from the enhancement but also shortening the recycling delay due to the shorter relaxation time of protons.

When sample is spun, the rotation of sample introduces modulations the orientation of the interaction tensor with respect to the external magnetic field which are oscillating at frequencies of ω_R and $2\omega_R$.^{7,8} If the difference of effective field is ω_R or $2\omega_R$, the dipolar coupling will be an effective average value over one rotor period and is the origin for polarization transfer. If spinning rate is much greater than chemical shift offset, the contribution from chemical shift offset can be neglected and the match condition becomes:

$$\sqrt{\Omega_I^2 + \omega_{1,I}^2} \pm \sqrt{\Omega_S^2 + \omega_{1,S}^2} = \sqrt{\omega_{1,I}^2} \pm \sqrt{\omega_{1,S}^2} = \gamma_I B_{1,I} \pm \gamma_S B_{1,S} = n\omega_R, \quad n = 1,2 \quad (1.80)$$

If the spinning rate is comparable to dipolar coupling, there exist some match conditions that chemical shifts (Ω) are partially responsible for polarization transfer, and the match condition is:

$$\sqrt{\Omega_I^2 + \omega_{1,I}^2} \pm \sqrt{\Omega_S^2 + \omega_{1,S}^2} = n\omega_R, \quad n = 1,2 \quad (1.81)$$

This equation indicates the basis for spectrally induced filtering in combination with cross polarization (SPECIFIC-CP). With the contribution of chemical shift, if the RF field is carefully chosen, the desired peak at certain isotropic chemical shift can be filtered out. The NCACX, NCOCX and CONCACX experiments are created based on it, the ¹⁵N polarization can be transferred to either the carbonyl carbons or alpha carbons by SPECIFIC-CP.

T₂ measurement---Hahn-echo

After a single 90° x pulse is applied to the sample, magnetization is rotated to y-axis from z-axis. Since the different portions of sample are subjected to external magnetic field differing slightly

from the inhomogeneity and the existence of chemical shift differences, the magnetization vectors will fan out. Other than that, the transverse relaxation is the loss of the phase coherence (phase coherence: the two signals have a constant relative phase) or randomization the spin in transverse plane the so the net magnetization of xy-plane becomes zero. If the vectors' spread due to inhomogeneous field and different chemical shift is recovered, the true transverse relaxation rate can be detected. Hahn-echo is a general method to detect transverse relaxation rate. A 180° y pulse is applied after an evolution time τ and an echo will occur after the same evolution time. By adjusting evolution time, a series FID can be acquired and the amplitude of the successive echoes decay exponentially. T_2 can be found from the envelope of the echoes.

$$90_x^\circ - [\tau - 180_y^\circ - \tau - \text{echo}]_n$$

The effect of Hahn-echo can be calculated as follows:

In the rotating frame, the Hamiltonian of \vec{B}_1 field is $\hat{H}_1 = -\vec{M} \cdot \vec{B}_1 = -\gamma B_1 \hat{I}_x$. An initial 90_x° pulse rotates the magnetization to y-axis and the density operator is $\hat{\rho}(0) = e^{i\gamma B_1 \hat{I}_x t} \hat{I}_z e^{-i\gamma B_1 \hat{I}_x t} = -\hat{I}_y$.

The density operator at the time of τ after 180_y° pulse is:

$$\hat{\rho}(t) = \hat{U}(t) \hat{\rho}(0) \hat{U}^{-1}(t) \quad (1.82)$$

$\hat{U}(t)$ is propagator which pushes the density operator ahead in time. Propagator at the time τ after 180_y° pulse is:

$$\begin{aligned} \hat{U}(2\tau) &= e^{-i\omega \hat{I}_z 2\tau} e^{i\pi \hat{I}_y} e^{-i\omega \hat{I}_z \tau} \\ &= e^{-i\omega \hat{I}_z \tau} e^{i\pi \hat{I}_y} e^{-i\omega \hat{I}_z \tau} e^{-i\pi \hat{I}_y} e^{i\pi \hat{I}_y} \\ &= e^{-i\omega \hat{I}_z \tau} e^{i\omega \hat{I}_z \tau} e^{i\pi \hat{I}_y} \\ &= e^{i\pi \hat{I}_y} \end{aligned} \quad (1.83)$$

$$\hat{\rho}(2\tau) = e^{i\pi \hat{I}_y} (-\hat{I}_y) e^{-i\pi \hat{I}_y} = -\hat{I}_y \quad (1.84)$$

Density operator at 2τ is the same at the density operator after initial pulse and the evolution caused by chemical shift has been recovered. Moreover, resonance offset and heteronuclear dipolar coupling which are interactions linear in \hat{I}_z , can be recovered by Hahn-echo as well.

Paramagnetic relaxation enhancement (PRE)

Paramagnetic relaxation enhancement is a very promising tool to probe long-range distance of biomolecules up to 35 Å depending on the paramagnetic source. The dipolar coupling between

unpaired electron of paramagnetic source and target nuclear isotope increase the relaxation rate of the nucleus of interests and this increase is distance dependent.

A paramagnetic center produces the dipolar coupling between the nuclear of interest and the unpaired electron of the paramagnetic species which has an impact on the relaxation rate. This enhancement is a distance- and orientation-dependent effect which can be used to explore structural information of biochemicals.⁹ In general, the paramagnetic species enhance the relaxation of the nuclear spins at well-defined locations, resulting in line broadening and intensity reduction, which is inversely proportional to each other. The information can be exploited to determine the location of small peptides in lipid.¹⁰ In my project, the ¹³C of acyl chain of POPC/POPG mixture is the nucleus of interest and Mn²⁺ served as paramagnetic origin. The dipolar interaction between the ¹³C in lipid acyl chain and the unpaired electrons of Mn²⁺ will contribute to accelerate T₂ relaxation. The enhancement of the relaxation rate can be described as:

$$\Gamma_2 = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \gamma_I^2 g_e^2 \mu_B^2 S(S+1) \frac{1}{r^6} [4J(0)+3J(\omega_I)]$$

$$J(\omega) = \frac{\tau_c}{1+(\omega\tau_c)^2} \quad (1.85)$$

$$\frac{1}{\tau_c} = \frac{1}{T_{2e}} + \frac{1}{\tau_r} + \frac{1}{\tau_m}$$

where Γ_2 is the relaxation rate enhancement, it can be found from $\Gamma_2 = R_2^{w/para} - R_2^{w/o para}$, μ_0 is the vacuum permeability; γ_I is the gyromagnetic ratio of ¹³C; g_e is the g factor of electron; μ_B is the electron Bohr magneton;¹¹ S is spin quantum number and is 5/2 for Mn²⁺; r is the electron-nucleus distance; J(ω) is the spectral density and $\frac{\omega_I}{2\pi}$ is the ¹³C Larmor frequencies. The correlation time τ_c is the inverse sum of the electronic spin-spin relaxation time T_{2e}; the rotational correlation time of the molecule τ_r and the residence time of the Mn²⁺ near ¹³C τ_m .

In my PRE experiment, Mn²⁺ was chosen as paramagnetic species since it only binds to membrane surface and does not penetrate lipid, making a well-defined location. ¹³Cs of acyl chain of POPC/POPG membranes will exhibit accelerated relaxation when Mn²⁺ presence. The enhancement of relaxation rate of a ¹³C is hypothesized to be proportional to the probability of chain protrusion into the headgroup region and the probability that a Mn²⁺ is bound to a headgroup close to the protruded chain. Since the concentration of Mn²⁺ is already known by sample preparation, it is possible to determine the distance between ¹³C and the location of Mn²⁺ once the

probability of chain protrusion is fixed.^{12,13} PRE can also be used as a semi-quantitative approach to estimate the population of molecules at different states. In our case, the measured Γ_2 is hypothesized to be proportional to probability of lipid tail protrusion and is a weighted average value from the larger Fp-adjacent and smaller more distant values. By fitting our PRE data, we found the ratio of the enhancement between Fp-adjacent and Fp-distant lipids which is consistent with some computational simulation results.¹⁴

Relaxation in solid state NMR

Unlike relaxation is widely used in characterizing dynamic in solution NMR, it is less common in solid state NMR. The measurement and data analysis are challenging, especially for transverse relaxation. Although MAS helps to average the CSA and dipolar couplings, the higher spinning frequency is required to eliminate the strong anisotropic interaction and dipolar couplings. The conventional MAS rate up to ~ 50 kHz is not fast enough to average out the multi-spin ^1H - ^1H dipolar couplings.¹⁵ The line broadening caused by dipolar coupling affect both longitudinal and transverse relaxation in solid samples. The presence of dipolar coupling affects the T_1 relaxation process by influencing the rate at which nuclear spins return to thermal equilibrium and the relaxation may occur at a slower rate. The dipolar coupling often shortens the transverse relaxation time by contributing dephasing of the transverse magnetization.¹⁶ MAS influences the dipolar coupling and the reduction dipolar coupling leads to longer transverse relaxation time.

Overall, the relaxation in solids is complicated that solution. But the relaxation can provide rich information about molecular motion and relaxation measurement is a powerful tool to explore biomolecules dynamics.¹⁷

Spin diffusion

Spin diffusion is Z magnetization transfer spatially driven by homonuclear dipolar coupling. When the inhomogeneous polarized has been created, which means the spins of the sample do not have a uniform or homogenous polarization distribution, the polarization can spread through the space mediated by homonuclear dipolar coupling. If the spin A is excited and spin B is in close proximity, the flip-flop term (term B of the dipolar coupling interaction) of the homonuclear dipolar coupling will transfer the polarization from A to B to reach an equilibrium of the magnetization of the two spins, so the polarization would be equally distribution across the sample. Spin diffusion is widely used in 2D homonuclear correlation measurements and 3D measurements in the mixing step. In multidimensional ssNMR experiments, a longer mixing time allows the polarization transfer to a

distant spin due to the nature of spatial diffusion. By manipulating mixing time, the polarization can spread to neighboring spins depending on distance.

Multidimensional NMR and the correlation peaks (cross peaks)

1D NMR spectrum of proteins is difficult to interpret due to the peak overlapping caused by large number of degenerated peaks. This can be solved by acquisition of multidimensional NMR spectra. Compared with 1D spectra in which all signals are superimposed in one dimension, signals are spread over a surface (2D) or in a three-dimensional space (3D,4D). The polarization transfer either through bond (scalar coupling in solution NMR) or space (dipolar coupling in ssNMR). A basic 2D NMR measurement is briefly described as following:

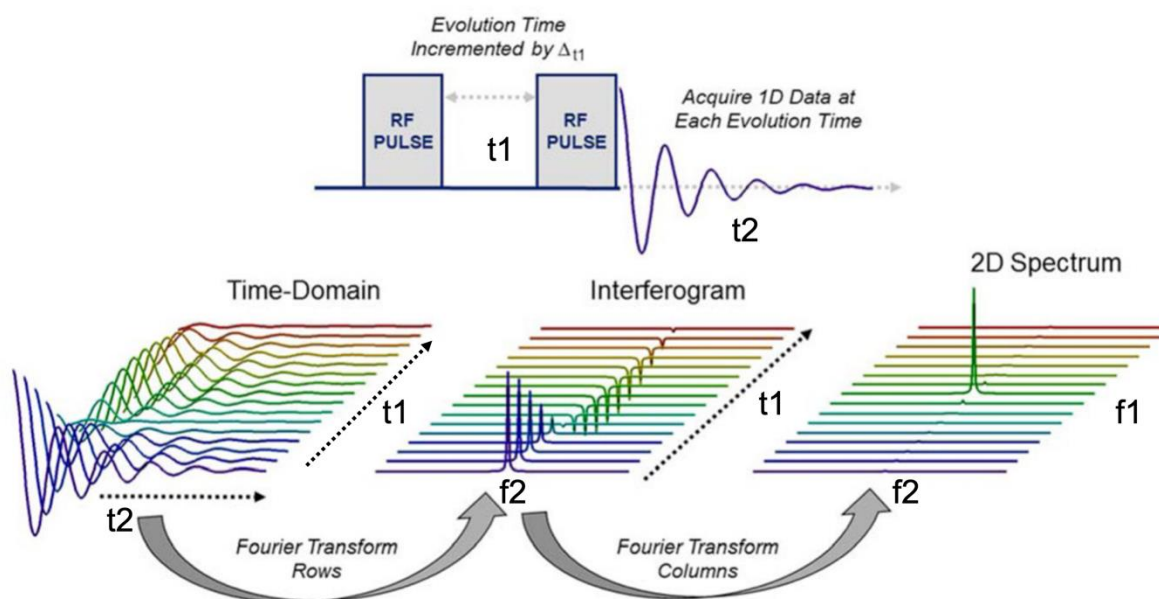


Figure 1.10. Illustration of 2D NMR experiments and data processing. The 2D NMR experiment usually consists of magnetization preparation, evolution, mixing and data acquisition blocks. Fourier transform is applied to rows data and then columns data. The final processed spectrum is double frequency labeled.¹⁸

First step is the magnetization preparation, and a simple way is to apply a 90° pulse to generate transverse magnetization. Then, the evolution time t_1 is incremented by Δt_1 systematically in a series of separate 1D experiments. More increments mean more data will be recorded so the spectral resolution can be improved but it takes a longer time. Following each evolution time t_1 , another R.F. pulse manipulates the coherence obtained from evolution time transferring into a detectable signal, which is often termed as mixing period. It is very common that in mixing period, the magnetization can be transferred via spin diffusion. Finally, the signal recorded at the

acquisition is frequency labeled. The time-domain data can be treated as a matrix. A row in the matrix is the data of t_2 and a column is the data of t_1 . To process the 2D NMR data, Fourier transform is firstly applied to the row data and then the column data (shown in Figure 1.10). Eventually, the final processed data is a 2D spectrum, with frequency ω_1 , corresponding to the evolution in t_1 , and ω_2 for t_2 .

The cross peaks of multidimensional spectrum can be interpreted to deliver the correlation information between spins and further translated to connectivity information. For solution NMR, the commonly used 2D experiments like COSY, HSQC (Heteronuclear Single Quantum Coherence), and HMBC (Heteronuclear Multiple Bond Correlation) can be used to elucidate bond connectivity. The magnetization transfer via scalar coupling. HSQC examines proton-carbon single bond correlations and HMBC examines the correlations between carbon and protons separated by two bonds.¹⁹ Similarly, NCACX, NCOCX, and CONCACX experiments in solid state are the measurements of bond connectivity through magnetization transfer via dipolar coupling.

Rotational-echo double resonance (REDOR)

To simplify and obtain a high-resolution ^{13}C spectrum of a biosolid sample, fast spinning speed of sample rotation is often required to reduce the number of sidebands arising from carbon shift anisotropy. However, if a weak dipolar coupling is desired to be measured, the fast spinning speed exceeding dipolar coupling will average it to zero. For example, the ^{13}C - ^{15}N dipolar coupling is $\sim 1\text{kHz}$ and spinning rate for high resolution condition is usually several kilohertz. Dipolar coupling constant is inversely proportional to the cube of the internuclear vector distance. Determination of dipolar coupling constant is of great interest as it is an indirect measurement of distance. REDOR is one of ssNMR pulse sequences to detect weak heteronuclear dipolar coupling.

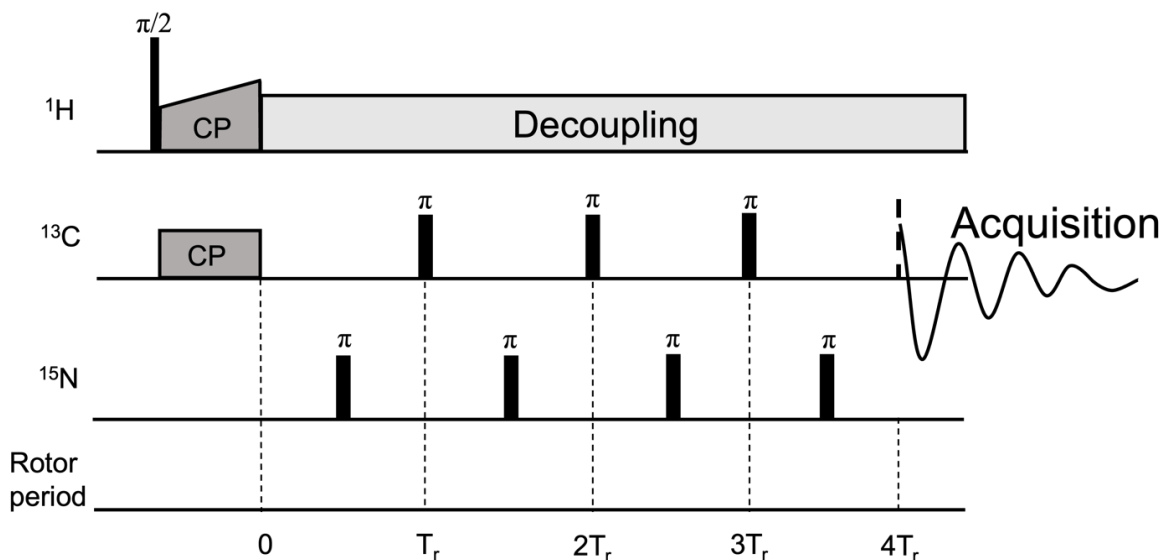


Figure 1.11. A typical ^{13}C - ^{15}N REDOR pulse sequence. The build-up curve of ^{13}C - ^{15}N heteronuclear dipolar coupling intensity can be created by accumulating the current block.

Under MAS, chemical shift anisotropy and heteronuclear dipolar coupling are often eliminated at the end of a rotor period. The time-dependent Hamiltonian for a heteronuclear pair of spin- $\frac{1}{2}$ nuclei with MAS is:

$$\hat{H}_D = d_c [\sin^2 \beta \cos 2(\omega_r t + \alpha) - \sqrt{2} \sin 2\beta \cos(\omega_r t + \alpha)] \hat{S}_z \hat{I}_z \quad (1.86)$$

where d_c is dipolar coupling constant; β is the polar angle between heteronuclear dipolar vector and rotating frame z-axis and α is the azimuthal angle with respect to the rotation frame.²⁰

The dipolar coupling is a product of spatial and spin part. Spatial part is controlled by spinning and spin is manipulated by pulses. As stated earlier, the Hanh-echo refocuses chemical shift as well as heteronuclear dipolar coupling. Based on this, if a π -pulse is applied to spin S at the half rotor period, the sign of \hat{S}_z will become opposite so the dipolar coupling reaches to an effective averaged value over one rotor period. By switching on and off the π -pulses at S channel, the spectra with and without heteronuclear dipolar coupling can be collected and their difference is heteronuclear dipolar coupling signal, demonstrated in Figure 1.12.²¹

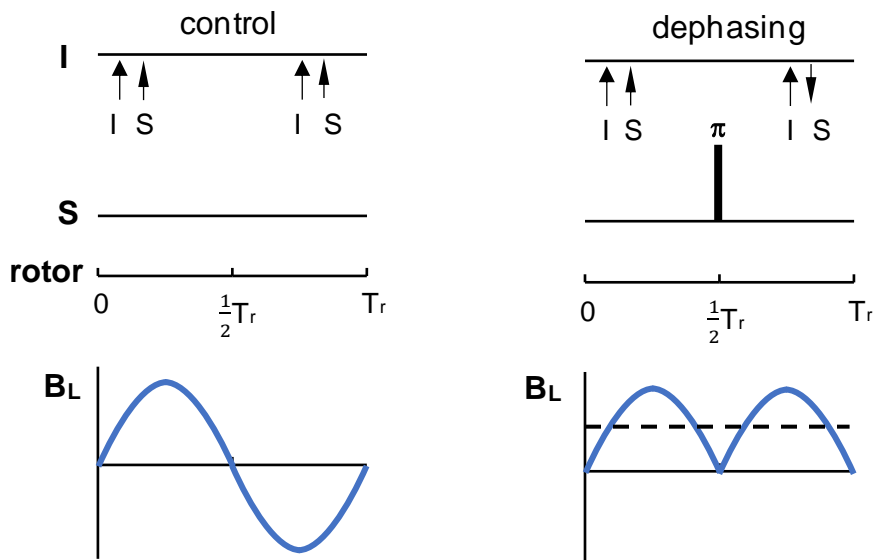


Figure 1.12. Illustration for the influence of a π -pulse at S channel on heteronuclear dipolar coupling of a spin- $\frac{1}{2}$ pair in REDOR. Without π -pulse, the integral of heteronuclear dipolar coupling is zero over one rotor period but an effective average dipolar coupling with a π -pulse applied.

As shown in Figure 1.11, in ^{13}C - ^{15}N dipolar experiment, a ramp CP is applied and transfers the polarization from ^1H to ^{13}C . The ramp CP is chosen because of the distribution of Larmor frequencies, and the resonance offset field. For a powder sample, the CP transfer efficiency would be different for each molecular orientation. A ramp CP would cover a certain range of frequency and improve the transfer efficiency. After $^1\text{H} \rightarrow ^{13}\text{C}$ CP, REDOR experiment is broken up with two parts: one with rotor synchronized π pulses applied at ^{15}N channel at the middle of each rotor period (S_1) and the other without pulses at ^{15}N channel (S_0). The π pulses at ^{13}C channel refocus the ^{13}C isotropic chemical shift and ^{13}C CSA is averaged out by MAS. The dephasing π pulses of ^{15}N recouple the ^{13}C - ^{15}N dipolar coupling averaged out in each rotor period by MAS in S_0 experiment. The dipolar coupling leads to an intensity reduction of S_1 compared with S_0 . The REDOR dephasing is defined as:

$$\frac{\Delta S}{S_0} = \frac{S_0 - S_1}{S_0} \quad (1.87)$$

The dephasing buildup curve of $\Delta S/S_0$ vs dephasing time τ can be achieved, where τ is the time period after CP but before FID acquisition controlled by rotor periods. The ^{13}C - ^{15}N dipolar coupling can be obtained by fitting using SIMSPON program. The dephasing of ^{13}C - ^{15}N dipolar coupling at different dephasing time can be simulated by SIMPSON and the simulated dephasing

can be used for population fitting (see more details in Chapter 6). The ^{13}C - ^{15}N internuclear distance can be calculated by:

$$d_{\text{CN}}(\text{Hz}) = 3080/r^3 (\text{\AA}) \quad (1.88)$$

1.2 Introduction of influenza virus

Influenza, commonly known as the flu, is an infectious disease caused by influenza virus. It is a respiratory pathogen and classified into four types (A, B, C and D) based on the difference of internal protein antigens (e.g., PA, PB, PB1), nucleoprotein (NP) and matrix protein (M) of the virus.²² Type A is found in a wide variety of warm-blooded animals including human while type B infects only humans. The type A viruses are the most virulent human pathogens among the four influenza types which can cause the severest disease and can be subdivided into different serotypes based on the antibody response to these viruses.²³ Influenza spreads around the world in yearly outbreaks, resulting in about three to five million cases of severe illness and about 290,000 to 650,000 deaths.²⁴ The most effectively method to prevent influenza virus from infection is taking vaccine. However, because of the high mutation rate of the virus, there is no particular influenza vaccine conferring protection more than a few years. The vaccine is reformulated each flu season to combat the current circulating strain.²⁴

1.2.1 Structure of Influenza A virus

Influenza A virus is an RNA virus belonging to the family Orthomyxoviridae. It is enveloped virus with viral capsid surrounded by host-derived lipid membranes (Figure 1.13). Morphologically, the virus can be a sphere with diameter of ~ 100 nm or a filament reaches up to $20 \mu\text{m}$ in length.²⁵

The hemagglutinin (HA) is glycoprotein of influenza virus with ~ 550 amino acid residues, located on the surface of the virus. The HA is synthesized as a single polypeptide and derived into a complex consisting of HA1 and HA2 subunits after cleavage. The HA1 subunit is responsible for binding the virus with the sialic acid receptor on the host membrane while the HA2 subunit plays critical role in fusion between host cell and virus. For all HA subtypes, the cytoplasmic tail, the transmembrane domain, the stalk region which is a triple-stranded coiled-coil of α helices extending from the membrane and the fusion peptide are the most conserved regions.^{26,27} The ~ 23 amino acids of N-terminus of HA2 are highly conserved commonly referring as Influenza fusion peptide (IFP).

Neuraminidase (NA) is a transmembrane glycoprotein present on the viral surface as a tetramer and its approximate ratio with HA is around 300HA:40NA, and the quantity ratio is influenced by

the subtype of the virus.²⁸ NA is very important to cleave the HA cellular binding receptor-terminal sialic acid from the cell-surface glycans and facilitate the release of budded virions. Compared with HA, NA-lipid interaction is less explored. Studies indicate that expression of NA alone in the absence of matrix protein and HA was sufficient to generate and release NA containing particles, indicating that NA is capable of inducing membrane curvature.^{26,29}

Moreover, there also exist matrix proteins, M1 and M2, where M1 is a major determinant of influenza virus morphology through its ability to modulate membrane curvature. Research illustrates that M1 interacts with the lipid bilayer producing an outward bending of the membrane and this is postulated to be the major driving force of influenza budding. M2 is a transmembrane homotetramer.² When the influenza virus is endocytosed into the host cell, virus-envelop-bound M2 ion channel is open in response to the low pH of the endosome and acidifies the virion, which in turn releases the viral ribonucleoprotein complex into the host cell.^{30,31}

In the viral core, there exist genetic materials including eight segments of negative-sense single-stranded RNA (i.e., complementary to the mRNA sense).²² These RNAs are associated with nucleoprotein (NP) and RNA polymerase complex (PA, PB1, PB2) to form viral ribonucleoprotein (vRNP) complex. The eight negative-sense RNA of influenza A virus encode 10 products, including PB1, PB2, PA polymerases, HA, NP, NA, M1 and M2 proteins, and nonstructural NS1 and NS2.

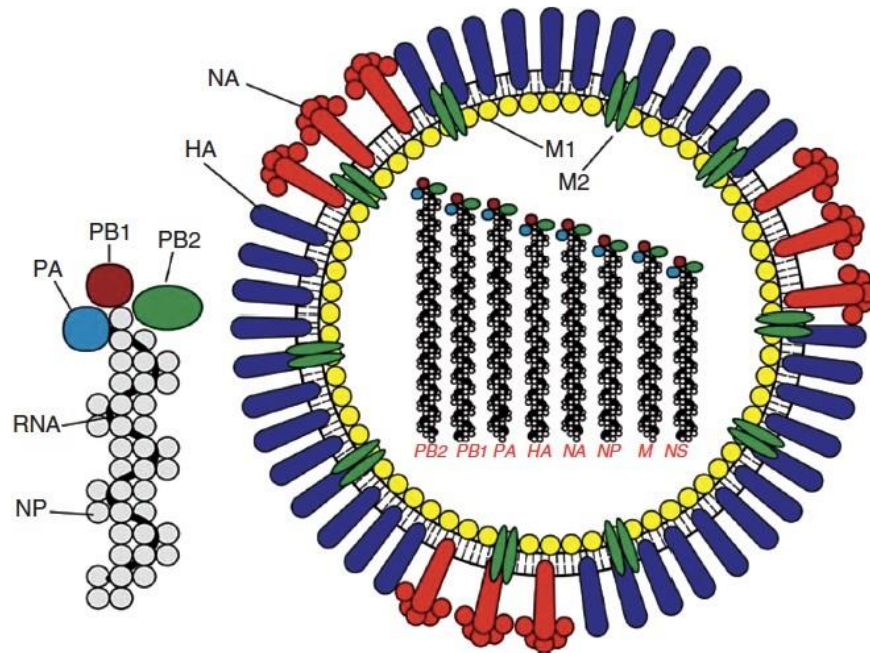


Figure 1.13. Structure of influenza A virus.²³ The viral capsid consists of eight single strand RNA genome is surrounded by the membrane containing three proteins.

1.2.2 Replication of influenza A virus

The overview of the replication of influenza virus is shown in Figure 1.14. The viral HA binds to sialic acid receptor on the surface of host membrane followed by endocytosis. Then acidic lysosomes fuse with the endosome and causes the pH of endosome dropping to ~6. With the endosome migrating towards nucleus, the endosomal pH is reduced to ~5.5 as late endosome. This low pH of the late endosome triggers conformational changes of trimeric HA2 subunit of HA, followed by fusion between viral and endosomal membranes.³²

In the meantime, the acidification of the endosome also makes the M2 proton channel open to acidify the viral core resulting in uncoating of the viral genome. At lower pH, the electrostatic force connected RNPs and M1 protein is strongly weakened so RNPs can be released into cytoplasm of host cell to transcribe and replicate.

After being released into the cytoplasm, RNPs enter the nucleus and viral genome will be replicated.³³ However, the negative-sense single-stranded RNA of influenza virus cannot be translated into protein directly. Instead, viral RNA polymerase enzyme helps these negative RNA converts into positive-sense RNA which then can act as mRNA to be translated into viral proteins in endoplasmic reticulum in the host cell cytosol. Newly synthesized PA, PB1, and PB2 ,NP and

M1 proteins are transported into the host nucleus and the new RNPs will be assembled.^{34,35} The assembled RNPs are transported out of the nucleus and to the plasma membrane for incorporation to form new virions.³⁶

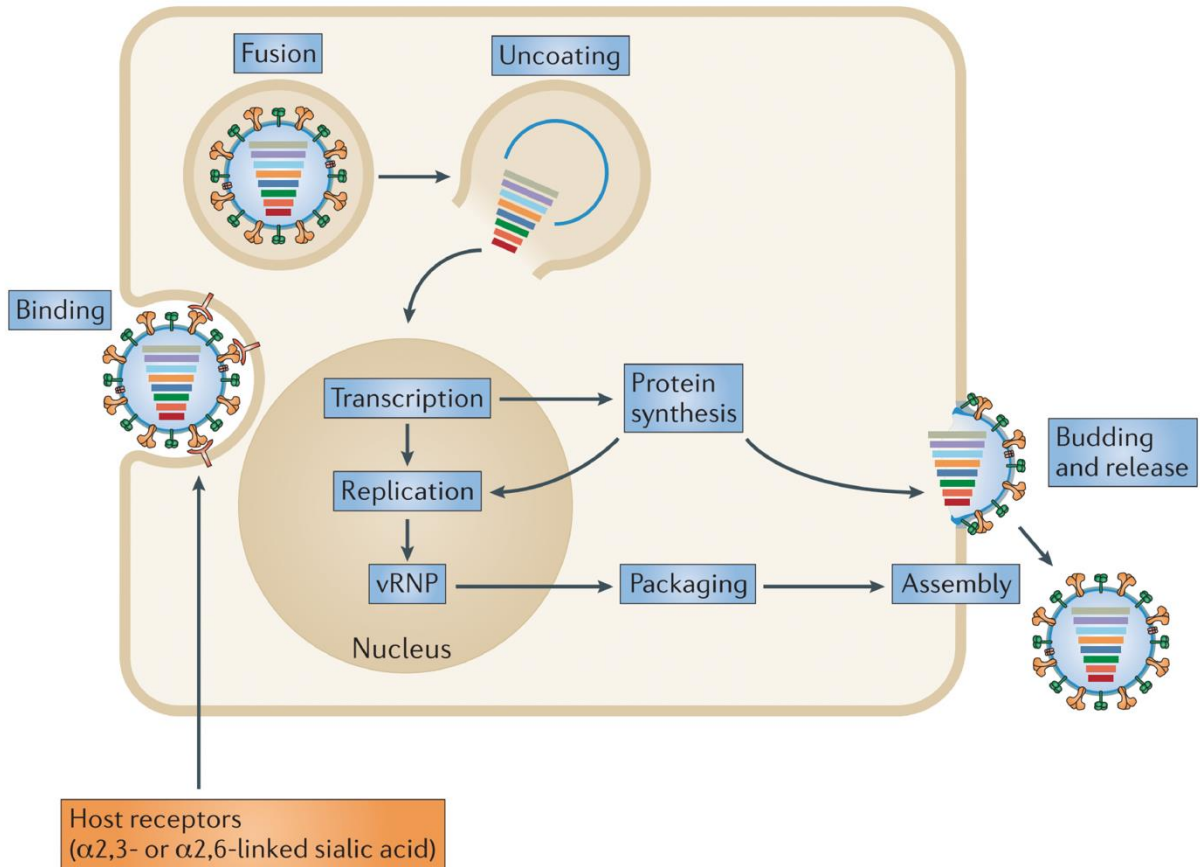


Figure 1.14. Replication of influenza virus.³⁷

1.2.3 Structure of HA

HA has receptor-binding sites and is the fusion glycoprotein of influenza virus and is also the target for infectivity-neutralizing antibodies.³⁸ The viral HA is initially synthesized as a fusion-inactive precursor of ~526 residues to prevent premature fusion. The crystallographic structure of soluble fragment of HA at neutral pH reveals that it is a single polypeptide as a trimer with two structurally distinct regions: a globular region of antiparallel β -sheet HA1 and a triple stranded coiled-coil of α -helices stalk region HA2.

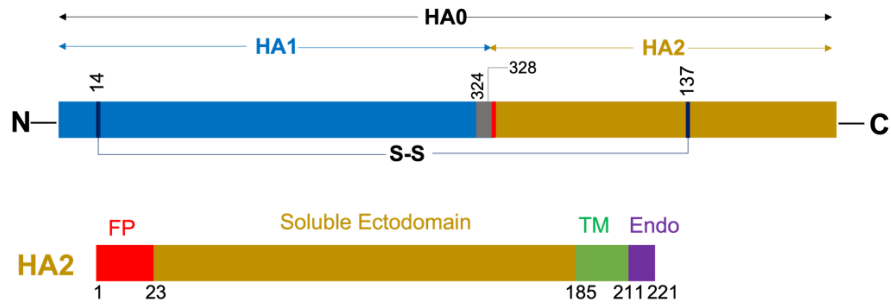


Figure 1.15. Schematic illustration of HA and HA2 sequences. TM and Endo represent transmembrane domain and Endo domain respectively.

In order to carry out its functions, HA must undergo a priming step which is proteolytic cleavage to render it fusion competent.⁶ Cleavage of HA0 at site of ~325 generates the C terminus of HA1 and N terminus of HA2.³⁸ After cleavage, the HA0 precursor is proteolytically cleaved to two subunits, HA1 and HA2 which remain covalently bound by a disulfide bond.²⁷

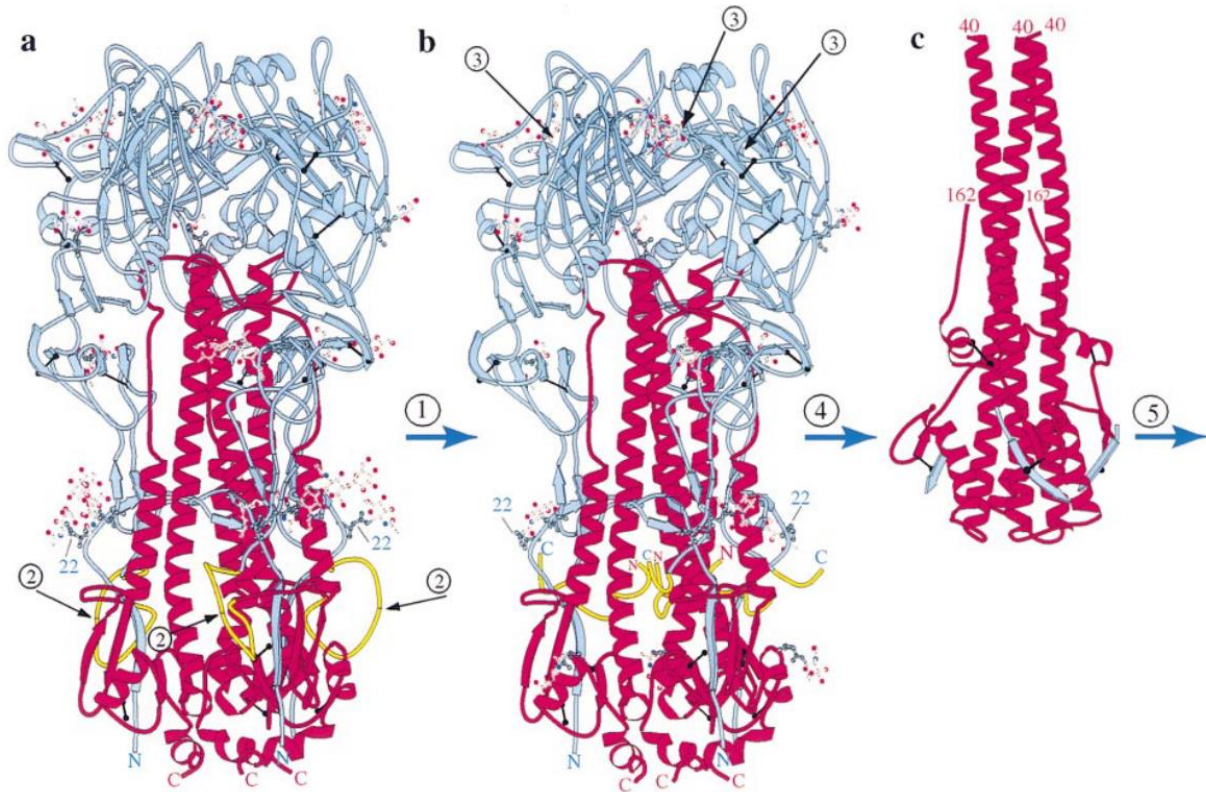


Figure 1.16. Three conformations of HA trimer. (a) Uncleaved precursor R239Q HA0 with arrow 2 pointing to the cleavage sites. Residues 323 of HA1 to HA2 12 are yellow. Disulfide bonds, HA1, HA2 are black, blue and red respectively. There are 6 disulfide bonds of HA monomer in neutral pH. The 12 cysteines connecting by disulfide bonds are: 64-76,97-139,281-305,52-277,473-477,14-466.³⁹ (b) Cleaved BHA with receptor binding sites marked as arrow 3. (c) Low-pH-induced conformation of thermolysin-solubilized TBHA2.⁴⁰

The HA1 subunit of the HA trimer bears the binding sites which is located 135 Å from the viral membrane, allowing the virus to attach to sialic receptors on the surface of host cell and initiate endocytosis. After entry of virus into host cell, fusion peptide of HA2 domain is exposed to endosomal membrane through conformational change and helps viral and endosomal membrane fuse together.⁶ Fusion peptide is considered as the only segment of HA inserted into host cell membrane which is proved by experiment that using the photoreactive lipid as the labeling reagent, the sole part of HA2 ectodomain that becomes labeled after fusion is fusion peptide.⁴¹ In general, cleavage occurs at the C-terminal end of a single basic residue for all HA subtypes. The fusion pH-induced, irreversible conformational change of the ectodomain of HA2 in the HA exposes the HA2 N-terminal fusion peptide.

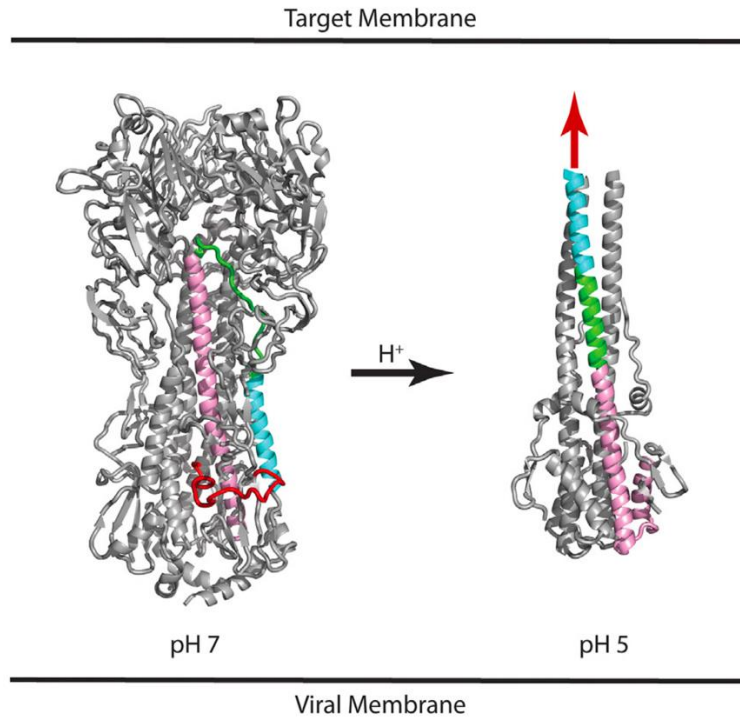


Figure 1.17. Structure of non-cleaved HA at neutral and fusion active pH. Fusion peptide (HA2 residues 1–21), red; N-helix (HA2 residues 37–57), cyan; Stem loop (HA2 residues 58–74), green; C-helix (HA2 residues 75–126), pink.⁴²

1.2.4 Conformational change of fusion peptide

The infection of influenza virus is a two-stage process that involves the entry of enveloped virus into the endosome through endocytosis followed by fusion between viral and host membranes. Viral fusion peptide is necessary and plays a vital role in fusion. According to the research of soluble trimers fragment's structure of HA, which includes soluble ectodomain and fusion peptide of HA2 and HA, it turns out that fusion peptide is ~ 100 Å from the distal tip and ~ 35 Å from the viral membrane end of the molecule.¹⁰ Neither host membrane nor viral membrane is close to the fusion peptide. To achieve fusion, either the viral or host membranes have to be moved to reachable region of fusion peptide. This is achieved by conformational change of HA2 subunit triggered by acidification of endosome after endocytosis. More specifically, when the pH becomes lower, the protonation of HA is enhanced and leads to electrostatic repulsion of trimeric globular HA1 subunits. When pH drops into fusogenic pH (pH 5-6), protonation of some amino acid residues, like histidine, glutamates, and aspartates, influence the protonation equilibrium of neighboring residues resulting in breaking of H-bonding network and the overall effect is conformational rearrangement of ectodomain of HA2. As a result, the exposure of the hydrophobic fusion peptide

facilitates the interaction between influenza virus and host cell membrane starting with insertion of fusion peptide into host membrane followed by membrane fusion.

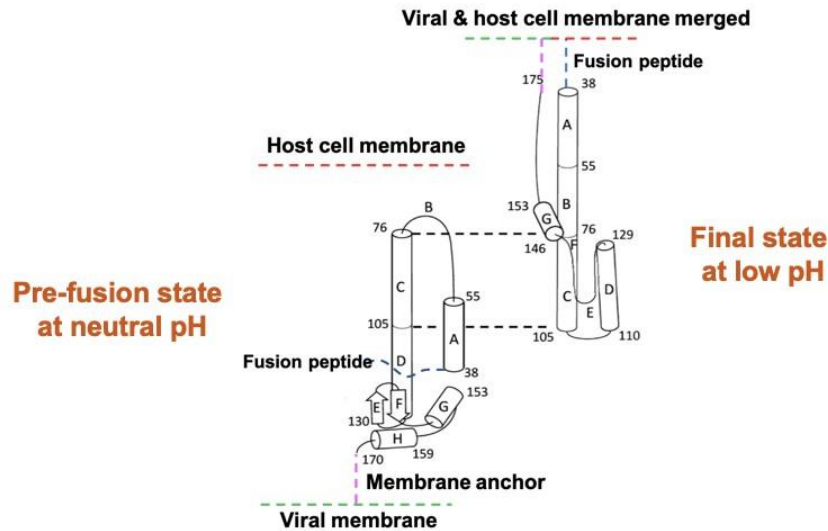


Figure 1.18. The structure of initial and final state monomer of HA2 ectodomain.⁴³

The three major rearrangements of HA2 are as follows: firstly, the short α -helix (A) and the extended loop (B) of the pre-fusion state become part of the long helix forming the coiled-coil structure of the final state. The fusion peptide is located at ~ 23 residues of the N-terminal. It is very clear that the rearrangement of the HA2 domain relocates the fusion peptide over 100 \AA from its previously buried position. Secondly, some residues in the middle part of the long α -helix region (between C and D) of native HA2 unfold to form a reverse turn to make helix D antiparallel to the long helix C. Lastly, the residues of antiparallel β sheet (E and F) at C-terminal of the initial state (G and H) is extended to a loop and becomes antiparallel with the groove between the adjacent α -helix in the center of coiled-coil structure. Besides, the α -helix H is completely extended in the final state. The overall effect of this refolding is to deliver the fusion peptide toward the target membrane and to bend the molecule in half so that the fusion peptide and the viral membrane anchor are the same end of the rod-shaped molecule.

1.2.5 HA mediated fusion mechanism

As for infection and replication of influenza virus, the bilayers of viral and host membrane have to be merged into one and then replication is initiated. During this time, the activation energy barrier mainly coming from two sources must be overcome. One is the strong repulsion when two negatively charged surfaces of bilayers come close, especially when the distance falls below 20 \AA . The other repulsion is the hydrophobic effect resisting the exposure of hydrophobic lipid tails to

the aqueous environment.⁴⁴ Fortunately, influenza fusion peptides serve as catalyst to reduce the energy barrier between two opposing phospholipid bilayers below 2 nm and disrupt the structure of the target membrane by insertion of fusion peptides, facilitating fusion and replication.

To date, four structural classes of viral fusion proteins have been identified based on studies of the three-dimensional structural change of viral glycoproteins in the initial and/or final fusion states. Class I model including influenza virus and HIV has a character that conformation of ectodomain of HA2 at final state has a signature trimer of α -helical hairpins with a central coiled-coil structure.

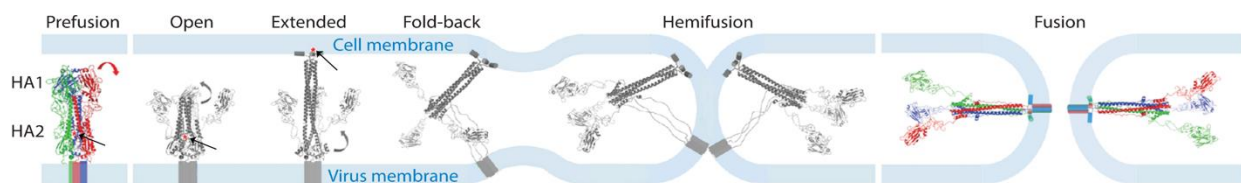


Figure 1.19. The proposed HA mediated membrane fusion pathway. Fusion peptide is represented by red asterisk (where the black arrow is pointing to). The known initial and final states are colored, and the intermediate states are shown in grey.⁴⁵

The virus particles are engaged with target cell by binding of the HA1 subunit to the sialic-acid receptor on the host cell. Previous study indicates that dissociation of HA1 from HA1 and HA2 complex which can be triggered by acidic pH, followed by a conformational rearrangement.⁴⁶ The conformational change induced by lowering pH delivers the fusion peptide located at the hydrophobic N-terminal of HA2 to target membrane from a pocket formed by C- and N-terminal ends of HA1. Afterwards, the formation of an extended coiled-coil structure drives fusion peptide inserting into the target bilayers to initiate fusion.^{47,48} The extended structure extends the 80 Å native coiled coil to form a 135 Å fusogenic structure. Consider the Fp is ~ 100 Å away from the distal tip of HA at pre-fusion state and is the only part of virus inserted into host cell membrane, the existence of extended structure can relocate the Fp and transport it to host cell.⁴⁹ The extended structure is indirectly evidence by single-molecule fluorescence resonance energy transfer (smFRET) experiment lately. The fluorophores attached at the positions 17 and 127 in the HA2 domain is reported to have a lower FRET efficiency on the transition from initial state to the coiled-coil conformation.⁵⁰

One of the commonly proposed mechanisms of HA mediated membrane fusion is lipid tail protrusion, shown in Figure 1.17. When fusion occurs, the two separated unfused membranes will be at initial close apposition, in which the distance between the membranes is less than 1nm. The outer leaflets of the bilayers then merged whereas the inner leaflets of the vial and host cell remain

separated, known as the stalk intermediate followed by hemifusion diaphragm where the outer leaflets merged while the inner leaflet touched. The hemifusion state is supported by single-particle fluorescence microscopy-based assays.^{51,52} Afterwards, the fusion pore is formed at the end of membrane fusion so that viral genome can enter to the host cell cytoplasm.

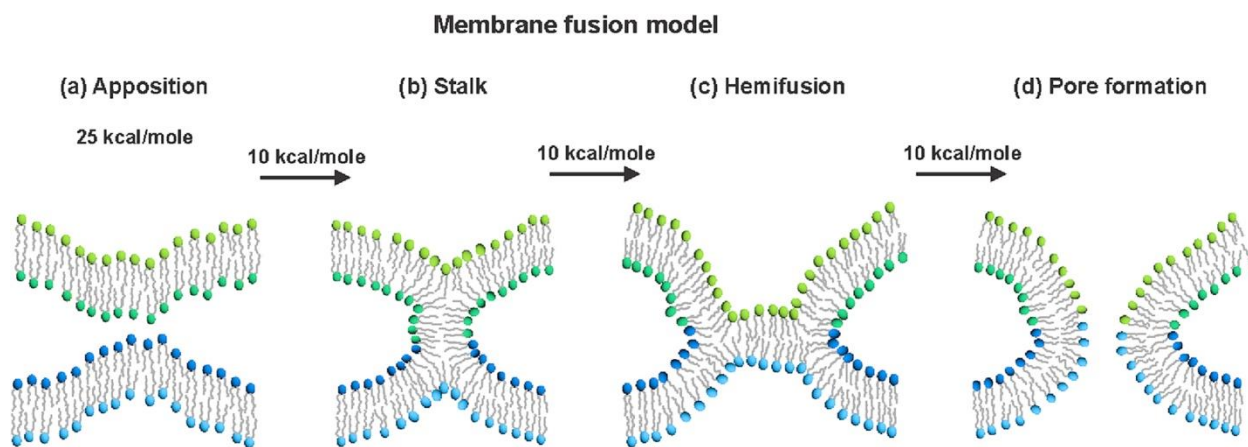


Figure 1.20. Schematic representation of intermediates during membrane fusion.

Fusion of two bilayer membranes is thermodynamically favorable (~ -20 kT) but with a high kinetic barrier.⁵³ The hydration force repulsion has been measured as the repulsive pressure with progressive water removal. It is reported that the direct repulsive pressure of egg lecithin bilayers is first detected at the separation of 27 Å and grows exponentially with a decay constant about 2.6 Å to reach 1500 atm at 3 Å separation. The repulsive force equals to a kinetic barrier that prevents lipids vesicles of 100 Å or larger to approach each other.⁵⁴ Some researchers assume that the native state of HA1/HA2 complex is at a metastable status supporting by being treated either with heat or chemical denaturant urea (4.5 M) at neutral pH resulting in the same pattern of acid-induced conformational rearrangement.⁵⁵ Epanand et al. made a controversial conclusion that HA at neutral pH is not in a metastable state based on neither the isolated HA nor the HA in the intact influenza virus exothermic process examined by differential scanning calorimetry.⁵⁶ Post-fusion state of HA2 is the most stable configuration in the absence of HA1 and the free energy released from drastically conformational change of HA2 between pre- and post-fusion is believed to be used to overcome kinetic barrier.⁵⁷ One of the universal features of post-fusion state for class-I fusion virus is the spatial proximity of the fusion peptide and the transmembrane anchor. The transmembrane domain (TMD) is assumed in proximity of fusion peptide and may be important in membrane fusion.⁵⁸ A newly reported SARS-CoV-2 spike protein structure determined by cryogenic electron microscopy evidence that the TMD wraps around the fusion peptide at the last

stage of membrane fusion.⁵⁹ Bentz proposed that the at least 8 HA trimers is needed to form the first fusion pore. Only 2 or 3 of these HA trimers is required to undergo conformational change and insert their fusion peptides into the target membrane to mediate fusion.⁶⁰ Rokonujjaman et al. examined fusion activity induced by Fp-HM, fusion peptide plus HM region (detail about HM can be found at chapter 5) by measuring the vesicle fusion of mixture consisting of different fractions of wild-type (WT) and V2E mutant. By fitting A (percent activity) = $100 \times (f_{WT})^n = 100 \times (1 - f_{V2E})^n$, the best fit n is reported to be 6 (0.39), corresponding to 2 fully WT trimers.⁶¹

1.3 HIV introduction

Human immunodeficiency virus (HIV) is a virus that attacks human body's immune system. If left not treated, HIV leads to acquired immunodeficiency syndrome (AIDS), a condition that the immune system fails to fight against life-threatening opportunistic infections. There doesn't exist an effective HIV cure. According to World Health Organization, 38.4 million people in the world were living with HIV in 2021. Two types of HIV have been characterized: HIV-1 and HIV-2. HIV-1 is more virulent and more infective than HIV-2 and causes global infections. The following introduction will be focused on HIV-1 denoted by HIV.

1.3.1 Structure of HIV

HIV is an enveloped virus with diameter of ~100-120nm. Its virion is wrapped by membrane proteins. The glycoprotein gp160 with 856 residues is the only protein existing on the surface of virion, referred as HIV enveloped protein (Env).⁶² It is a precursor of the surface glycoprotein gp120 (M.W. 120 kDa) and transmembrane glycoprotein gp41 (M.W. 41 kDa). Gp160 is cleaved by host furin-like protease to yield a complex of non-covalently associated receptor-binding subunit gp120 and fusion protein subunit gp41. The heterodimer complex forms a spike on the HIV surface consisting of a trimerized metastable gp41 transmembrane subunit with three gp120 surface subunits.^{63,64} There are approximate 10-14 spikes per virion according to electron tomography analysis.⁶⁵ There exists matrix protein forming a sphere underneath membrane. The two identical positive single-stranded RNAs are encapsulated by a conical capsid.

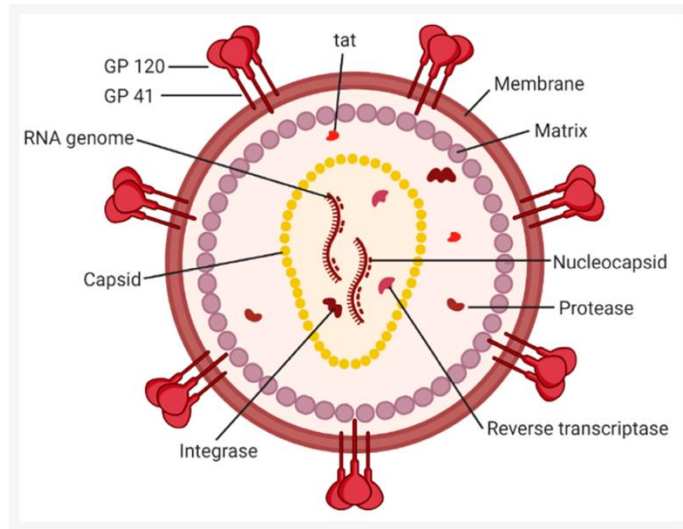


Figure 1.21. Schematic structure of HIV. Enveloped proteins gp120 and gp41 are distributed on the host-cell derived membrane surface.⁶⁶

1.3.2 Viral entry

The infection stages are shown in Figure 1.22. The entry of HIV into cells requires a sequential interaction of gp120 with receptors on the cell surface. It is initiated by gp120 binding to the primary receptor CD4 on the surface of T cells. The primary binding induces the conformational changes of gp120 resulting in the exposure and/or formation of a binding site for specific chemokine receptors, CCR5 and CXCR4, serving as co-receptors for virus entry. Only when the co-receptor has engaged to binding, the full transition proceeds.^{58,67,68} The virial-membrane fusion is mediated by fusion protein and then HIV viral core is delivered into the host cell cytoplasm by extensively conformational rearrangement. The gp120/gp41 timer is metastable and the potential energy stored in pre-triggered could be used to overcome the activation energy to from intermediate stages during fusion. Single-molecule fluorescence resonance energy transfer studies evidenced that Env can transition from the metastable status to transient, CD4- and co-receptor-stabilized configuration.⁶⁹

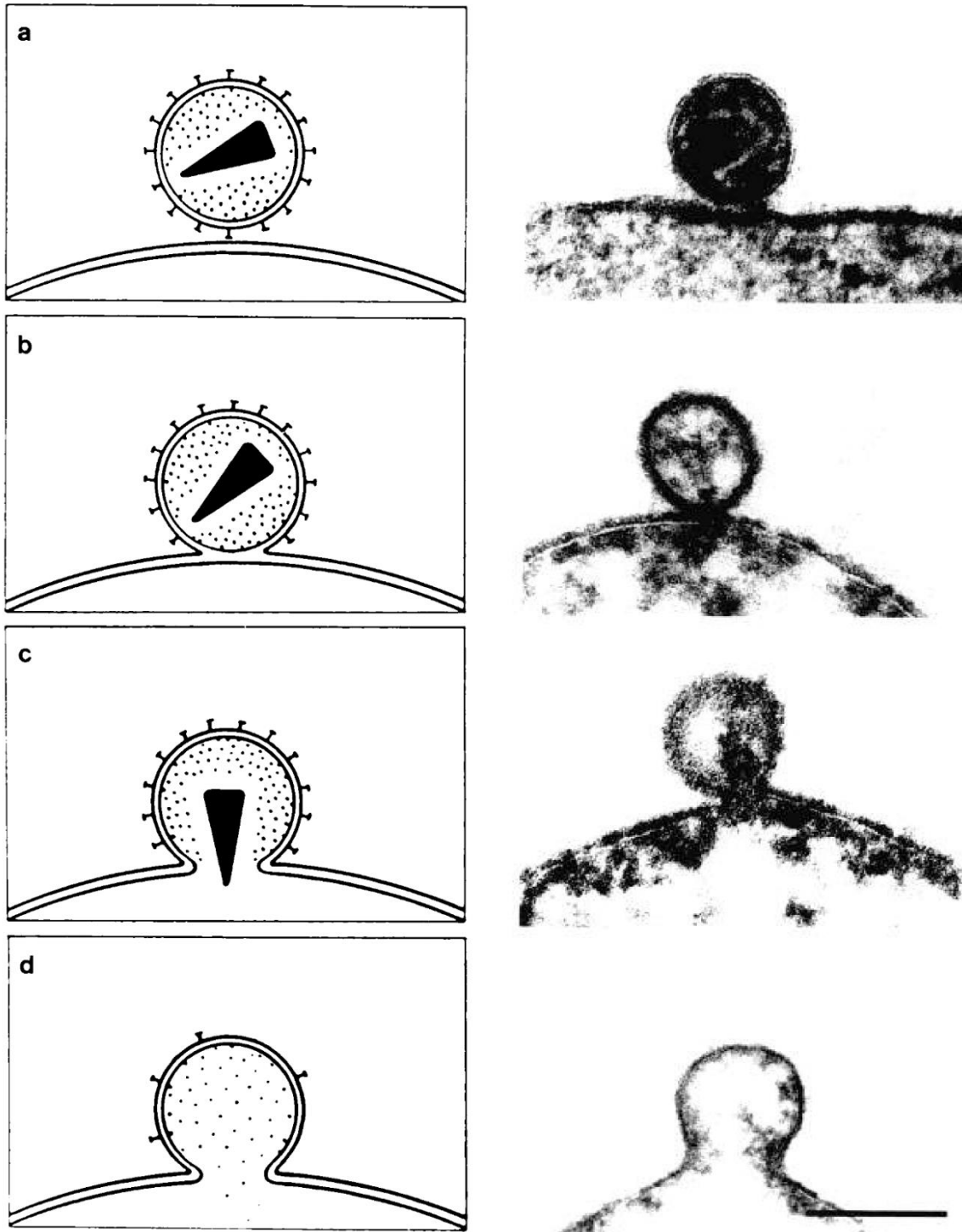


Figure 1.22. The schematic infection stages of HIV (left panel) and their corresponding electron microscopy images (right panel). (a) HIV binding to host cell, (b) the hemifusion intermediate, (c) large viral pore formation and (d) complete fusion.⁷⁰

1.3.3 Fusion mediated by gp120/gp41 complex

The gp41 ectodomain has several functional groups, fusion peptide, N-heptad repeat (NHR), C-heptad repeat (CHR), membrane proximal extracellular region (MPER), along with a loop to connect N-helix and C-helix, transmembrane domain, and endo domain.

The fusion mechanism isn't completely clear because lacking experimental evidence for intermediate states. One of the proposed mechanisms is that the HIV virus-cell fusion is mediated by gp120/gp41 complex and gp41 is subjected to a substantially conformational changes during fusion. The conformation of gp41 can be grouped into four states: (i) native, (ii) pre-hairpin intermediate, (iii) fusogenic, and (iv) post-fusion. The conformational rearrangement (Figure 1-20.) is initiated by gp120 binding with co-receptor. As following, the fusion peptide at the N-terminal of gp41 is repositioned and then insert into target membrane at pre-hairpin intermediate state.

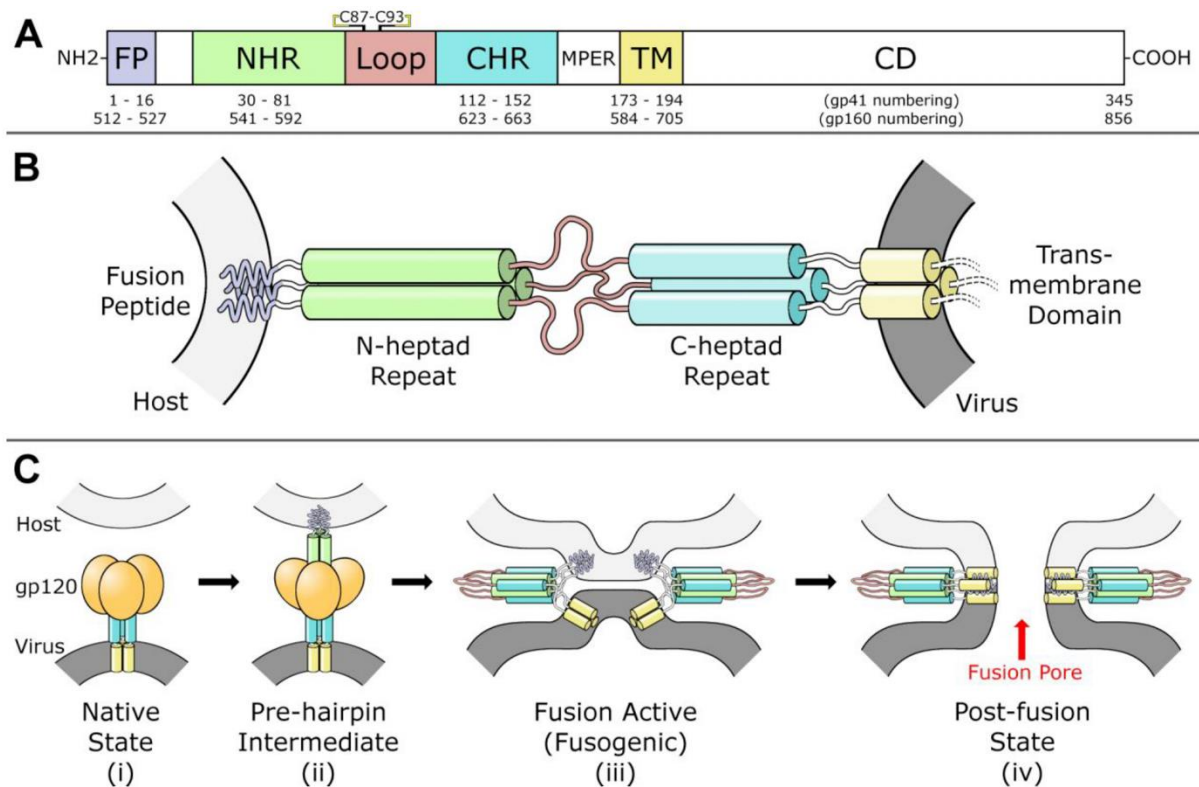


Figure 1.23. A. Diagram of gp41 sequence. B. gp41 in the pre-hairpin state. Fusion peptide inserts into the host cell and transmembrane domain anchors in the viral membrane. C: Model of gp120-mediated membrane fusion. Gp120 is omitted for clarity in fusogenic and post-fusion state.⁷¹

Conformational change of gp41 from energy-rich prefusion state to low energy post-fusion state releases free energy to overcome the kinetic barriers arising from bringing two opposing

membranes into proximity. The fusion peptide is repositioned by ~ 70 Å to interact with the host cell and the fusion-intermediate conformation bridging viral and target cell membrane is assumed to have a length of ~ 110 Å. Then the refolding of C-helix and N-helix generates the six-helix bundle core structure to drag the viral and cell membrane into close apposition for fusion.⁷² At the post-fusion state, the NHR is more compactly packed compared with the native state and the NHR/CHR six helical bundle is formed.

REFERENCES

- (1) Schmidt-Rohr, K.; Spiess, H. W. *Multidimensional Solid-State NMR and Polymers*; Academic Press, 1994. DOI: 10.1016/C2009-0-21335-3.
- (2) Harris, R. K. *Nuclear Magnetic Resonance Spectroscopy - a Physicochemical View*; Longman Scientific & Technical, 1986; Vol. 127. DOI: 10.1016/0022-2860(85)80025-9.
- (3) Interaction, Z. 6 – *NMR Interactions : Zeeman and CSA*.
- (4) Nardelli, F.; Borsacchi, S.; Calucci, L.; Carignani, E.; Martini, F.; Geppi, M. Anisotropy and NMR Spectroscopy. *Rendiconti Lincei* 2020, 31 (4), 999–1010. DOI: 10.1007/s12210-020-00945-3.
- (5) Ulrich, A. S.; Grage, S. L. Chapter 6.2 2H NMR. *Studies in Physical and Theoretical Chemistry* 1998, 84 (C), 190–211. DOI: 10.1016/S0167-6881(98)80010-4.
- (6) Polenova, T.; Gupta, R.; Goldbourt, A. Magic Angle Spinning NMR Spectroscopy: A Versatile Technique for Structural and Dynamic Analysis of Solid-Phase Systems. *Anal Chem* 2015, 87 (11), 5458–5469. DOI: 10.1021/ac504288u.
- (7) Schnell, I. Merging Concepts from Liquid-State and Solid-State NMR Spectroscopy for the Investigation of Supra- and Biomolecular Systems. *Curr Anal Chem* 2005, 1 (1), 3–27. DOI: 10.2174/1573411052948415.
- (8) Jenczyk, J. Magic Angle Spinning and Truncated Field Concept in NMR. *Concepts Magn Reson Part A Bridg Educ Res* 2019, 2019. DOI: 10.1155/2019/5895206.
- (9) Koehler, J.; Meiler, J. Expanding the Utility of NMR Restraints with Paramagnetic Compounds: Background and Practical Aspects. *Progress in Nuclear Magnetic Resonance Spectroscopy*. Elsevier B.V. 2011, pp 360–389. DOI: 10.1016/j.pnmrs.2011.05.001.
- (10) Buffy, J. J.; Hong, T.; Yamaguchi, S.; Waring, A. J.; Lehrer, R. I.; Hong, M. Solid-State NMR Investigation of the Depth of Insertion of Protegrin-1 in Lipid Bilayers Using Paramagnetic Mn²⁺. *Biophys J* 2003, 85 (4), 2363–2373. DOI: 10.1016/S0006-3495(03)74660-8.
- (11) Bellomo, G.; Ravera, E.; Calderone, V.; Botta, M.; Fragai, M.; Parigi, G.; Luchinat, C. Revisiting Paramagnetic Relaxation Enhancements in Slowly Rotating Systems: How Long Is the Long Range? *Magnetic Resonance* 2021, 2 (1), 25–31. DOI: 10.5194/mr-2-25-2021.
- (12) Liang, S. STRUCTURE AND FUNCTION STUDY OF HIV AND INFLUENZA FUSION PROTEINS, Michigan state university, 2017.
- (13) Larsson, P.; Kasson, P. M. Lipid Tail Protrusion in Simulations Predicts Fusogenic Activity of Influenza Fusion Peptide Mutants and Conformational Models. *PLoS Comput Biol* 2013, 9 (3). DOI: 10.1371/journal.pcbi.1002950.
- (14) Clore, G. Marius; Iwahara, J. Theory, Practice and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low Population States of Biological

- Macromolecules and Their Complexes. *Chem Rev* 2009, 109, 4108–4139. DOI: 10.1037/a0030561.Striving.
- (15) Reif, B.; Ashbrook, S. E.; Emsley, L.; Hong, M. Solid-State NMR Spectroscopy. *Nature Reviews Methods Primers* 2021, 1 (1). DOI: 10.1038/s43586-020-00002-1.
 - (16) Rovó, P. Recent Advances in Solid-State Relaxation Dispersion Techniques. *Solid State Nucl Magn Reson* 2020, 108 (May). DOI: 10.1016/j.ssnmr.2020.101665.
 - (17) Schanda, P.; Ernst, M. Studying Dynamics by Magic-Angle Spinning Solid-State NMR Spectroscopy: Principles and Applications to Biomolecules. *Prog Nucl Magn Reson Spectrosc* 2016, 96, 1–46. DOI: 10.1016/j.pnmrs.2016.02.001.
 - (18) Delaglio, F.; Walker, G. S.; Farley, K.; Sharma, R.; Hoch, J.; Arbogast, L.; Brinson, R.; Marino, J. P. Non-Uniform Sampling for All: More NMR Spectral Quality, Less Measurement Time. *Am Pharm Rev* 2017, 20 (4).
 - (19) Junker, J. Theoretical NMR Correlations Based Structure Discussion. *J Cheminform* 2011, 3 (7), 3–6. DOI: 10.1186/1758-2946-3-27.
 - (20) Gullion, T.; Schaefer, J. Detection of Weak Heteronuclear Dipolar Coupling by Rotational-Echo Double-Resonance Nuclear Magnetic Resonance. *Advances in Magnetic and Optical Resonance* 1989, 13 (C), 57–83. DOI: 10.1016/B978-0-12-025513-9.50009-4.
 - (21) Gullion, T. Rotational-Echo, Double-Resonance NMR. *Modern Magnetic Resonance* 2007, 1 (1), 713–718. DOI: 10.1007/1-4020-3910-7_89.
 - (22) Webster, R. G.; Bean, W. J.; Gorman, O. T.; Chambers, T. M.; Kawaoka, Y. Evolution and Ecology of Influenza A Viruses. *Microbiol Rev* 1992, 56 (1), 152–179.
 - (23) Cox, R. J.; Brokstad, K. A.; Ogra, P. Influenza Virus: Immunity and Vaccination Strategies. Comparison of the Immune Response to Inactivated and Live, Attenuated Influenza Vaccines. *Scand J Immunol* 2004, 59 (1), 1–15. DOI: 10.1111/j.0300-9475.2004.01382.x.
 - (24) Nuwarda, R. F.; Alharbi, A. A.; Kayser, V. An Overview of Influenza Viruses and Vaccines. *Vaccines (Basel)* 2021, 9 (9). DOI: 10.3390/vaccines9091032.
 - (25) Dou, D.; Revol, R.; Östbye, H.; Wang, H.; Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology*. 2018, p 1581. <https://www.frontiersin.org/article/10.3389/fimmu.2018.01581>.
 - (26) Chlanda, P.; Zimmerberg, J. Protein–Lipid Interactions Critical to Replication of the Influenza A Virus. *FEBS Lett* 2016, 590, 1940–1954. DOI: 10.1002/1873-3468.12118.
 - (27) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the Haemagglutinin Membrane Glycoprotein of Influenza Virus at 3 Å Resolution. *Nature* 1981, 289, 366–373.
 - (28) Gaymard, A.; Le Briand, N.; Frobert, E.; Lina, B.; Escuret, V. Functional Balance between Neuraminidase and Haemagglutinin in Influenza Viruses. *Clinical Microbiology and Infection* 2016, 22 (12), 975–983. DOI: 10.1016/j.cmi.2016.07.007.
 - (29) Lai, J. C. C.; Chan, W. W. L.; Kien, F.; Nicholls, J. M.; Peiris, J. S. M.; Garcia, J. M.

- Formation of Virus-like Particles from Human Cell Lines Exclusively Expressing Influenza Neuraminidase. *Journal of General Virology* 2010, 91 (9), 2322–2330. DOI: 10.1099/vir.0.019935-0.
- (30) Pinto, L. H.; Lamb, R. A. The M2 Proton Channels of Influenza A and B Viruses. *Journal of Biological Chemistry* 2006, 281 (14), 8997–9000. DOI: 10.1074/jbc.R500020200.
- (31) Hong, M.; Su, Y. Structure and Dynamics of Cationic Membrane Peptides and Proteins: Insights from Solid-State NMR. *Protein Science* 2011, 20 (4), 641–655. DOI: 10.1002/pro.600.
- (32) Hamilton, B. S.; Whittaker, G. R.; Daniel, S. Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses* 2012, 4 (7), 1144–1168. DOI: 10.3390/v4071144.
- (33) Samji, T. Influenza A: Understanding the Viral Life Cycle. *Yale Journal of Biology and Medicine* 2009, 82 (4), 153–159.
- (34) Jones, I. M.; Reay, P. A.; Philpott, K. L. Nuclear Location of All Three Influenza Polymerase Proteins and a Nuclear Signal in Polymerase PB2. *EMBO J* 1986, 5 (9), 2371–2376. DOI: 10.1002/j.1460-2075.1986.tb04506.x
- (35) Neumann, G.; Castrucci, M. R.; Kawaoka, Y. Nuclear Import and Export of Influenza Virus Nucleoprotein. *J Virol* 1997, 71 (12), 9690–9700. DOI: 10.1128/jvi.71.12.9690-9700.1997.
- (36) Eisfeld, A. J.; Neumann, G.; Kawaoka, Y. At the Centre: Influenza A Virus Ribonucleoproteins. *Nat Rev Microbiol* 2015, 13 (1), 28–41. DOI: 10.1038/nrmicro3367.
- (37) Shi, Y.; Wu, Y.; Zhang, W.; Qi, J.; Gao, G. F. Enabling the “Host Jump”: Structural Determinants of Receptor-Binding Specificity in Influenza A Viruses. *Nat Rev Microbiol* 2014, 12 (12), 822–831. DOI: 10.1038/nrmicro3362.
- (38) Skehel, J. J.; Wiley, D. C. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annual Reviews in Biochemistry* 2000, 69, 531–569. DOI: 10.1128/JVI.06147-11.
- (39) Maggioni, M. C.; Liscaljet, I. M.; Braakman, I. A Critical Step in the Folding of Influenza Virus HA Determined with a Novel Folding Assay. *Nat Struct Mol Biol* 2005, 12 (3), 258–263. DOI: 10.1038/nsmb897.
- (40) Chen, J.; Lee, K. H.; Steinhauer, D. A.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. Structure of the Hemagglutinin Precursor Cleavage Site, a Determinant of Influenza Pathogenicity and the Origin of the Labile Conformation. *Cell* 1998, 95 (3), 409–417. DOI: 10.1016/S0092-8674(00)81771-7.
- (41) Durrer, P.; Galli, C.; Hoenke, S.; Corti, C.; Glück, R.; Vorherr, T.; Brunner, J. H⁺-Induced Membrane Insertion of Influenza Virus Hemagglutinin Involves the HA2 Amino-Terminal Fusion Peptide but Not the Coiled Coil Region. *Journal of Biological Chemistry* 1996, 271 (23), 13417–13421. DOI: 10.1074/jbc.271.23.13417.
- (42) Caffrey, M.; Lavie, A. PH-Dependent Mechanisms of Influenza Infection Mediated by

- Hemagglutinin. *Front Mol Biosci* 2021, 8 (December), 1–6. DOI: 10.3389/fmolb.2021.777095.
- (43) Bullough, P. A.; Hughson, F. M.; Skehel, J. J.; Willey, D. C. Structure of Influenza HA at the PH of Membrane Fusion. *Nature* 1994, 371 (September), 37–43.
- (44) Boonstra, S.; Blijleven, J. S.; Roos, W. H.; Onck, P. R.; Van Der Giessen, E.; Van Oijen, A. M. Hemagglutinin-Mediated Membrane Fusion: A Biophysical Perspective. *Annu Rev Biophys* 2018, 47, 153–173. DOI: 10.1146/annurev-biophys-070317-033018.
- (45) Podbilewicz, B. Virus and Cell Fusion Mechanisms. *Annu Rev Cell Dev Biol* 2014, 30 (1), 111–139. DOI: 10.1146/annurev-cellbio-101512-122422.
- (46) Carr, C. M.; Chaudhry, C.; Kim, P. S. Influenza Hemagglutinin Is Spring-Loaded by a Metastable Native Conformation. *Proceedings of the National Academy of Sciences* 2002, 99 (26), 14306–14313. DOI: 10.1073/pnas.99.26.14306.
- (47) Floyd, D. L.; Ragains, J. R.; Skehel, J. J.; Harrison, S. C.; van Oijen, A. M. Single-Particle Kinetics of Influenza Virus Membrane Fusion. *Proceedings of the National Academy of Sciences* 2008, 105 (40), 15382–15387. DOI: 10.1073/pnas.0807771105.
- (48) Blijleven, J. S.; Boonstra, S.; Onck, P. R.; van der Giessen, E.; van Oijen, A. M. Mechanisms of Influenza Viral Membrane Fusion. *Seminars in Cell and Developmental Biology*. Academic Press December 1, 2016, pp 78–88. DOI: 10.1016/j.semcd.2016.07.007.
- (49) Carr, C. M.; Kim, P. S. A Spring-Loaded Mechanism for the Conformational Change of Influenza Hemagglutinin. *Cell* 1993, 73 (4), 823–832. DOI: 10.1016/0092-8674(93)90260-W.
- (50) Das, D. K.; Govindan, R.; Nikić-Spiegel, I.; Krammer, F.; Lemke, E. A.; Munro, J. B. Direct Visualization of the Conformational Dynamics of Single Influenza Hemagglutinin Trimers. *Cell* 2018, 174 (4), 926–937.e12. DOI: 10.1016/j.cell.2018.05.050.
- (51) Otterstrom, J.; Van Oijen, A. M. Visualization of Membrane Fusion, One Particle at a Time. *Biochemistry* 2013, 52 (10), 1654–1668. DOI: 10.1021/bi301573w.
- (52) Melikyan, G. B.; White, J. M.; Cohen, F. S. GPI-Anchored Influenza Hemagglutinin Induces Hemifusion to Both Red Blood Cell and Planar Bilayer Membranes. *Journal of Cell Biology* 1995, 131 (3), 679–691. DOI: 10.1083/jcb.131.3.679.
- (53) Ryham, R. J.; Klotz, T. S.; Yao, L.; Cohen, F. S. Calculating Transition Energy Barriers and Characterizing Activation States for Steps of Fusion. *Biophys J* 2016, 110 (5), 1110–1124. DOI: 10.1016/j.bpj.2016.01.013.
- (54) Parsegian, V. A.; Fuller, N.; Rand, R. P. Measured Work of Deformation and Repulsion of Lecithin Bilayers. *Proc Natl Acad Sci U S A* 1979, 76 (6), 2750–2754. DOI: 10.1073/pnas.76.6.2750.
- (55) Carr, C. M.; Chaudhry, C.; Kim, P. S. Influenza Hemagglutinin Is Spring-Loaded by a Metastable Native Conformation. *Proc Natl Acad Sci U S A* 1997, 94 (26), 14306–14313.

DOI: 10.1073/pnas.94.26.14306.

- (56) Epanand, R. F.; Epanand, R. M. Irreversible Unfolding of the Neutral PH Form of Influenza Hemagglutinin Demonstrates That It Is Not in a Metastable State. *Biochemistry* 2003, 42 (17), 5052–5057. DOI: 10.1021/bi034094b.
- (57) Harrison, S. C. Viral Membrane Fusion. *Nat Struct Mol Biol* 2008, 15 (7), 690–698. DOI: 10.1038/nsmb.1456.
- (58) Harrison, S. C. Viral Membrane Fusion. *Virology* 2015, 479–480, 498–507. DOI: 10.1016/j.virol.2015.03.043.
- (59) Shi, W.; Cai, Y.; Zhu, H.; Peng, H.; Voyer, J.; Rits-Volloch, S.; Cao, H.; Mayer, M. L.; Song, K.; Xu, C.; Lu, J.; Zhang, J.; Chen, B. Cryo-EM Structure of SARS-CoV-2 Postfusion Spike in Membrane. *Nature* 2023, 619 (7969), 403–409. DOI: 10.1038/s41586-023-06273-4.
- (60) Bentz, J. Minimal Aggregate Size and Minimal Fusion Unit for the First Fusion Pore of Influenza Hemagglutinin-Mediated Membrane Fusion. *Biophys J* 2000, 78 (1), 227–245. DOI: 10.1016/S0006-3495(00)76587-8.
- (61) Rokonujjaman, M.; Sahyouni, A.; Wolfe, R.; Jia, L.; Ghosh, U.; Weliky, D. P. A Large HIV Gp41 Construct with Trimer-of-Hairpins Structure Exhibits V2E Mutation-Dominant Attenuation of Vesicle Fusion and Helicity Very Similar to V2E Attenuation of HIV Fusion and Infection and Supports: (1) Hairpin Stabilization of Membrane Apposition. *Biophys Chem* 2023, 293 (November 2022), 106933. DOI: 10.1016/j.bpc.2022.106933.
- (62) Wang, Q.; Finzi, A.; Sodroski, J. The Conformational States of the HIV-1 Envelope Glycoproteins. *Trends Microbiol* 2020, 28 (8), 655–667. DOI: 10.1016/j.tim.2020.03.007.
- (63) Cai, L.; Gochin, M.; Liu, K. Biochemistry and Biophysics of HIV-1 Gp41 - Membrane Interactions and Implications for HIV-1 Envelope Protein Mediated Viral-Cell Fusion and Fusion Inhibitor Design. *Curr Top Med Chem* 2011, 11 (24), 2959–2984. DOI: 10.2174/156802611798808497.
- (64) McCaul, N.; Quandt, M.; Bontjer, I.; van Zadelhoff, G.; Land, A.; Crooks, E. T.; Binley, J. M.; Sanders, R. W.; Braakman, I. Intramolecular Quality Control: HIV-1 Envelope Gp160 Signal-Peptide Cleavage as a Functional Folding Checkpoint. *Cell Rep* 2021, 36 (9), 109646. DOI: 10.1016/j.celrep.2021.109646.
- (65) Zhu, P.; Chertova, E.; Bess, J.; Lifson, J. D.; Arthur, L. O.; Liu, J.; Taylor, K. A.; Roux, K. H. Electron Tomography Analysis of Envelope Glycoprotein Trimers on HIV and Simian Immunodeficiency Virus Virions. *Proc Natl Acad Sci U S A* 2003, 100 (26), 15812–15817. DOI: 10.1073/pnas.2634931100.
- (66) Rossi, E.; Meuser, M. E.; Cunanan, C. J.; Cocklin, S. Structure, Function, and Interactions of the HIV-1 Capsid Protein. *Life* 2021, 11 (2), 1–25. DOI: 10.3390/life11020100.
- (67) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structure of an HIV Gp 120 Envelope Glycoprotein in Complex with the CD4 Receptor and a Neutralizing Human Antibody. *Nature* 1998, 393 (6686), 648–659. DOI: 10.1038/31405

- (68) Engelman, A.; Cherepanov, P. The Structural Biology of HIV-1: Mechanistic and Therapeutic Insights. *Nat Rev Microbiol* 2012, *10* (4), 279–290. DOI: 10.1038/nrmicro2747.
- (69) Munro, J. B.; Gorman, J.; Ma, X.; Zhou, Z.; Arthos, J.; Burton, D. R.; Koff, W. C.; Courter, J. R.; Smith, A. B.; Kwong, P. D.; Blanchard, S. C.; Mothes, W. Conformational Dynamics of Single HIV-1 Envelope Trimers on the Surface of Native Virions. *Science (1979)* 2014, *346* (6210), 759–763. DOI: 10.1126/science.1254426.
- (70) Grewe, C.; Beck, A.; Gelderblom, H. R. HIV: Early Virus-Cell Interactions. *J Acquir Immune Defic Syndr (1988)* 1990, *3* (10), 965–974.
- (71) Allen, W. J.; Rizzo, R. C. Computer-Aided Approaches for Targeting HIVgp41. *Biology (Basel)* 2012, *1* (2), 311–338. DOI: 10.3390/biology1020311.
- (72) Caillat, C.; Guilligay, D.; Torralba, J.; Friedrich, N.; Nieva, J. L.; Trkola, A.; Chipot, C. J.; Dehez, F. L.; Weissenhorn, W. Structure of Hiv-1 Gp41 with Its Membrane Anchors Targeted by Neutralizing Antibodies. *Elife* 2021, *10*, 1–26. DOI: 10.7554/ELIFE.65005.

Chapter 2 Materials and methods

2.1 Materials

The DNA plasmids containing HM gene was ordered from GenScript (Piscataway, NJ). The protein expression cell *Escherichia coli* BL21(DE3) strain was purchased from Novagen (Gibbstown, NJ). The lipids POPC and POPG were purchased from Avanti Lipids (Alabaster, AL). The IFP were purchased from GL Biochem (Shanghai, China). 1,3-¹³C-glycerol and 2-¹³C-glycerol were ordered from Sigma-Aldrich (St. Louis, MO). Other reagents were typically purchased from Sigma-Aldrich (St. Louis, MO).

2.2 HM protein expression and purification

Synthesizing and regulating proteins in living organisms is known as protein expression. Bacterial protein expression systems are widely used for producing recombinant proteins. *E. coli* has been commonly chosen as the expression host because the fast growth rate and low cost for the culture related reagents.¹ To make *E. coli* cells express desired proteins, the DNA encoding for the target protein is replicated rapidly through polymerase chain reaction (PCR) followed by transforming into *E. coli* cells. After transformation, *E. coli* cells incorporating with the gene of target protein will express recombinant proteins at the certain circumstance.^{2,3}

E. coli cells are grown in a culture medium and the stage of growth is estimated by measuring the optical density at 600nm (OD₆₀₀). The light scattering caused by the presence of cells is used to estimate cell concentration.⁴ A typical cell growth stage is shown in Figure 2.1.

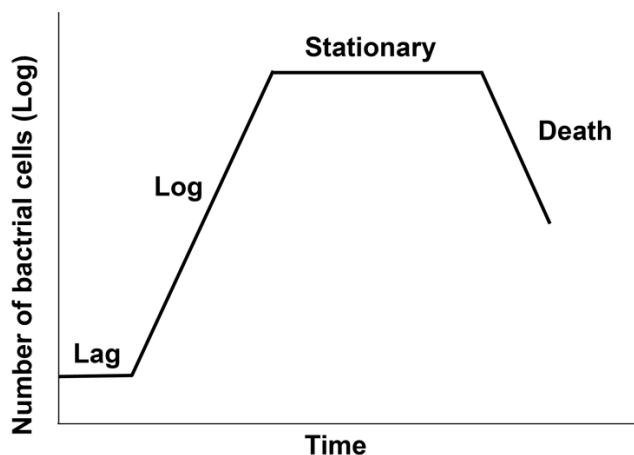


Figure 2.1. A typical cells growth curve.

After being transferred to a culture medium, cells take time to reach a physiological state being capable of rapid cell growth and division, which is lag phase. Log phase or exponential phase is

where cell growth is initiated, and cells start to rapidly divide. DNA replication, RNA transcription, and protein production are at a rapid, constant rate at log phase so the cell growth in a constant rate. Cells enter the stationary phase after reaching the maximum cell density depending on the supply of nutrient in the culture medium.⁵ Dead cells are replaced by the new cells in a very short time results in the number of cells being constant. Nutrient depletion leads the cells entering the final phase, death phase. Cells die and the number of cells decreases in death phase.⁶

In general, protein expression is initiated by adding inducer, e.g., isopropyl β -D-thiogalactopyranoside (IPTG) to the culture medium at log phase ($OD_{600} = 0.5 - 0.8$). The optimal point of OD_{600} is dependent on expressed protein, expression system and culture medium condition.

2.2.1 Expression of HM inclusion bodies proteins

HM is N- and C- helix region connecting by a non-native loop along with membrane proximal external region of gp41. Schematic diagram of full-length gp41 and HM construct and its amino acid sequence is shown in Figure 2.2.

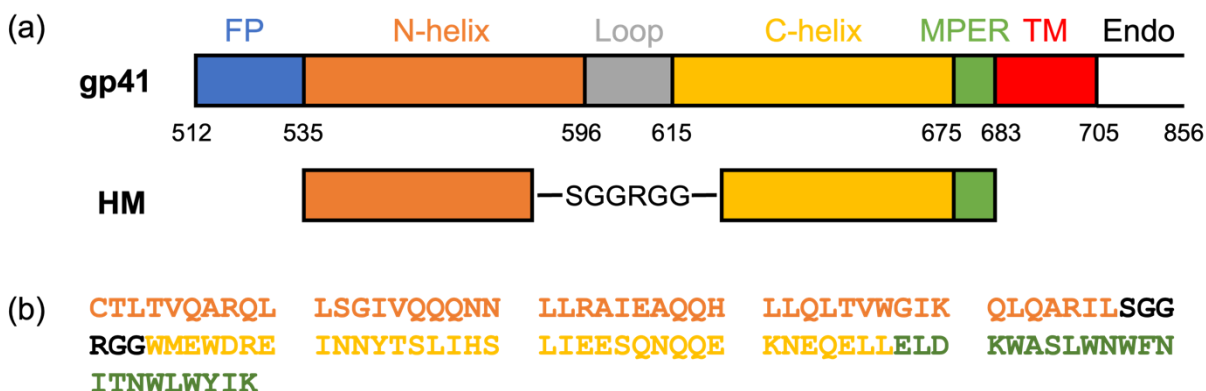


Figure 2.2. (a) Schematic diagram of full-length gp41 and HM. (b) Amino acid sequence of HM. FP, TM and Endo represents fusion peptide, transmembrane domain, and endo domain respectively. The -SGGRGG- is the replacement of the native loop which does not affect the SHB assembly.

The DNA encoded for HM was subcloned into pET-24a(+) vector containing Lac operon and kanamycin resistance. The plasmid was transferred into *E.coli* BL21 (DE3) stain and then grew in desired medium overnight. The Minimal medium was used for my project. The *E.coli* stain containing HM plasmid was preserved for future use and the stock aliquots were prepared by mixing 1mL culture and 0.5 mL 50% glycerol stored at -80°C .⁷

A typical HM IBs protein expression started with adding 1.5 mL glycerol (prepared in Minimal medium) to 500 mL Minimal medium (M9 minimal salts, 1 M MgSO_4 , 100 mM CaCl_2 , 4g/L

glucose) containing 50 mg/L kanamycin. The cell grew at 37 °C and 120 rpm for 12-14 hours. When OD₆₀₀ reached ~0.8, 2 mM IPTG was added to induce protein expression for at least 7 hours, 37 °C and the cell was harvest by centrifugation at 9000g, 4 °C for 30 min.

To get high yield of labeled HM IBs proteins, HM was firstly grown in unlabeled medium. After overnight growth, the cell was collected and then transferred to labeled medium. After cells were resuspended in the labeled medium by shaking at 150 rpm for 30min, 2 mM IPTG was added to the culture to induction for at least 7 hours.

Several different labeled media were prepared for this project. All labeled media were ¹⁵N labeled which made by following recipe:

For 500 mL minimal culture, 3g Na₂HPO₄, 1.5g KH₂PO₄, 0.25g NaCl, 0.5g ¹⁵NH₄Cl. Labeled ¹³C source, including ¹³C-glucose, 1,3-¹³C-glycerol, 2-¹³C-glycerol, was added to the labeled minimal culture before cells were transferred from unlabeled medium. For reversely labeled samples, the unlabeled amino acid (500 mg/L) corresponding to amino acid which is expected not to be labeled was added to the cell culture before protein expression. Samples prepared for chapter 6 and their associated labels are present in Table 2.1.

Table 2.1. The samples, the associated labels, and the labeled materials used for sample preparation.

Samples	Labeling	Labeled materials
U-HM	Uniformly ¹³ C and ¹⁵ N labeled	D-glucose- ¹³ C ₆ ; ¹⁵ NH ₄ Cl
Leu-Rev-HM ^a	All amino acids of HM are uniformly ¹³ C and ¹⁵ N labeled except for leucine	D-glucose- ¹³ C ₆ ; ¹⁵ NH ₄ Cl
1,3- ¹³ C-Glyc-HM	HM is labeled by ¹⁵ NH ₄ Cl and 1,3- ¹³ C-Glycerol	1,3- ¹³ C-Glycerol; ¹⁵ NH ₄ Cl
2- ¹³ C-Glyc-HM	HM is labeled by ¹⁵ NH ₄ Cl and 2- ¹³ C-Glycerol	2- ¹³ C-Glycerol; ¹⁵ NH ₄ Cl

^a Unlabeled leucine 200 mg/L was added to culture medium before protein expression.

2.2.2 HM inclusion bodies proteins purification

Purification of HM was basically sonicating wet cells harvested by centrifugation from 2.2.1 in appropriate buffered in an ice bath. All buffers should be stored at 4 °C. Each sonication round lasted for 1 min with 0.8s on followed by 0.2s off, 80 % amplitude. Wet cells (~ 5g) were firstly

subjected to three rounds of sonication in 40 mL PBS buffer (10 mM Na₃PO₄, 2 mM K₃PO₄, 137 mM NaCl, 3 mM KCl, pH 7.4) followed by centrifugation (100000g, 4 °C, 30min). The supernatant was discarded, and the pellet was saved for next step. After three times of PBS wash, the soluble protein, molecules, and suspended membrane fragments which are only effectively precipitated by > 100000g should be removed and the insoluble HM protein as IBs formation should be left as pellet. The pellet was then tip sonicated in 40 mL wash buffer (50 mM Na₃PO₄, 300 mM NaCl, 1% w/w Triton X-100, pH 8.0). The buffer with detergent dissolved membrane proteins and lipids. Then, the pellet was lyophilized and packed into ssNMR rotor for ssNMR experiments.

2.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is an electrophoresis method that can separate proteins by molecular mass. Sodium dodecyl sulfate (SDS) is a detergent having a strong protein-denaturing effect with negative charges which can bind to protein backbone at a constant molar ratio. Incubation with SDS, protein will be solubilized and denatured to linear chains with negative charges proportional to the polypeptide chain length. Negative charged proteins travelling through matrix made by polyacrylamide towards the anode under electric voltage is termed as Poly Acrylamide Gel Electrophoresis (PAGE).⁸ Polyacrylamide is linked by a cross-linking agent, bis-acrylamide (BIS) to form a three-dimensional network serving as a molecular mesh. Gel polymerization is initiated with ammonium persulfate (APS) and accelerated by addition of the catalyst *N, N, N', N'*-tetramethylethylenediamine (TEMED). The concentrations of acrylamide and BIS determine the length of the polymers and the extent of the crosslinking and most importantly, pore size. Separation gels with different acrylamide concentrations is shown in Table 2.2.⁹

Table 2.2. Gel concentrations and the separation weight range.

Gel concentration, %	Molecular weight range, kDa
<5	>200
5-12	20-200
10-15	10-100
>15	<15

When proteins move towards positive electrode, the one with smaller size will move faster than the larger size proteins due to less resistance at the time of electrophoresis. The rate of migration reflects the structure and the charge of the protein. Since the usage of SDS eliminates influence

from protein structure and make the protein charged proportional to the backbone, the migration rate will only be determined by protein length, i.e., molecular weight of protein in this case. The average molecular weight of amino acid is ~ 100 Da and the backbone length is the same for each residue of a protein. So the backbone length is approximately proportional to the molecular weight of the protein.

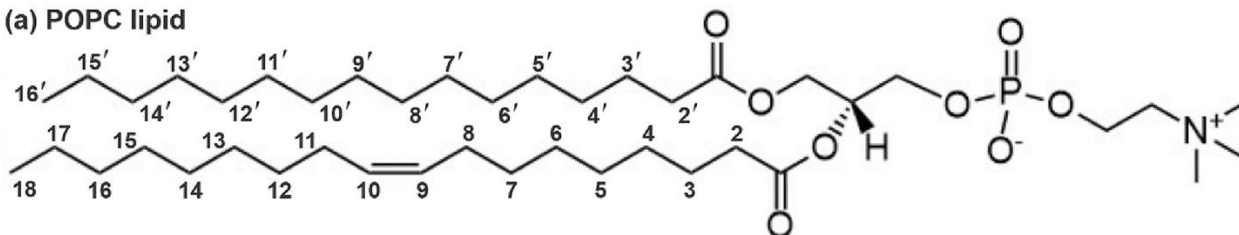
Before running the SDS-PAGE for HM proteins, HM was boiled in 200-300 μ L PBS buffer (PBS buffer + 5% SDS) for at least 30 minutes. Then loaded 3 μ L of the solution to the sample well for electrophoresis. 4–15% Mini-PROTEAN[®] TGX precast polyacrylamide gel purchased from Bio-Rad were used for this project.

2.3 Lipid sample preparation for ssNMR

2.3.1 Lipids and IFP

Figure 2.3 displays the numbering of lipids (POPC) (POPG). The IFP sequence is GLFGAIAGFIENGWEGMIDGGGKKKKKG. The underlined segment is the 20 N-terminal residues of the Ha2 subunit of the hemagglutinin protein (H3 subtype) and the C-terminal segment is non-native residues that greatly increase IFP solubility in aqueous solution so that IFP binding to membrane can be done without organic solvent or detergent additives. The IFP was prepared by Fmoc solid-phase peptide synthesis and purified by reverse-phase HPLC. GL Biochem stated that IFP purity >95% and this statement is consistent with the electrospray ionization mass spectrum that we acquired (Figure B 1).

(a) POPC lipid



(b) POPG lipid

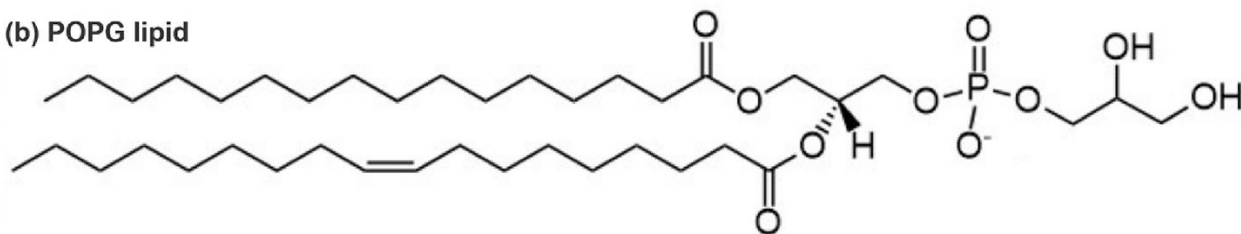


Figure 2.3. Chemical structures of POPC and POPG lipids with site numbering of the acyl chains with prime (') for the palmitoyl chain and no prime for the oleoyl chain.

2.3.2 Membrane samples

(1) Lipid (~50 μmole) was dissolved in 2 mL chloroform:methanol (9:1 v/v) and the solvent removed by nitrogen gas and then overnight vacuum. (2) The dried lipid was suspended in ~2 mL of 10 mM HEPES/5 mM MES buffer at pH 5.0 and the suspension was subjected to freeze-thaw cycles (~10 \times). (3) The suspension was subjected to ultracentrifugation at 150000 g at 4 $^{\circ}\text{C}$ for 2 h. (4) The harvested lipid pellet was lyophilized. (5) A hydrated lipid sample was prepared in a 3.2 mm outer-diameter NMR rotor by adding in sequence a ~ 10 μL aliquot of water, a portion of the lipid pellet, and then another aliquot of water. The total volume was ~40 μL . The top cap was placed on the rotor followed by overnight incubation at ambient temperature for membrane hydration.

2.3.3 Membrane samples with Mn^{2+}

An aqueous solution was prepared with $[\text{MnCl}_2]$ either ~40 mM or ~4 mM and an aliquot added to the lipid suspension after step 2. The suspension was then subjected to freeze-thaw cycles (~5 \times) to promote homogeneous distribution of Mn^{2+} in the sample.¹⁰ After step 3 ultracentrifugation, $[\text{Mn}^{2+}]_{\text{free}}$ was detected in the supernatant using an Agilent/Varian AA240 atomic absorption spectrometer with air-acetylene flame and 279.5 nm wavelength. Instrument calibration was done with MnCl_2 standard solutions in HEPES/MES buffer at pH 5. The Mn^{2+} not in the supernatant was considered bound to the membrane and $\% \text{Mn}^{2+} = (\text{mole bound } \text{Mn}^{2+})/(\text{mole lipid}) \times 100$.

2.3.4 Membrane samples with Fp

A solution was prepared with $[\text{Fp}] \approx 1$ mM in HEPES/MES buffer at pH 5.0. The solution was added dropwise to the lipid suspension after step 2 so that the Fp:lipid $\approx 1:30$ mole:mole ratio. The Fp/lipid suspension was subjected to freeze/thaw cycles (~5 \times) and then gently agitated overnight. If Mn^{2+} was to be included in the sample, the MnCl_2 solution was added before the freeze/thaw cycles. After step 3 ultracentrifugation, $[\text{Fp}]_{\text{free}}$ was measured in the supernatant using A280 and if MnCl_2 had been added, $[\text{Mn}^{2+}]_{\text{free}}$ was measured in the supernatant using flame atomic absorption spectroscopy.

2.4 Solid-state NMR experiments

2.4.1 CP/MAS- T_2 experiment

To investigate the effect of Fp on lipid protrusion, T_2 relaxation rates of lipids need to be determined and CP/MAS- T_2 sequence have been used for this purpose. Spectra were acquired on an NMR spectrometer with 9.4 T magnet, Bruker Neo console, and Bruker Efree magic angle

spinning (MAS) probe designed for lower dielectric heating of aqueous samples and for a rotor with 3.2 mm outer diameter. NMR data were acquired at 298 K with 8.0 kHz MAS frequency, ^{13}C transmitter at 100.0 ppm, and ^1H transmitter at 3.5 ppm. Figure 2.3 displays the pulse sequence with $^1\text{H} \rightarrow ^{13}\text{C}$ cross polarization followed by dephasing with Hahn echo, and then ^{13}C acquisition, with ^1H decoupling during dephasing and acquisition.^{11,12} CP parameters were varied to yield highest aliphatic intensity and typically included a $2.5 \mu\text{s}$ $^1\text{H} \pi/2$ pulse followed by 1.4 ms contact time with 42 kHz ^{13}C radiofrequency field and ^1H rf field with a linear ramp between 44 and 60 kHz. The Hahn echo was $\tau/2$ - $^{13}\text{C} \pi$ pulse- $\tau/2$. Data were collected for a range of τ typically between ~ 2 and ~ 40 ms and the $10.0 \mu\text{s}$ $^{13}\text{C} \pi$ pulse was rotor-synchronized with the start of dephasing. There was 50 kHz rf field of the ^1H SPINAL-64 decoupling, acquisition with 25 μs dwell time and 1400 complex points, 1 s recycle delay, and sum of 8 or 16 K scans. The phases of $^1\text{H} \pi/2$, ^{13}C CP, $^{13}\text{C} \pi$, and receiver were cycled: (y, x, x, x); (-y, x, -x, -x); (y, -x, -x, -x); (-y, -x, x, x); (y, y, y, y); (-y, y, -y, -y); (y, -y, -y, -y); (-y, -y, y, y). The 8-step phase cycle includes 180° alternation of the $^1\text{H} \pi/2$ phase and correlated quadrature alternation of the ^{13}C CP, $^{13}\text{C} \pi$, and receiver phases. Spectra were referenced using the methylene peak of adamantane at 40.5 ppm peak and the terminal lipid chain $^{13}\text{CH}_3$ peak at 13.9 ppm.

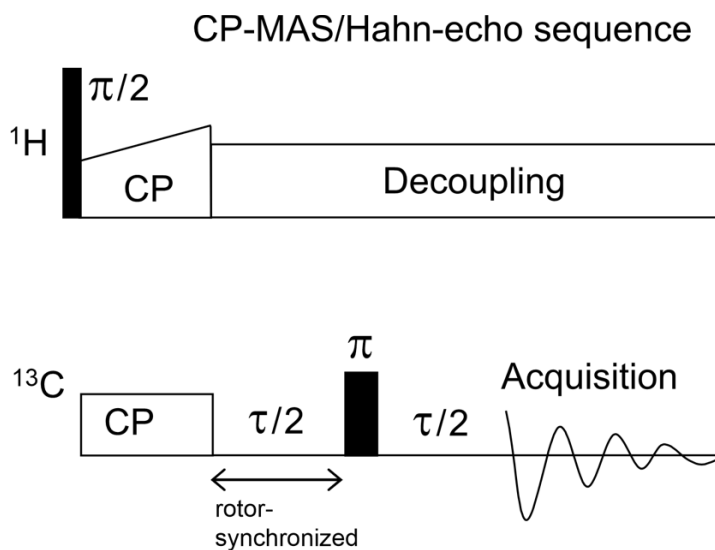


Figure 2.4. CP-MAS/Hahn-echo sequence displayed as R.F. field vs. time.

2.4.2 Multidimensional correlation ssNMR experiments

In general, there are more than one CP transfer involved in the multidimensional correlation NMR spectra. The maximum polarization transfer efficiency depends on the gyromagnetic ratio of the

related nuclei. To ensure a decent signal at the observable channel, each CP should be optimized and a tri-peptide standard, such as MLF, is necessary for optimization.

To assign multidimensional correlation spectra of a protein, $^{13}\text{C} \rightarrow ^{13}\text{C}$ 2D correlation spectra should be acquired at first. Polarization is transferred from protons to carbons and then diffuses to other nearby carbons. Compared with other 3D correlation sequences, there is only one CP involved in $^{13}\text{C} \rightarrow ^{13}\text{C}$ 2D correlation spectrum. The polarization is generated on ^1H then transferred to nearby ^{13}C s. The 3D experiments usually contain one more CP block to transfer the polarization from current ^{13}C s to their neighboring ^{13}C via spin diffusion. Therefore, $^{13}\text{C} \rightarrow ^{13}\text{C}$ 2D correlation spectra have more substantial signals than 3D measurements. $^{13}\text{C} \rightarrow ^{13}\text{C}$ correlation spectra can be used for chemical shift assignment as long as peaks are distinguished from one other. However, for most of protein samples, residues from the same amino acid type have very similar chemical shifts. The superposition of the peak causes the ^{13}C - ^{13}C correlation spectra less informative to assign residues but still be helpful for amino acid determination.

The general protocol for 2D & 3D correlation spectra acquisition is described as following: (details available in chapter 3)

1. First, the magic angle should be checked with KBr.
2. Switch to MLF or other standard, tune the probe. $^1\text{H} \rightarrow ^{13}\text{C}$ CP and $^1\text{H} \rightarrow ^{15}\text{N}$ CP should be optimized by loading hC and hN sequence. The hC and hN sequence are the CP from ^1H to ^{13}C or ^{15}N . Optimize necessary parameters to reach the strongest peak intensity. Parameters need to be optimized (useful commands for optimization: ‘dpl’: selective region will be displayed in optimization window; ‘popt’: optimization): ^{13}C and ^{15}N 90° pulse width (using zg flag); CP shape, amplitude and contact time. Typically, ^{13}C transmitter frequency offset is at 100 ppm and 120 ppm for ^{15}N . After optimization, collect a $^1\text{H} \rightarrow ^{13}\text{C}$ CP and $^1\text{H} \rightarrow ^{15}\text{N}$ CP spectrum.
3. Load hNCA sequence from BioSolid. The sequence includes ^1H - ^{15}N CP and ^{15}N - ^{13}C SPECIFIC CP. Compared with hNCACX, no ^{13}C - ^{13}C spin diffusion is preferable during optimization due to a better signal. First, the optimized parameter of ^1H - ^{15}N can be used as a start. Then, find the $^{15}\text{N} \rightarrow \text{CA}$ CP match condition. The transmitter frequency offset of ^{13}C for $^{15}\text{N} \rightarrow \text{CA}$ CP should be set at the center frequency of CA region, and it can be determined from $^1\text{H} \rightarrow ^{13}\text{C}$ CP spectrum, usually at ~ 55 ppm. The power level of C and N channel can be initially set based on $B_1, ^{13}\text{C} \pm B_1, ^{15}\text{N} = \omega_R$. If a ramped CP is selected, be careful with the percentage of the ramp power level if applicable. For example, if a 50% ramp is selected, the effective R.F. field applied to the CP is

only 50% of the ramp. Then, start a coarse array to find the ^{15}N - ^{13}C signal. Once the match condition has been found, a fine array should be carried out to optimize the match condition. Decoupling power level (typically ~ 80 - 110 kHz), contact time (~ 3 - 5 ms) should also be optimized. ^1H - ^{15}N CP can be re-optimized if needed.

4. Load hNCACX sequence. Load parameters optimized for hNCA. The only difference between those two sequences is that hNCACX has a spin diffusion part to allow the polarization transfer from CA to other side chain carbons. The longer the mixing time is, the farther the polarization can be transferred.

Dipolar assisted rotational resonance (DARR) was used for spin diffusion. The magnetization transferred from CA to other side chain carbons and the transferring distance depends on mixing time. The longer mixing time is, the further magnetization can be transferred. Usually, 30ms would be chosen for intra-residue transferring. But for residues containing aromatic carbons, a longer mixing time is recommended. 30ms, 50ms, 100ms, 250ms and 500ms are the common choices for mixing time.

5. NCOCX experiment is very similar to NCACX. The difference is the transmitter offset was set to ~ 175 ppm instead of 55ppm.

6. CONCA: parameters can be referred to NCA and NCO. There is no CON experiment as the result of very low sensitivity of ^{15}N detection. Ideally, NCO and CON should be very similar, so the parameter of NCO is informative. Compared with NCACX and NCOCX, CONCACX has the weakest signal caused by triple CP transfer.

REFERENCES

- (1) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front Microbiol* 2014, 5 (APR), 1–17. DOI: 10.3389/fmicb.2014.00172.
- (2) Makrides, S. C. Strategies for Achieving High-Level Expression of Genes in Escherichia Coli. *Microbiol Rev* 1996, 60 (3), 512–538. DOI: 10.1128/mmbr.60.3.512-538.1996.
- (3) Engineering, C. Recombinant Protein Expression in Escherichia Coli (LIDO . FALA DE PROMOTORES).Pdf. 1999, 60, 411–421.
- (4) Stevenson, K.; McVey, A. F.; Clark, I. B. N.; Swain, P. S.; Pilizota, T. General Calibration of Microbial Growth in Microplate Readers. *Sci Rep* 2016, 6 (December), 4–10. DOI: 10.1038/srep38828.
- (5) Bunch, A. W. High Cell Density Growth of Micro-Organisms. *Biotechnol Genet Eng Rev* 1994, 12 (1), 535–561. DOI: 10.1080/02648725.1994.10647921.
- (6) Pletnev, P.; Osterman, I.; Sergiev, P.; Bogdanov, A.; Dontsova, O. Survival Guide: Escherichia Coli in the Stationary Phase. *Acta Naturae* 2015, 7 (4), 22–33. DOI: 10.32607/20758251-2015-7-4-22-33.
- (7) Zheng, H.; Bai, Y.; Jiang, M.; Tokuyasu, T. A.; Huang, X.; Zhong, F.; Wu, Y.; Fu, X.; Kleckner, N.; Hwa, T.; Liu, C. General Quantitative Relations Linking Cell Growth and the Cell Cycle in Escherichia Coli. *Nat Microbiol* 2020, 5 (8), 995–1001. DOI: 10.1038/s41564-020-0717-x.
- (8) Al-Tubuly, A. A. SDS-PAGE and Western Blotting. *Methods in Melocular Medicine* 2000, 40, 391–405. DOI: 10.1385/1-59259-076-4:391.
- (9) George, A. J. T.; Urch, C. E. *Diagnostic and Therapeutic Antibodies*; 2000.
- (10) Su, Y.; Mani, R.; Hong, M. Asymmetric Insertion of Membrane Proteins in Lipid Bilayers by Solid-State NMR Paramagnetic Relaxation Enhancement: A Cell-Penetrating Peptide Example. *J Am Chem Soc* 2008, 130 (27), 8856–8864. DOI: 10.1021/ja802383t.
- (11) E.O.Stejeskal, J. S. and. Carbon-13 Nuclear Magnetic Resonance of Polymers Spinning at the Magic Angle. *J Am Chem Soc* 1976, 98 (4), 1031–1032.
- (12) Metz, G.; Ziliox, M.; Smith, S. O. Towards Quantitative CP-MAS NMR. *Solid State Nucl Magn Reson* 1996, 7 (3), 155–160. DOI: 10.1016/S0926-2040(96)01257-X.

Chapter 3 Multidimensional ssNMR correlation experiment optimization

This chapter introduces the detail about multidimensional ssNMR correlation experiment optimization with MLF (Methionine – Leucine – Phenylalanine) as standard compound. All spectra were recorded on an NMR spectrometer with a 9.4 T magnet.

1. Find magic angle

The magic angle is critical for common ssNMR sequence requiring magic angle spinning. KBr is a very suitable chemical to find the magic angle 54.7° . The reasons of choosing KBr is that the resonance frequency of ^{79}Br is very close to ^{13}C . Another advantage is that ^{79}Br is a quadrupolar nucleus with $I = 3/2$. The face-centered symmetry of ^{79}Br results in very minimal quadrupolar broadening. The isotropic peak is sharp and easily to be recognized.¹ There would be an extensive sideband manifold arising from spinning modulation of $-3/2 \leftrightarrow -1/2$ and $1/2 \leftrightarrow 3/2$ transition of ^{79}Br when the rotation axis is very near or equal to magic angle. The sidebands are broadened into noise level when the rotation axis deviates far from the magic angle. Therefore, the manifold of spinning sideband is very sensitive to magic angle resulting in KBr being a perfect candidate to calibrate magic angle.² It has a very sharp isotropic peak, and its spinning sidebands are easily to be recognized resulting from its cubic crystal symmetry. In practical, KBr powder is packed into the ssNMR rotor and spun at desired spinning rate. The magic angle is where the spectrum displays spinning sidebands as many as possible. Because the number of spinning sidebands is directly related to the spinning rate, the spinning rate should not be too high. Insert KBr into the probe, tube the probe, and find the magic angle.

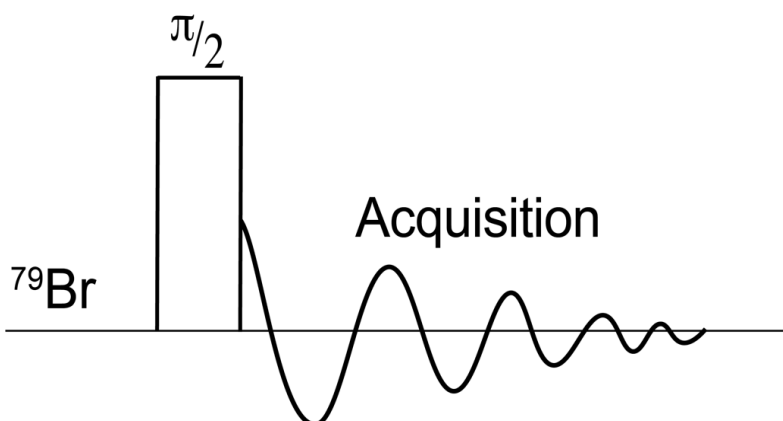


Figure 3.1. Br 90 pulse sequence used for ^{79}KBr magic angle calibration with $4.5 \mu\text{s}$ pulse width, 55.6 kHz pulse field, acquisition 30 ms .

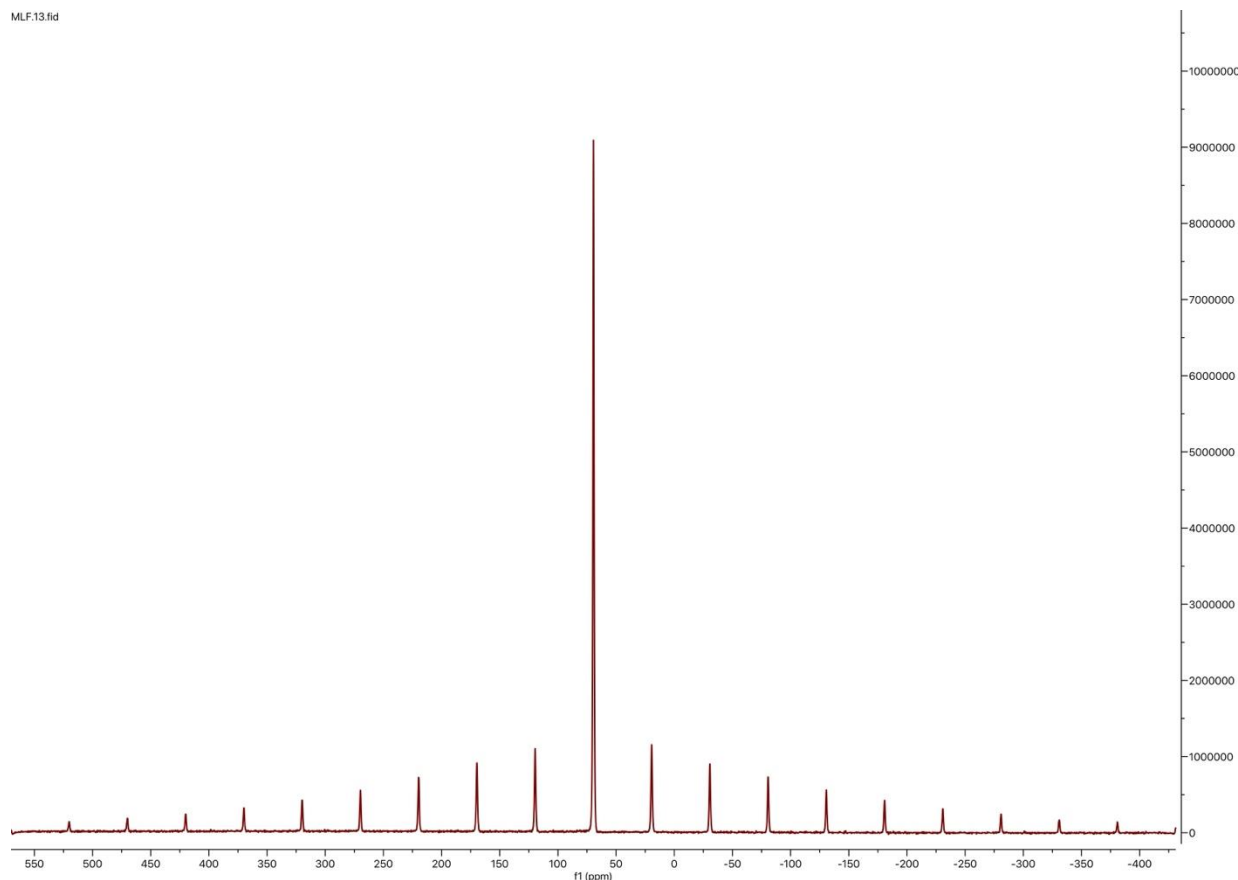


Figure 3.2. ^{79}Br spectrum with 8 scans, at spinning speed of 8 kHz, 298K. 55.6 kHz pulse field with 4.5 μs pulse width for Br $\pi/2$ pulse, No apodization used.

2. Referencing the spectrum

For solution NMR, referencing compound can be added to the sample as internal reference. But for ssNMR, spectra are generally referenced with an external referencing compound. In this chapter, two referencing compounds for ^{13}C chemical shift will be introduced. The referencing of ^{15}N can be achieved by calculating the gyromagnetic ratio between ^{13}C and ^{15}N of interest.

(a) Referencing by adamantane. Adamantane, $(\text{CH})_4(\text{CH}_2)_6$, is a commonly used reagent to reference the spectrum in ssNMR (structure shown in Figure 3.2). The ^{13}C peaks of adamantane are very sharp under MAS. The high symmetry helps the transition experience very limited CSA broadening. Adamantane is the most stable isomer of $\text{C}_{10}\text{H}_{16}$, and it is solid which can be easily packed into ssNMR rotors. Adamantane crystallizes a face-centered cubic structure and rapidly changes orientation at ambient temperature producing a narrow linewidth.³ Overall, adamantane

is a very suitable sample for accurate referencing.⁴ The methylene peak is referenced to 40.5 ppm to directly compare with the solution NMR data.⁵

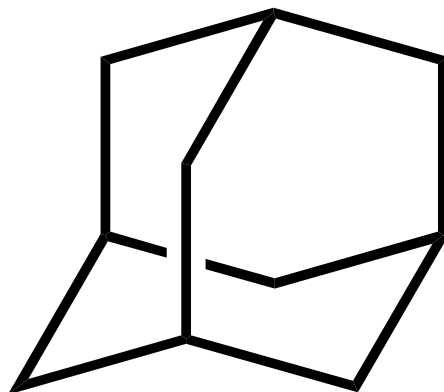


Figure 3.3. Structure of adamantane (<https://en.wikipedia.org/wiki/Adamantane>).

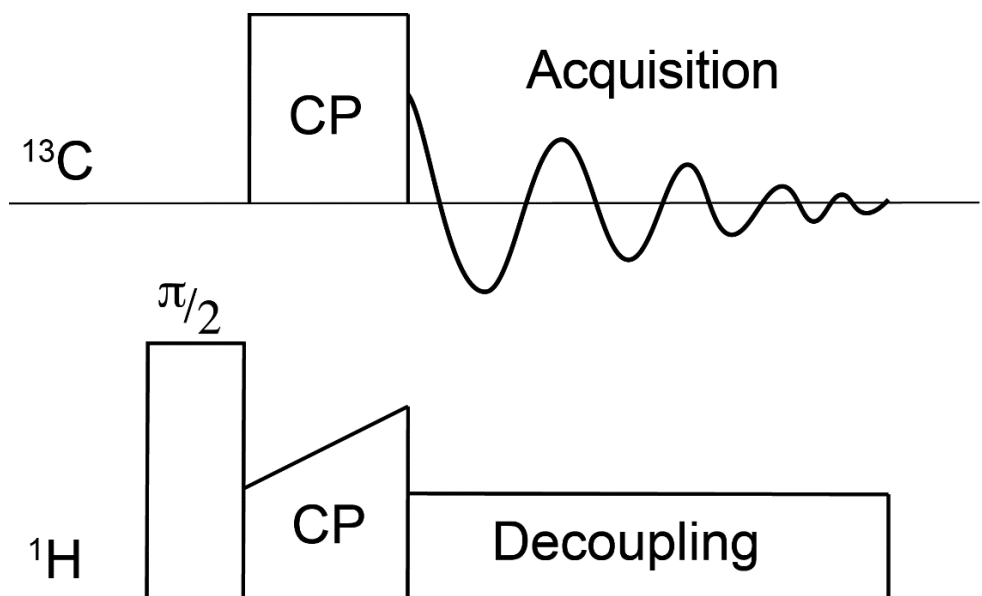


Figure 3.4. ¹H-¹³C CP pulse sequence used for adamantane. Parameters are: 2.5 μ s ¹H $\pi/2$ pulse, 1.8 ms CP contact time, CP linear ramp between 37 and 75 kHz for ¹H, decoupling 50 kHz, ¹³C CP rf field is 47 kHz, acquisition time 40 ms.

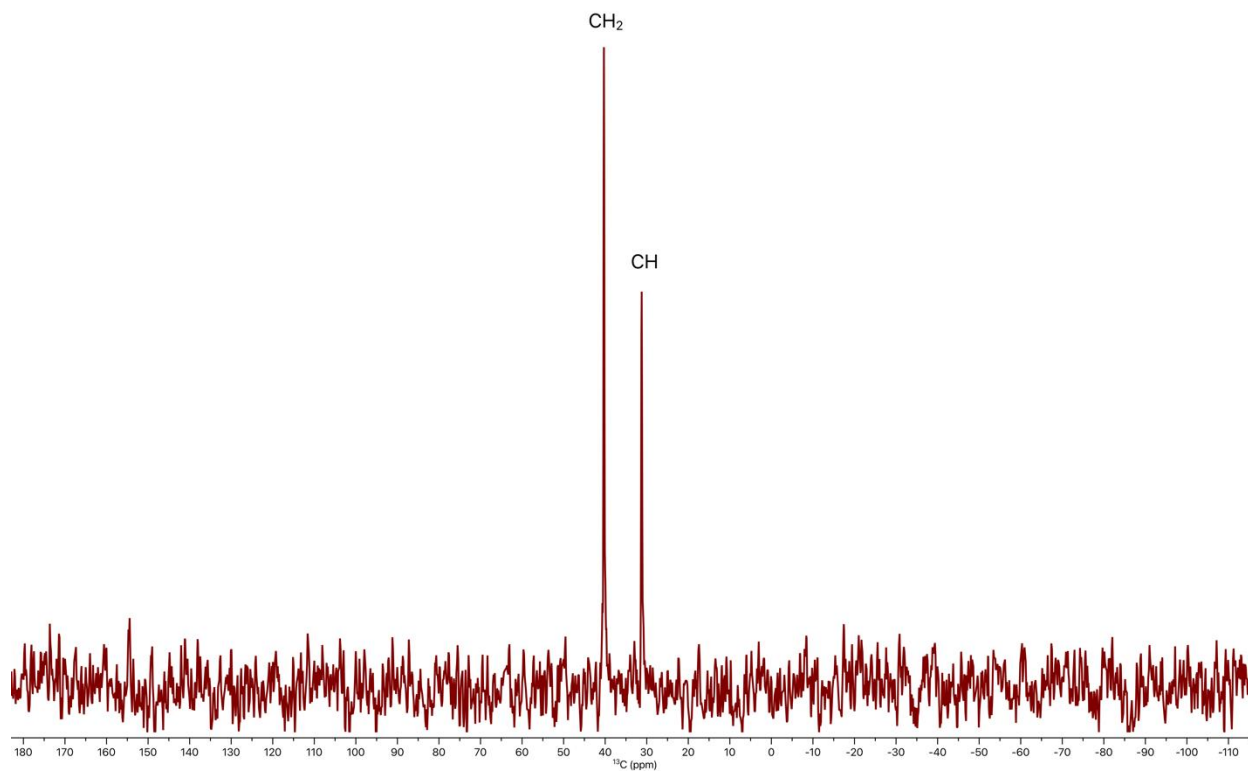


Figure 3.5. ^{13}C spectrum of adamantane with 32 scans under 8 kHz MAS, 298 K. The CH_2 peak is reference to 40.5 ppm. The spectrum is processed with 20 Hz exponential broadening.

(b) Referencing by ^{13}C -alanine. Other than adamantane, ^{13}C labeled alanine is another referencing compound for protein samples. The ^{13}C -alanine was used for reference in this case. The CO peak should be 177.905 ppm relative to tetramethylsilane (TMS) and 179.7 ppm relative to sodium trimethylsilylpropanesulfonate (DSS).

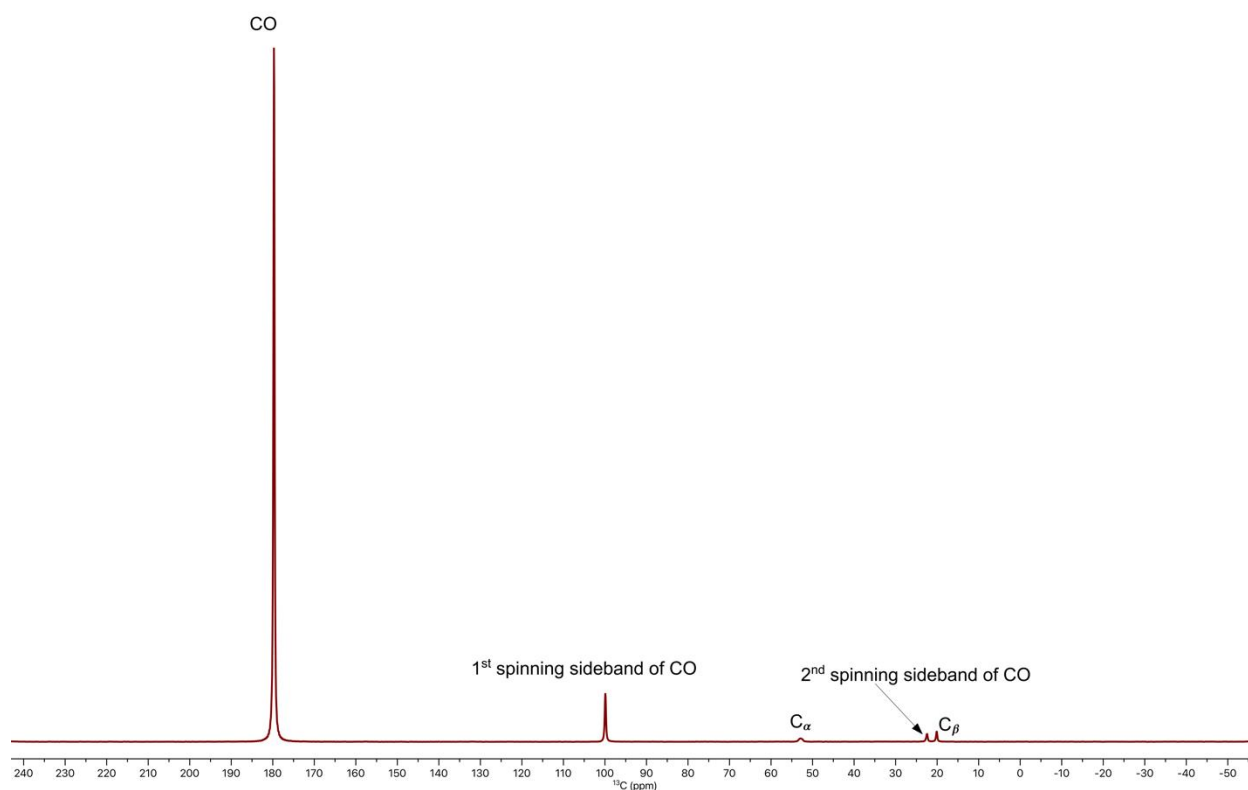


Figure 3.6. ^{13}C spectrum of ^{13}CO -Alanine processed with 20 Hz exponential broadening at 298 K, 8 kHz spinning rate, 128 scans. The o1p was set to 175 ppm. ^1H - ^{13}C CP Pulse sequence and typical parameters can be found at Figure 3.4.

3. Once the referencing is finished, switch to MLF or another standard, re-tune the probe. ^{15}N and ^{13}C pulse width should be calibrated. When calibrate the pulse width, -DC90 or -DN90 should be added in zg option of hC or hN sequence. The calibration pulse sequence is similar to typical ^1H - ^{13}C sequence with an additional 90° pulse after CP block to switch the ^{13}C or ^{15}N magnetization at xy-plane back to z axis. The 90° pulse is the point where no signal can be detected. A typical calibration pulse sequence is shown in Figure 3.7 .

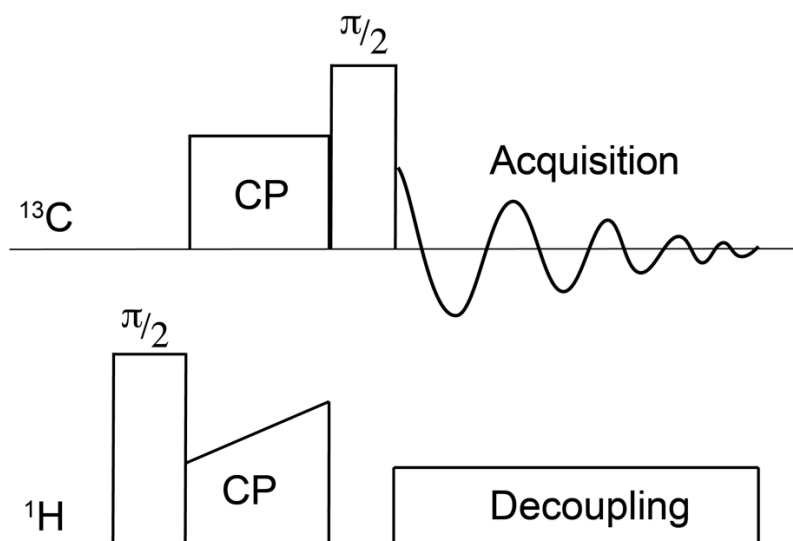


Figure 3.7. A typical pulse sequence for ^{13}C pulse width calibration. $2.5 \mu\text{s}$ ^1H 90° pulse, $4.2 \mu\text{s}$ ^{13}C 90° pulse, 1.8 ms CP contact time, CP linear ramp between 37 and 75 kHz for ^1H , decoupling 50 kHz , ^{13}C CP R.F field is 47 kHz for CP and 59.5 kHz for 90° pulse, acquisition time 40 ms . For ^{15}N pulse calibration, the R.F pulse is applied on ^{15}N channel, and the parameters are: ^{15}N $\pi/2$ pulse with 29 kHz pulse field, pulse width $6 \mu\text{s}$.

Then, $^1\text{H} \rightarrow ^{13}\text{C}$ CP and $^1\text{H} \rightarrow ^{15}\text{N}$ CP can be optimized. Load hC or hN sequence and find the best match condition. Parameters need to be optimized: ^{13}C and ^{15}N 90° pulse width (using zg flag) and pulse widths should be determined at first; CP shape, amplitude, contact time and other necessary parameters. Typically, the transmitter frequency offset (o1p) is 100 ppm for ^{13}C and 120 ppm for ^{15}N . After optimization, collect a $^1\text{H} \rightarrow ^{13}\text{C}$ CP and $^1\text{H} \rightarrow ^{15}\text{N}$ CP spectrum. The spectra can be used for transfer efficiency calculation. For example, the ratio of either the peak intensity or integral of the peak between $^1\text{H} \rightarrow ^{13}\text{C}$ CP spectrum and hNCA spectrum can be used to evaluate the transfer efficiency of $^{15}\text{N} \rightarrow ^{13}\text{C}$. The transfer efficiency of $^{15}\text{N} \rightarrow ^{13}\text{C}$ of MLF is $\sim 20\%$ (Useful commands for optimization: ‘dpl’: selective region will be displayed in optimization window; ‘popt’: optimization command).

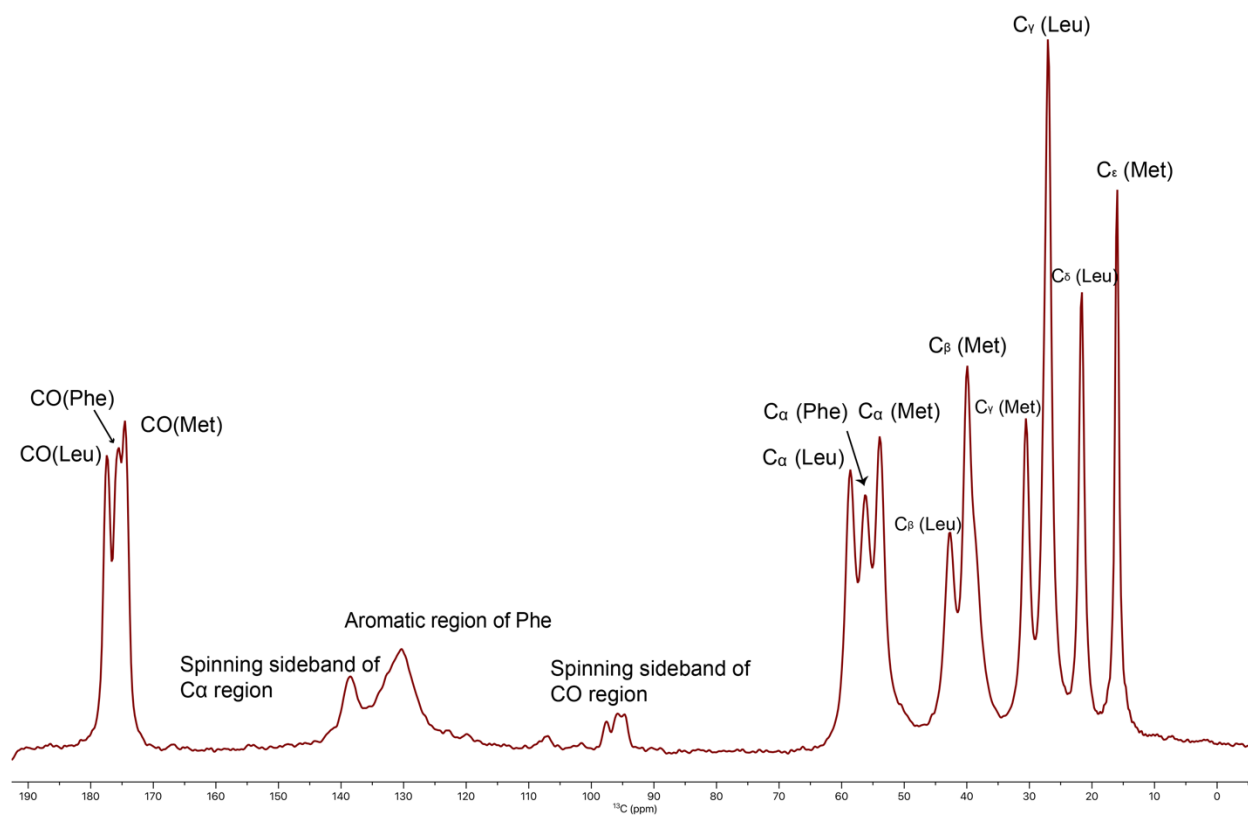


Figure 3.8. ^{13}C spectrum of MLF at 278 K, 8 kHz spinning rate, 16 scans. The o1p was at 100 ppm and the spectrum was processed with 40 Hz exponential broadening.

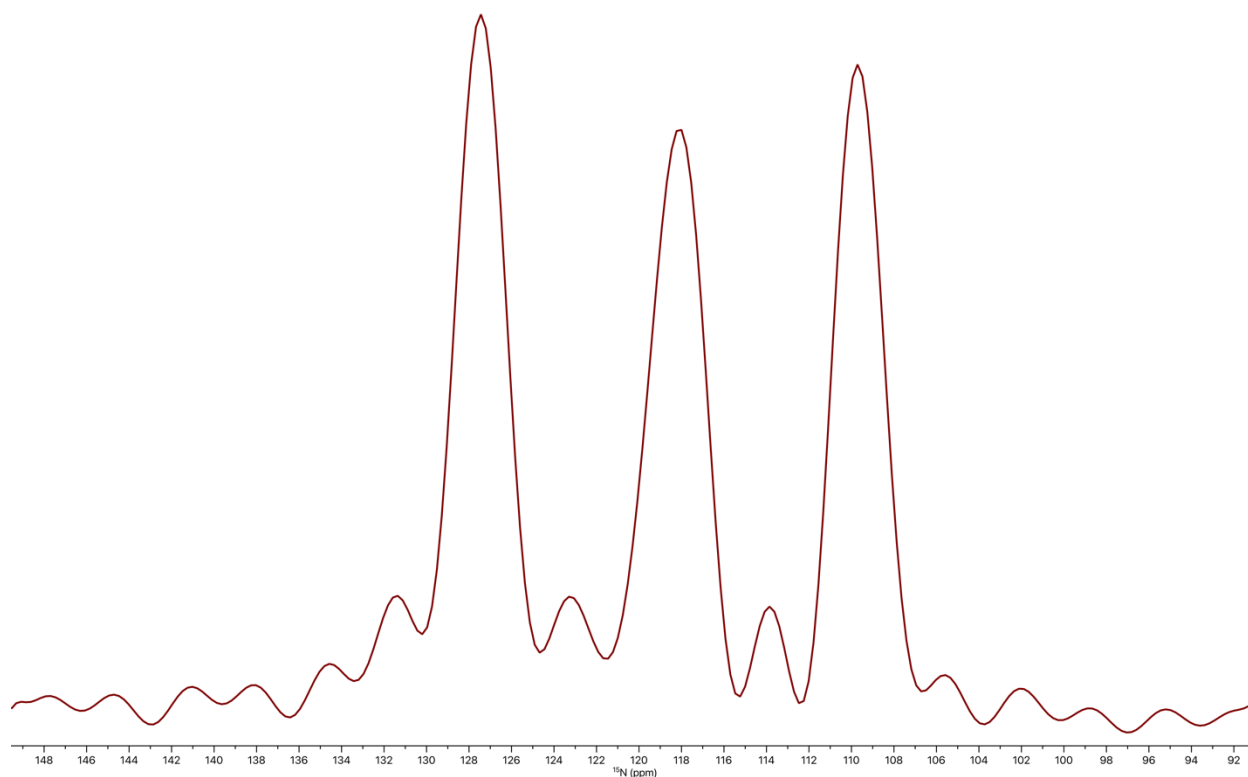


Figure 3.9. ^{15}N spectrum of MLF acquired with 16 scans under 8 kHz MAS at 278K. The o1p was at 130 ppm and the spectrum was processed with 30 Hz exponential broadening. The peak is corresponding to Met, Leu, and Phe from left to right.

4. Collect ^{13}C - ^{13}C 2D correlation spectra. The parameters are similar to $^1\text{H}\rightarrow^{13}\text{C}$ CP and the only difference is there existing spin diffusion in ^{13}C - ^{13}C experiment. Optimize parameter if needed, especially CP contact time and power level. Multiple mixing times can be used, depending on the structure of the sample. The commonly mixing times are: 30ms, 50ms, 100ms, 150ms, 300ms, 500ms, etc. 30ms and 50ms is aiming for intra-residual signal transfer and 300ms for inter-residual signal transfer. For example, if the aromatic region cross peak is expected to be visible, the longer mixing time with more scans is required. Dipolar assisted rotational resonance (DARR) was used for spin diffusion in the presented multidimensional ssNMR experiment. A small field of ^1H is applied in DARR to recouple the $^1\text{H}\rightarrow^{13}\text{C}$ dipolar coupling. A typical field is equal to spinning frequency and the irradiation of ^1H during mixing period suppresses the averaging of $^1\text{H} - ^1\text{H}$ and $^1\text{H} - ^{13}\text{C}$ coupling by MAS and enhances the magnetization transfer between ^{13}C spins.⁶

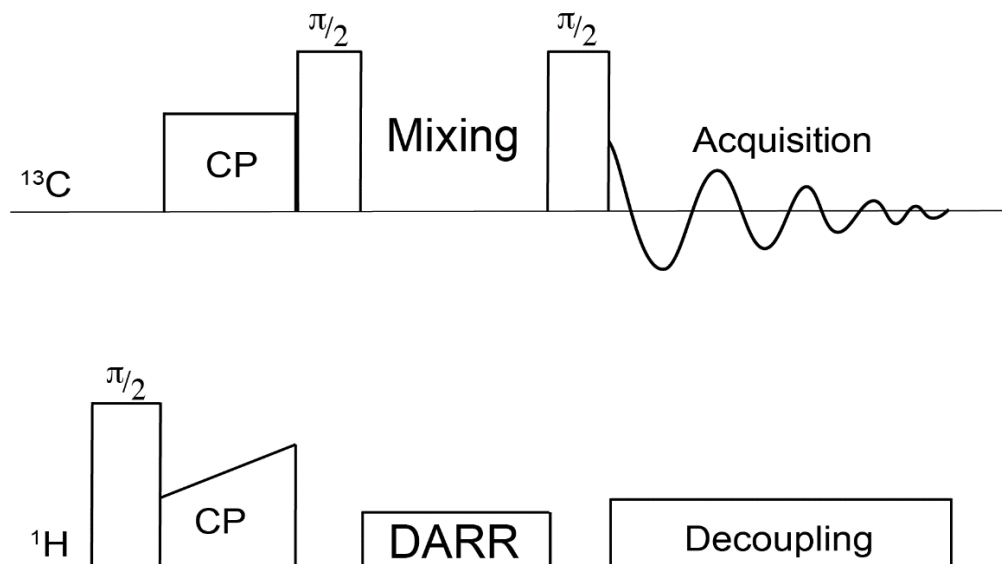


Figure 3.10. Pulse sequence for ^{13}C - ^{13}C homonuclear correlation experiment using DARR to facilitate spin diffusion. Typical parameters are: $2.5 \mu\text{s}$ ^1H 90° pulse, $4.2 \mu\text{s}$ ^{13}C 90° pulse, 1.5 ms CP contact time, mixing time 30 ms ; CP linear ramp between 37 and 75 kHz for ^1H , decoupling 50 kHz , ^{13}C field is 47 kHz for CP and 59.5 kHz for 90° pulse, 8 kHz for DARR; acquisition time 40 ms , recycle delay 1 s .

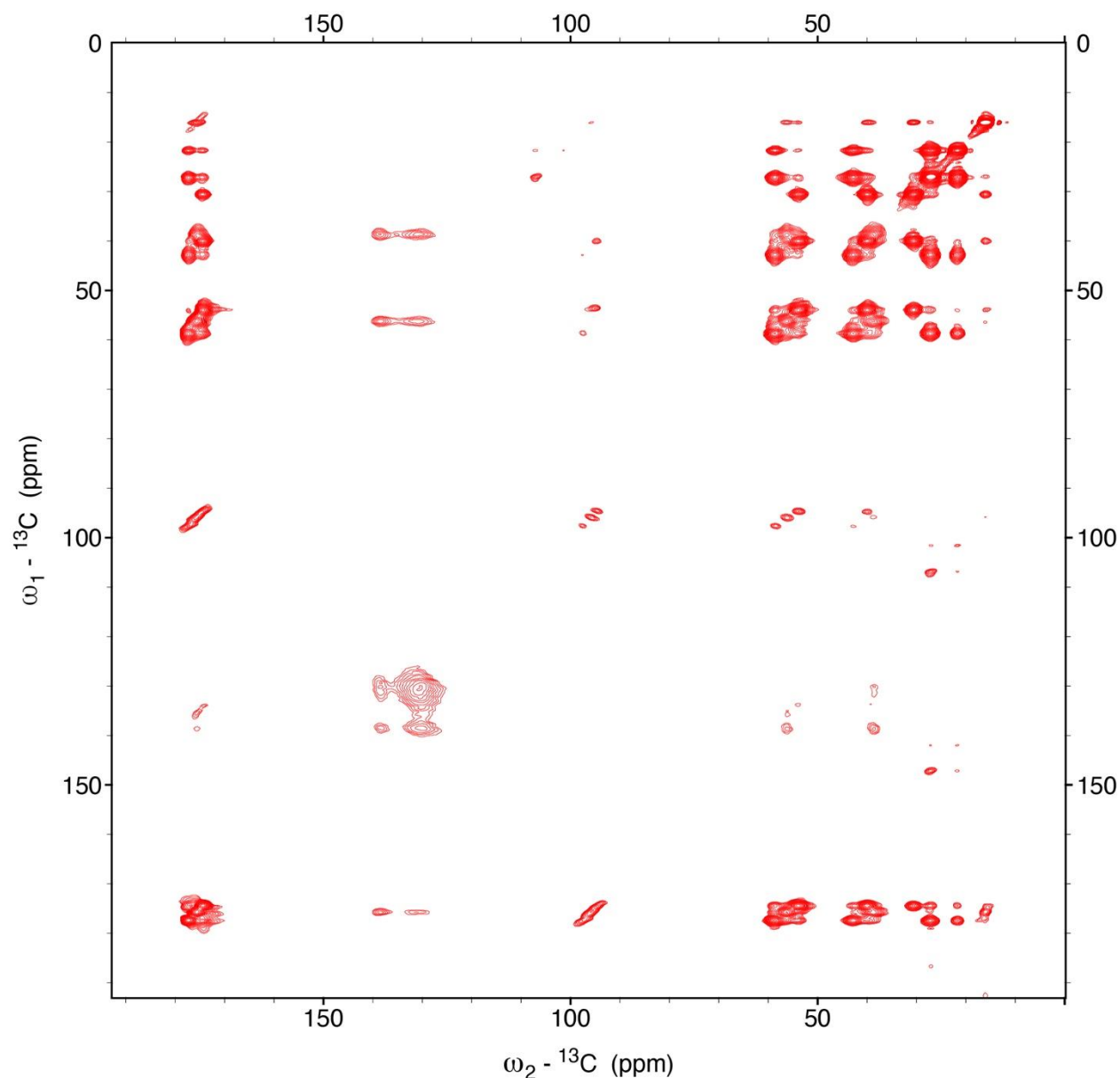


Figure 3.11. $^{13}\text{C} - ^{13}\text{C}$ homonuclear correlation spectrum of MLF using DARR with 16 scans at 278 K, mixing time 10ms. The size of FID for both dimensions are 400 and spectral width 200 ppm with transmitter frequency offset set as 100 ppm. The increment of delay for indirect dimension is 49.8625 μs . The spectrum is processed with no zero-filling and with Qsine for both dimensions, SSB = 2.

5. Load hNCA sequence from BioSolid. The sequence contains two parts: $^1\text{H} \rightarrow ^{15}\text{N}$ CP and $^{15}\text{N} \rightarrow ^{13}\text{C}$ SPECIFIC CP. Compared with hNCACX, no $^{13}\text{C} - ^{13}\text{C}$ spin diffusion is involved and preferable during optimization due to a stronger signal. First, the optimized parameter of $^1\text{H} - ^{15}\text{N}$ can be used as a start. Then, find the $^{15}\text{N} \rightarrow \text{CA}$ CP match condition. The transmitter frequency offset of ^{13}C for $\text{N} \rightarrow \text{CA}$ CP should be set at the center frequency of CA region, and it can be

determined from $^1\text{H} \rightarrow ^{13}\text{C}$ CP spectrum, usually at ~ 55 ppm. The power level of C and N channel can be initially set based on $B_{1,^{13}\text{C}} \pm B_{1,^{15}\text{N}} = \omega_R$. If a ramped CP is selected, the percentage of CP amplitude should be considered for power level calculation. Then, start a coarse array of $^{15}\text{N} \rightarrow ^{13}\text{C}$ CP power to find the $^{15}\text{N} \rightarrow ^{13}\text{C}$ signal. Once the match condition has been found, a fine array should be carried out to optimize the match condition. Decoupling power level (typically ~ 80 - 110 kHz), contact time (~ 3 - 5 ms) should also be optimized. $^1\text{H} \rightarrow ^{15}\text{N}$ CP can be re-optimized if needed.

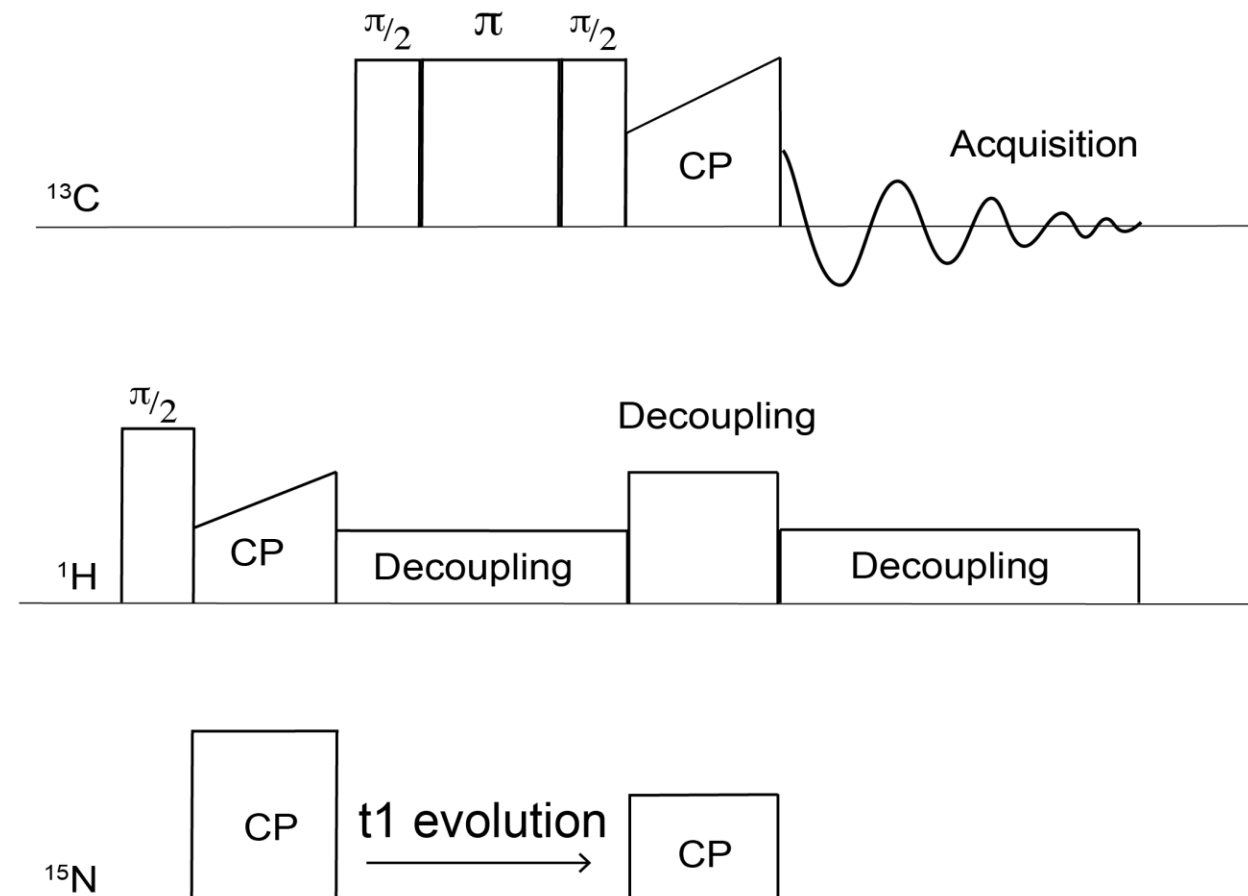


Figure 3.12. Pulse sequence for NCA. Typical parameters are: $2.25 \mu\text{s}$ ^1H 90° pulse, $4.2 \mu\text{s}$ ^{13}C 90° pulse, $8.4 \mu\text{s}$ ^{13}C 180° pulse, 1.4ms $^1\text{H} \rightarrow ^{15}\text{N}$ CP contact time, 4ms $^{15}\text{N} \rightarrow ^{13}\text{CA}$ CP contact time; $^1\text{H} \rightarrow ^{15}\text{N}$ CP: $37 - 75 \text{kHz}$ for ^1H with ^{15}N field as of 29kHz ; $^{15}\text{N} \rightarrow ^{13}\text{CA}$ CP: 11.47 - 12.68kHz for CA with 23.15kHz for ^{15}N , decoupling 75kHz for $^{15}\text{N} \rightarrow ^{13}\text{CA}$ CP and 66kHz for others; acquisition time 40ms , recycle delay 1s .

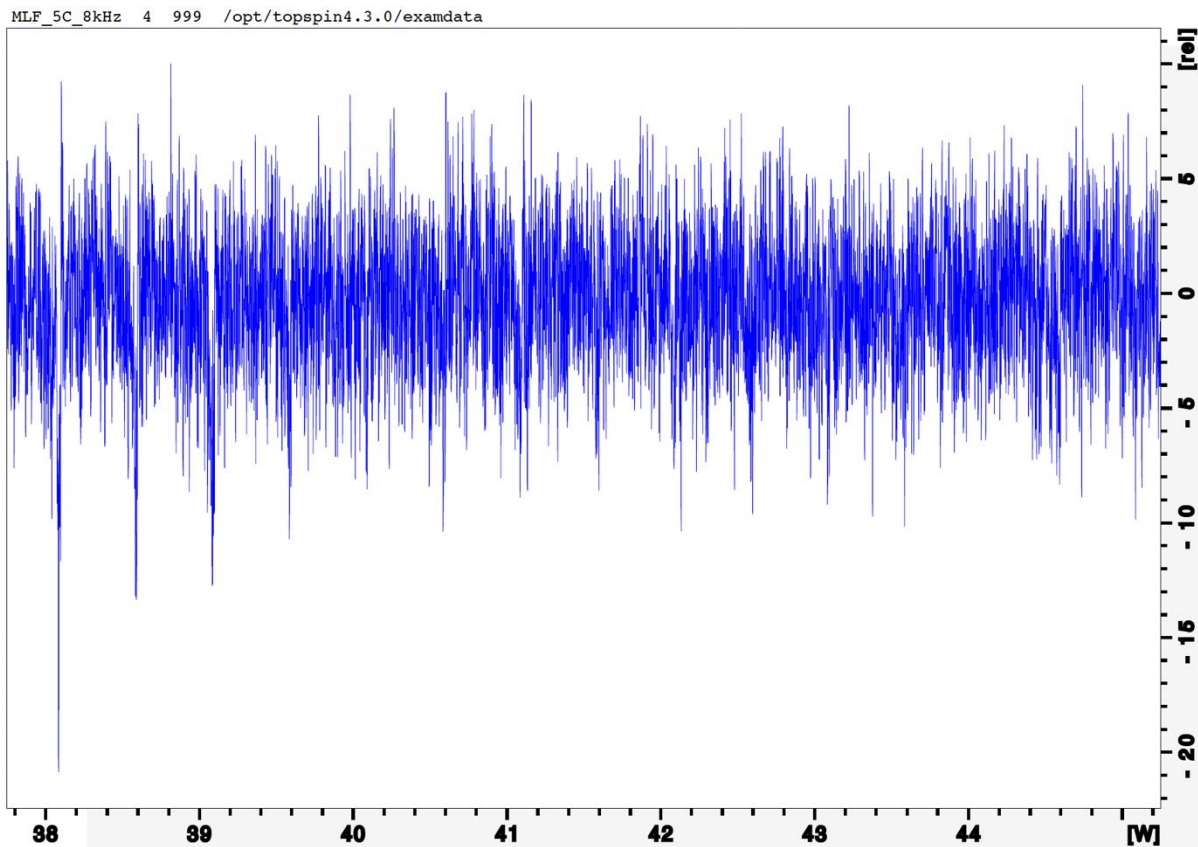


Figure 3.13. Optimization of N-CA CP power level of MLF. The dpl window was set at CA region. The array aimed to find the best power level of ^{15}N for NCA CP with the array range of 38-45 watts. The NCA CP signal presents when ^{15}N power is less or equal to 39 watts. Another fine array was performed to find the best CP power level of ^{15}N . The negative peak appears at the 38-40 watts region are unphased ^{15}N peak, and the one which shows the highest absolute intensity is the strongest FID.

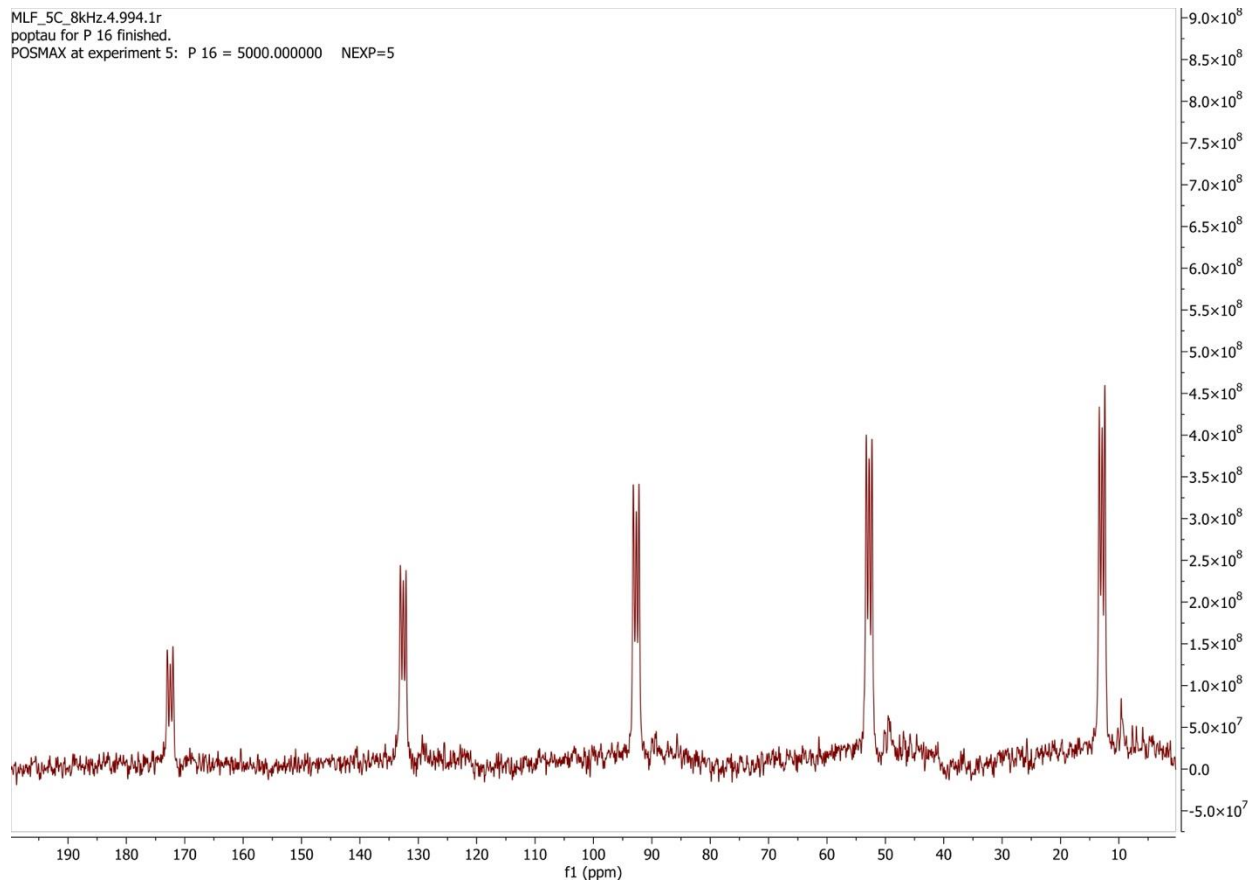


Figure 3.14. Optimization of contact time from 1ms to 5ms with increment 1ms.

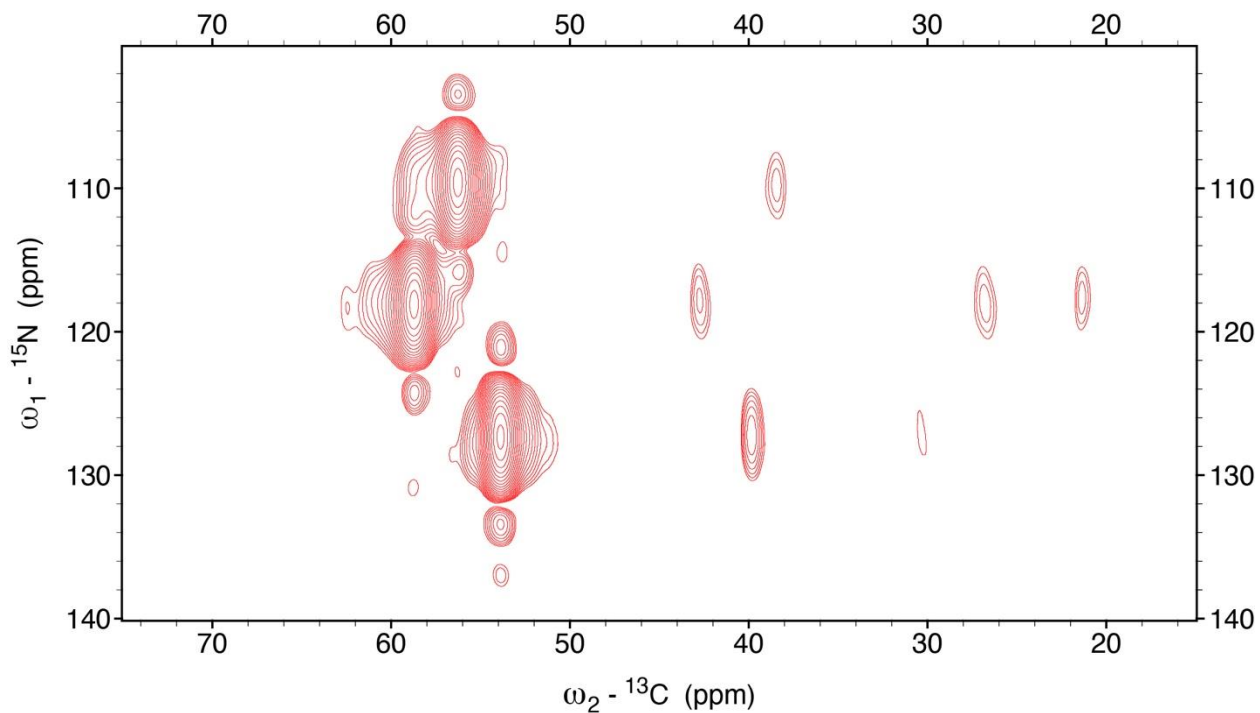


Figure 3.15. 2D NCA spectrum of MLF, contact time 4 ms. Other acquisition parameters can be found in Figure 3.12. The size of FID for N is 24 and 480 for CA. The increment for delay of ${}^{15}\text{N}$ dimension is 618.7375 μs . Spectral width is 200 ppm for CA and 40 ppm for N. The transmitter frequency offset was set 100 ppm for CA and 130 ppm for N. The spectrum was processed without zero-filling and with 25 Hz exponential broadening for CA dimension and SINE, SSB = 0 for ${}^{15}\text{N}$ dimension.

6. After optimizing hNCA, load hNCACX sequence. Load parameters optimized for hNCA. The only difference between those two sequences is that hNCACX has a spin diffusion part to allow the polarization transfer from CA to other side chain carbons. The longer the mixing time is, the farther the polarization can be transferred. DARR was used for spin diffusion. The magnetization transferred from CA to other side chain carbons and the transferring distance depends on mixing time. The longer mixing time is, the further magnetization can be transferred. Usually, 30ms would be chosen for intra-residue transferring. But for residues containing aromatic carbons, a longer mixing time is recommended. 30ms, 50ms, 100ms, 250ms and 500ms are the common choices for mixing time. From the overlaid 2D NCACX spectra, 10 ms mixing time led the polarization transferring to more carbons and stronger cross-peak intensities.

Number of points: number of points of indirect dimension means number of 2D spectra will be acquired in each indirect dimension. More number of points indicates longer data collecting time. But increase number of points will increase the spectrum resolution. Usually, N window can be set from 110 to 140 ppm (44 – 56 kHz in 9.4 T magnetic field) and CA window can be set from 45 ppm to 75ppm.

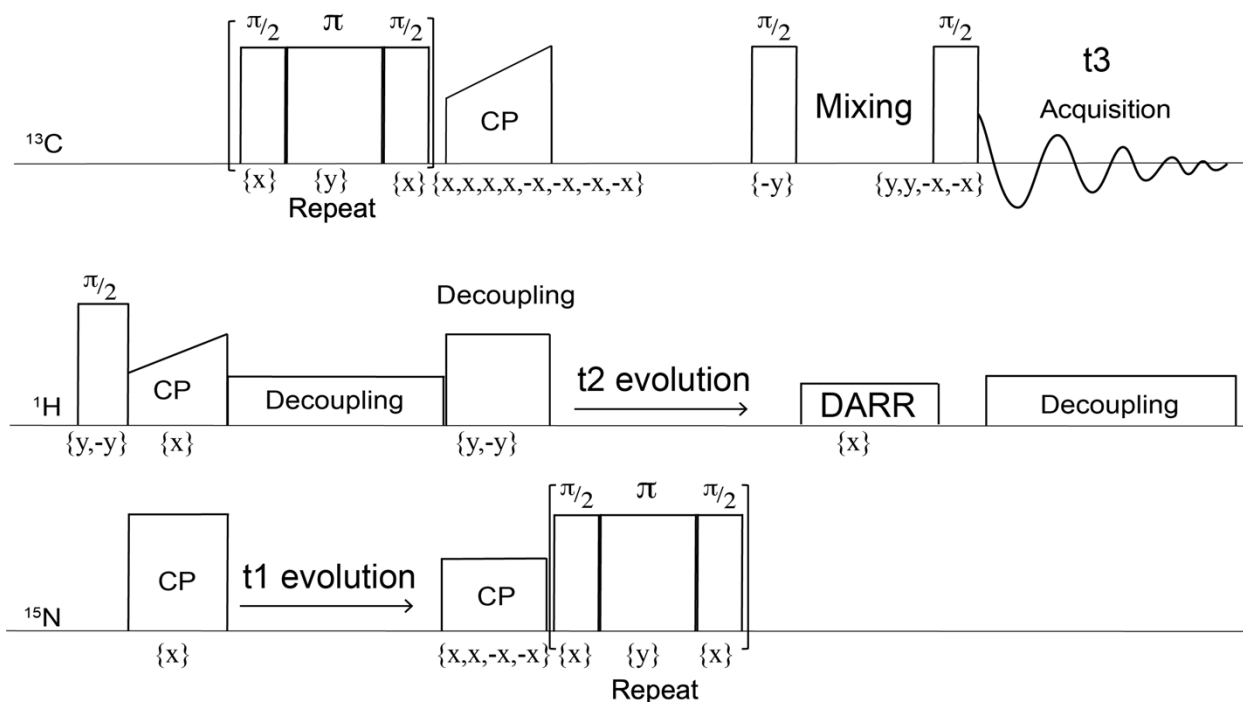


Figure 3.16. Pulse sequence for NCAcX. Parameters are the same as noted in Figure 3.12. Additional parameters are mixing time 30 ms and DARR field 8 kHz. The $\frac{\pi}{2}(x)-\pi(y)-\frac{\pi}{2}(x)$ sequence on ^{13}C channel is the WALTZ (Widely Adiabatic Low-Power Phase-Modulated) sequence consists of a series of R.F. pulses with specific phase and amplitude modulation) sequence to decouple the ^{13}C - ^{15}N dipolar coupling and scalar coupling.

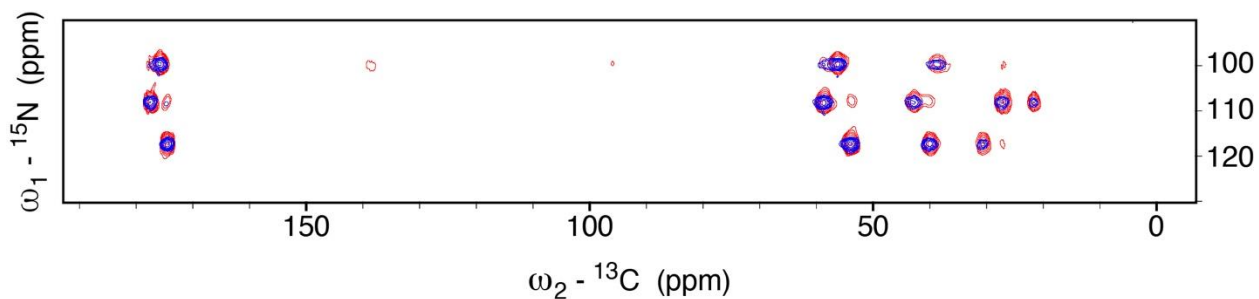


Figure 3.17. Overlaid spectra of 2D NCAcX of MLF with 32 scans under 8 kHz MAS at 278K. Blue — 5ms mixing time; Red — 10ms mixing time. The sizes of FID are 400 for CX and 24 for N. The spectral widths are 200 ppm with transmitter frequency offset set as 100 ppm for CX and 40 ppm with transmitter frequency offset centered at 130 ppm for N. The increment of delay for indirect dimension is 618.7375 μs . The spectrum is processed with no zero-filling and with Qsine for both dimensions, SSB = 2.

7. Collect a NCOCX spectrum. NCOCX experiment is very similar to NCAcX. The difference is the transmitter offset was set to ~ 175 ppm instead of 55 pm. Parameters can be re-optimized if needed.

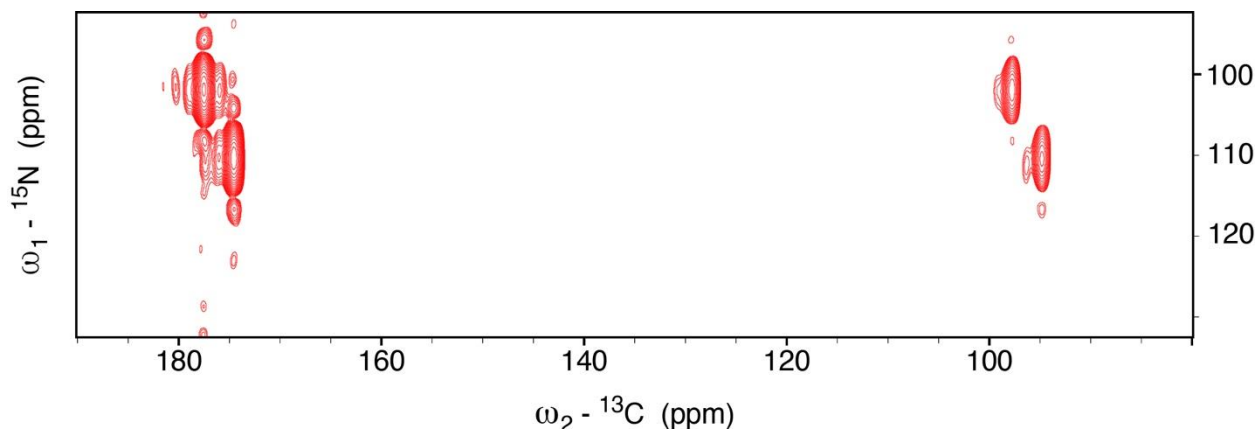


Figure 3.18. 2D NCOCX spectrum of MLF with 32 scans under 8 kHz MAS at 278K. Experimental parameters are: $2.25 \mu\text{s}$ ^1H 90° pulse, $4.2 \mu\text{s}$ ^{13}C 90° pulse, $8.4 \mu\text{s}$ ^{13}C 180° pulse, 1.4 ms $^1\text{H} \rightarrow ^{15}\text{N}$ CP contact time, 4 ms $^{15}\text{N} \rightarrow ^{13}\text{CO}$ CP contact time; $^1\text{H} \rightarrow ^{15}\text{N}$ CP: $37 - 75 \text{ kHz}$ for ^1H with ^{15}N field as of 20 kHz ; $^{15}\text{N} \rightarrow ^{13}\text{CO}$ CP: $11.47 - 12.68 \text{ kHz}$ for CA with 23.50 kHz for ^{15}N , decoupling 75 kHz for $^{15}\text{N} \rightarrow ^{13}\text{CA}$ CP and 66 kHz for others; acquisition time 40 ms , recycle delay 1 s . The sizes of FID are 480 for CX and 24 for N. The spectral widths are 200 ppm with transmitter frequency offset set as 100 ppm for CX and 40 ppm with transmitter frequency offset centered at 130 ppm for N. The increment of delay for indirect dimension is $618.7375 \mu\text{s}$. The spectrum is processed with no zero-filling and with Qsine for both dimensions, $\text{SSB} = 2$.

8. Collect CONCA spectrum. Parameters can be referred to NCA and NCO. There is no CON experiment as the result of very low sensitivity of ^{15}N detection. Ideally, NCO and CON should be very similar, so the parameter of NCO is informative. Compared with NCACX and NCOCX, CONCACX has the weakest signal caused by triple CP transfer.

Assigning spectra can be done either with 2D correlation spectra or 3D correlation spectra. As long as the sample have distinguished resonance in spectra, assigning peaks in 2D spectrum is possible. For example, MLF has three residues, and each residue is different from one another. The resonance of each carbon nucleus is distinguishable. Consequently, it is possible to assign MLF using ^{13}C - ^{13}C spectrum.

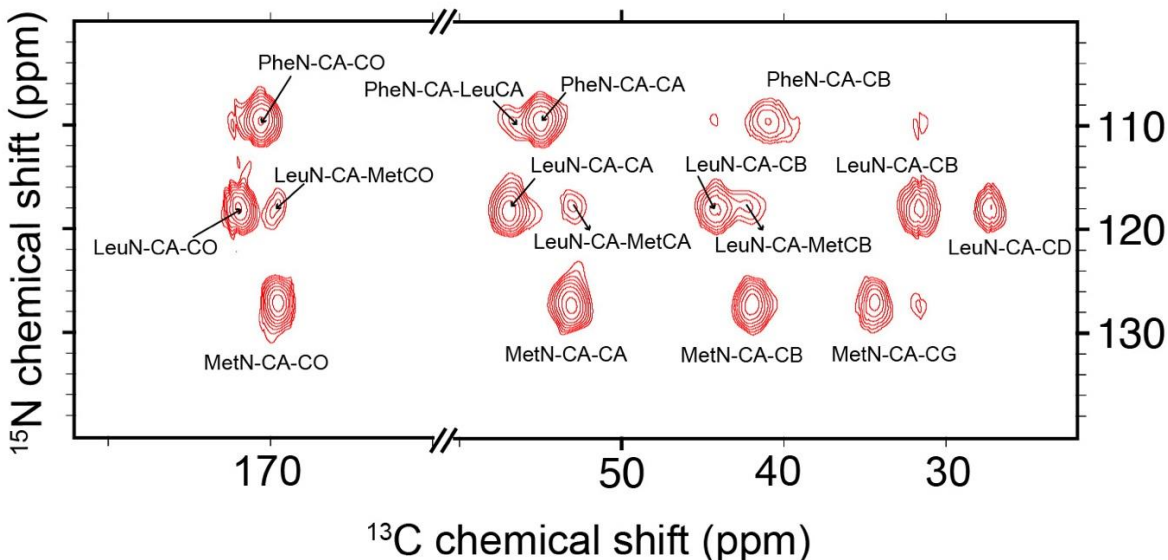


Figure 3.19. Assigned 2D NCACX spectrum of MLF.

To sequentially assign the spectra, 3D NCACX should be assigned firstly. It helps to determine the amino acid type. Chemical shifts of each amino acid are summarized in Table 3.1.

Table 3.1. Chemical shift (ppm) of MLF assigned from 3D NCACX at 298K.

Residue type	N	C α	C β	C γ	C δ	CO
Met	127.3	53.9	40.0	30.6	---	174.7
Leu	118.7	58.8	42.8	27.2	21.7	177.4
Phe	109.9	56.3	38.8	---	---	175.8

The sequential order of the residues can be determined with assist of inter-residue correlation ssNMR spectra assignment. For example, the Co of Met is at 174.7 ppm. Besides, there exist peaks corresponding to Leu at Co chemical shift of 174.7ppm in CONCACX spectrum. Therefore, it can be concluded that Leu is the following residue connecting after Met. Similarly, Leu's nitrogen is at 118.7ppm. At the N projection of 118.7ppm of NCOCX spectrum, the existence of Met indicating that Met is the preceding residue of Leu. Finally, the sequence is assigned to be M-L-F.⁷

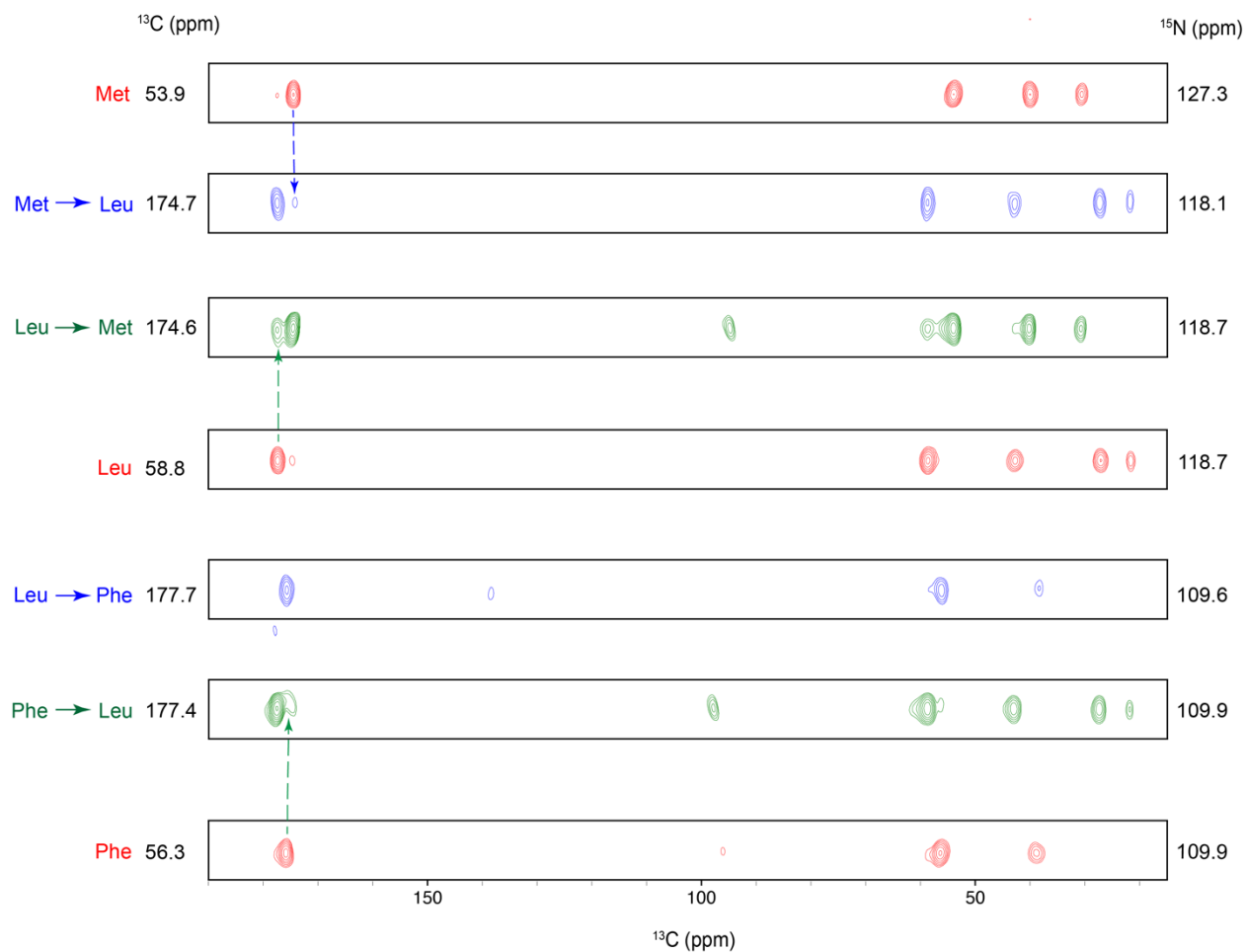
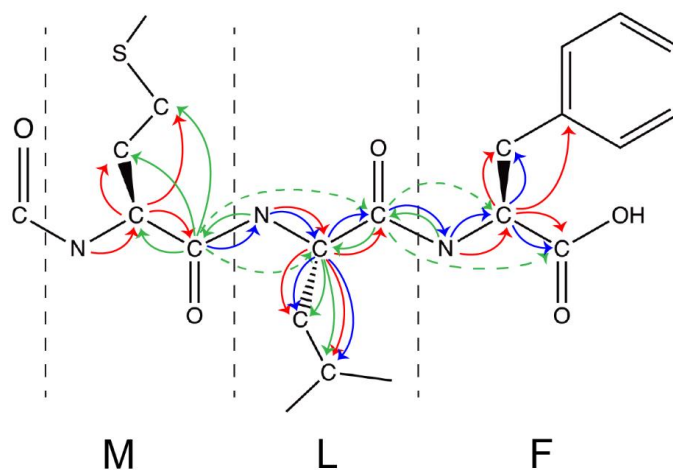


Figure 3.20. 3D ssNMR spectra in strip plot for sequential assignment of MLF. The data were collected by 3D NCACX (red), 3D NCOCX (green), and 3D CONCACX (blue). Parameters are the same as noted in Figure 3.12 except for ^{15}N pulse field becomes 28.9 kHz for $^{15}\text{N} \rightarrow ^{13}\text{CO}$ CP. The spectra were processed with widow function QSINE, SSB=2.

What described earlier is a simple example about sequential assignment of proteins. The assignment is much easier because there are only three residues in MLF. Typically, the starting point of assignment should be the isolated resonance which can be certainly assigned to a particular amino acid type. Then, the preceding and following residues can be found by tracing the connectivity.⁸ However, large proteins usually have more residues and repeating residues. Some proteins are heterogenous, and different structures give rise to different resonance. The ambiguous spectra making assignment more complicated.

REFERENCES

- (1) Chapman, R. P.; Widdifield, C. M.; Bryce, D. L. Solid-State NMR of Quadrupolar Halogen Nuclei. *Prog Nucl Magn Reson Spectrosc* 2009, 55 (3), 215–237. DOI: 10.1016/j.pnmrs.2009.05.001.
- (2) Frye, J. S.; Maciel, G. E. Setting the Magic Angle Using a Quadrupolar Nuclide. *Journal of Magnetic Resonance (1969)* 1982, 48 (1), 125–131. DOI: 10.1016/0022-2364(82)90243-8.
- (3) Hoffman, R. Solid-State Chemical-Shift Referencing with Adamantane. *Journal of Magnetic Resonance* 2022, 340, 107231. DOI: 10.1016/j.jmr.2022.107231.
- (4) Hayashi, S.; Hayamizu, K. Shift References in High-Resolution Solid-State NMR. *Bull Chem Soc Jpn* 1989, 62 (7), 2429–2430. DOI: 10.1246/bcsj.62.2429.
- (5) Morcombe, C. R.; Zilm, K. W. Chemical Shift Referencing in MAS Solid State NMR. *Journal of Magnetic Resonance* 2003, 162 (2), 479–486. DOI: 10.1016/S1090-7807(03)00082-X.
- (6) Asakura, T.; Suzuki, Y.; Nakazawa, Y.; Yazawa, K.; Holland, G. P.; Yarger, J. L. Silk Structure Studied with Nuclear Magnetic Resonance. *Prog Nucl Magn Reson Spectrosc* 2013, 69, 23–68. DOI: 10.1016/j.pnmrs.2012.08.001.
- (7) Hong, M.; Griffin, R. G. Resonance Assignments for Solid Peptides by NMR. *J Am Chem Soc* 1998, 7863 (28), 7113–7114.
- (8) Wang, S.; Matsuda, I.; Long, F.; Ishii, Y. Spectral Editing at Ultra-Fast Magic-Angle-Spinning in Solid-State NMR: Facilitating Protein Sequential Signal Assignment by HIGHLIGHT Approach. *J Biomol NMR* 2016, 64 (2), 131–141. DOI: 10.1007/s10858-016-0014-4.

Chapter 4 Lipid acyl chain protrusion induced by the influenza virus hemagglutinin fusion peptide detected by NMR paramagnetic relaxation enhancement

4.1 Introduction

Many zoonotic diseases including AIDS, influenza, and COVID are caused by viral pathogens that are membrane-enveloped.¹⁻⁵ An initial step in cellular infection is fusion (joining) of the viral and target cell membranes with consequent deposition of the viral capsid in the cytoplasm. Enveloped viruses have glycoprotein spikes whose protein have a receptor-binding subunit (RbSu) followed by a fusion subunit (FsSu), with typical proteolytic cleavage between the two subunits.^{1,6-10} The FsSu has a single transmembrane domain and a large N-terminal ectodomain (Ed) outside the virus membrane. Each spike contains a core with a defined number (often 3) of non-covalently-associated Ed's of FsSu's, and the same number of RbSu's that are non-covalently bound with this core. After the virus is in the host, RbSu's bind to specific receptor molecules on the exterior of target cells, and for some viruses, there is subsequent endocytosis. The RbSu's move away from the FsSu Ed core, and the core changes to a new structure, typically a thermostable trimer-of-hairpins with $T_m > 0^\circ\text{C}$.¹¹⁻¹⁵ There isn't sequence homology among the RbSu's of different virus families which can be partly understood because the RbSu's of different virus families bind different molecules. More surprisingly, there also isn't sequence homology or sequence-length homology among the FsSu's of different virus families. As noted above, the final Ed structure is typically a hairpin, but there are substantial length and structural differences between the hairpins of different families.¹⁶⁻²⁰

There are also large geometric changes of the membranes during fusion, including intermediate structures, but there aren't yet clear experimental data about the relative timings of changes in membrane vs. FsSu Ed structure. Figure 4.1 displays a common model for the membrane changes. These changes are in time-sequence: **(a)** initial close (nm) apposition of the viral and target membranes; **(b)** stalk intermediate that connects and is contiguous with the outer leaflets of the two membranes; **(c)** hemifusion diaphragm with contiguous inner leaflets of the two membranes; **(d)** pore formation in the diaphragm; and **(e)** pore expansion with final state of contiguous membranes and viral contents in the cytoplasm.^{4,21,22} There are some experimental data that support this model as well as computational studies. The computational consensus estimates for energy barriers of uncatalyzed fusion are ~ 25 kcal/mol for step **a** and ~ 10 kcal/mol between the step **a** \rightarrow **b**, **b** \rightarrow **c**, and **c** \rightarrow **d** states.⁴

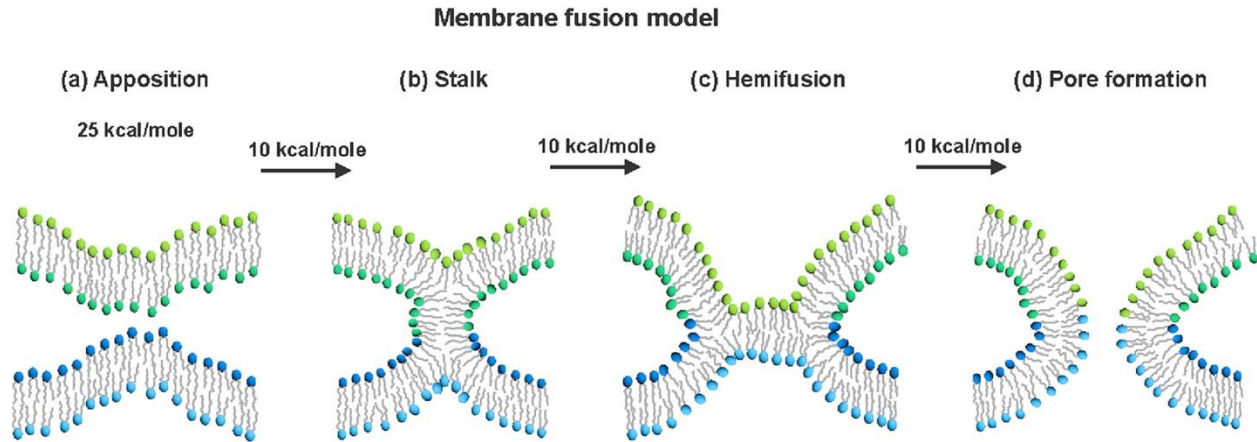


Figure 4.1. Pictorial representation of a common membrane fusion model that includes (a) initial close apposition of the viral and target membranes; (b) stalk formed from the outer leaflets of the two membranes; (c) hemifusion diaphragm that is contiguous with the inner leaflets of the two membranes; and (d) pore formation. The estimates of the free energy barriers for membrane apposition and for transformation between membrane intermediates are from computational studies of uncatalyzed fusion. The figure doesn't show the final step of pore expansion that precedes full contents mixing. The different colors of the headgroups are meant to visually enhance the changes in membrane topology during fusion but don't describe the locations of specific lipids during fusion. During the ~20 s estimated lifetime of a membrane intermediate structure in viral fusion, a lipid molecule could diffuse over $\sim 10^{10} \text{ \AA}^2$ leaflet area.

FsSu's have a N-terminal region (Ntr) that is not part of the final hairpin structure. The Ntr is often folded within the initial spike and then released as the hairpin forms.^{6–10} The Ntr length varies among FsSu's from different viral families, and the range of lengths is typically between 30 and 250 residues. Within a Ntr, there are one or more proposed “fusion peptide” segments that are hypothesized to bind the target membrane during fusion.^{23–26} The membrane-bound fusion peptide(s) may reduce the 25 kcal/mol apposition barrier, in conjunction with the more C-terminal hairpin structure of the Ed and viral transmembrane domain. In addition, a membrane may be modified by fusion peptide so that there is also reduction in the 10 kcal/mol barriers between subsequent membrane intermediates.^{27–38} A fusion peptide segment has typically been identified by observation of mutations that reduce viral fusion and/or infection without affecting initial spike structure.^{15,24,39–44} In addition, a fusion peptide sequence is typically highly-conserved and should bind membrane.^{41,42,45,46}

The present study is specifically focused on the fusion peptide (Fp) of the influenza virus FsSu which is subunit 2 of the hemagglutinin protein (Ha2). The influenza RbSu (Ha1) binds sialic acid followed by endocytosis of the virus and then endosome maturation that includes pH reduction to ~ 5 .¹ At low pH, Ha1 separates from the Ha2 Ed and the Ed then changes to the final trimer-of-

hairpins structure with accompanying fusion between the viral and endosome membranes. The Ha2 Fp is the ~20 N-terminal and highly-conserved residues of Ha2. The Fp has been identified by: (1) significant attenuation of fusion when specific Fp residues are mutated; (2) very high sequence conservation among different influenza subtypes; and (3) observation of membrane-bound Fp after influenza virus fusion.^{15,23,40,44,46} The Ha2 Fp often adopts helical hairpin structure and is a mixture of: (i) closed structure in which the two antiparallel helices are in van der Waals contact; and (ii) semi-closed structure in which the Phe-9 sidechain is inserted between the two helices.^{47,48}

The effects of the Ha2 Fp on membrane have been studied by computer simulations by several different groups. These simulations have typically been done using a membrane with Fp peptide without the rest of Ha2. One commonly-observed effect is a higher (~4 – 20 ×) probability for chain protrusion by lipids next to vs. further from the Fp (Figure 4.2).^{49–51} Protrusion is specifically defined as one or more carbons of the lipid chain being at least 1 Å closer to the aqueous phase than the P nucleus of the lipid headgroup. Protrusion is a functionally-interesting motion. As depicted in Figure 4.1, an early fusion step is the topological transition from (a) initial apposition of viral and target membranes to (b) stalk that connects the two membranes and is contiguous with the outer leaflets of these membranes. This step requires protrusion by some outer leaflet lipids of both membranes. The hypothesized correlation between increased lipid chain protrusion near Fp and stalk formation is supported by coarse-grained computer simulations of fusion that begin with full-length Ha2 with final trimer-of-hairpins structure and with Fp's in one membrane and transmembrane domains in the other membrane.⁵² In the absence of Fp, simulations show that at any given time, ~1% of the lipids have a protruded chain.^{49–51} For simulations with Fp, ~ 4–20 increased protrusion probability is observed for both chains of a lipid that is Fp-adjacent and for a variety of Fp structures.^{49–51} Both interfacial and transmembrane locations of the Fp have been observed as well as a variety of geometries of the protruded lipid relative to the Fp. These include: (1) “straddling” of the protruded chain over the Fp; and (2) hydrogen bonding between the headgroup phosphate oxygen and one of the four N-terminal residues of the Fp with associated headgroup intrusion into the bilayer.^{49,50}

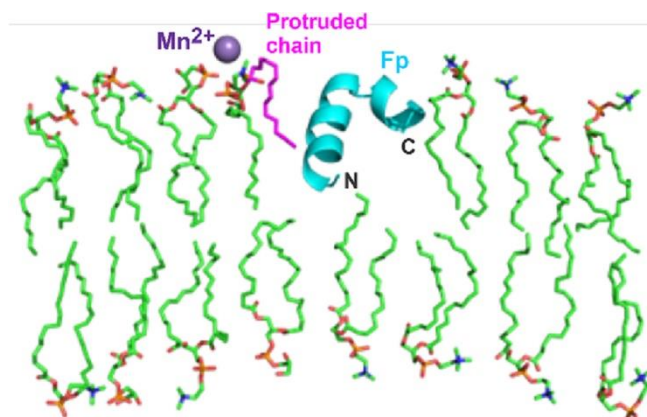


Figure 4.2. Representative picture of lipid acyl chain protrusion near a Fp. A set of 128 POPC and 32 POPG lipids in a pre-assembled bilayer were energy- minimized in a water box using the CHARMM/Membrane Builder/Bilayer Builder/Membrane Only System molecular dynamics program. A membrane cross-section is displayed with lipid acyl chains in light green. A representative protruded chain in magenta is next to a Fp backbone in turquoise and near a Mn^{2+} in purple that is bound to a lipid headgroup. The picture shows a protruded palmitoyl chain and a Fp with semi-closed structure with marked N- and C- termini. There is increased protrusion in simulations for both palmitoyl and oleoyl chains and for lipids next to a variety of Fp structures. In addition, Fp in simulations is observed with both interfacial and transmembrane locations and protruded lipids exhibit a variety of geometries relative to the neighboring Fp.

Despite its potential significance, to our knowledge there hasn't yet been direct experimental observation of increased lipid protrusion for membrane with vs. without Fp. If chain protrusion is increased with Fp, there will be associated decreases in chain order parameters, and such decreases were observed and quantified in some of the computational simulations with membrane with Fp peptide.^{49,51} However, these computational predictions were contradicted by electron paramagnetic resonance (EPR) and fluorescence spectra that showed increases in chain order parameters for samples with Fp.^{28,33} Such increases were also observed for membrane with putative fusion peptides from other virus families.^{34–36} The sequences are non-homologous with the Ha2 Fp sequence. However, more recently, ²H nuclear magnetic resonance (NMR) spectra of membrane with perdeuterated lipids showed decreases in chain order parameters with vs. without Fp.³⁷ The NMR-derived order parameters were in semi-quantitative agreement with the simulation-derived parameters.^{49,51,53} The seeming contradiction between EPR and NMR/simulation may be explained by: (i) EPR detects order for the spin labels and only 0.005 mol fraction of the sample lipids are spin-labeled, whereas (ii) NMR and simulation detect order for all lipids in the sample. The EPR-detected ordering of the spin label with Fp may be due to preferential binding of Fp to the spin label. This hypothesis is supported by the dose response of

ordering with respect to Fp mole fraction. The ordering reaches its maximum value when the Fp: lipid mole ratio 0.002, which is similar to the spin-labeled lipid: total lipid mole ratio.

There is other indirect experimental support for increased lipid protrusion from analysis of the large increases in ^2H NMR transverse relaxation rates (R_2 's) of deuterated lipids in samples with vs. without Fp.³⁸ The increases are interpreted to be due to modulation of the ^2H NMR frequency as the lipid laterally diffuses in the membrane leaflet. For lipid next to vs. further from Fp, there are larger vs. smaller amplitudes of mean-squared ^2H quadrupolar fields that are correlated with the smaller vs. larger chain order parameters. The experimental increases in R_2 's with vs. without Fp are in semi-quantitative agreement with values calculated using experimentally-based estimates of order parameters and the time for a lipid to diffuse past a Fp. Both the decreases in chain ^2H order parameters and large increases in ^2H R_2 's were also observed for a membrane sample with the putative fusion peptide of the HIV F₅Su (gp41).³⁸ The gp41 fusion peptide and the Ha2 Fp have non-homologous sequences and also adopt very different membrane-bound structures.^{54–58} Although these earlier experimental data are consistent with increased chain motion for lipid next to vs. further from Fp, they don't directly evidence increased protrusion (Figure 4.2). This knowledge gap motivated the present experimental NMR study in which chain protrusion is probed using comparison of chain ^{13}C R_2 's in samples with vs. without the paramagnetic Mn^{2+} species.^{59–61} The Mn^{2+} binds to the phosphate oxygens of the lipid headgroup, so a chain ^{13}C R_2 is augmented when the ^{13}C site is protruded into the headgroup region. The change in protrusion probability with vs. without Fp is probed by the difference in Mn^{2+} -associated increase in R_2 , i.e. paramagnetic relaxation enhancement (PRE). There is spectral resolution of some of the NMR signals from different $-\text{CH}_2$ sites, so the approach also yields information about how Fp affects protrusion for $-\text{CH}_2$ groups closer to the headgroup vs. the chain terminus.⁶²

4.2 Results

4.2.1 NMR sample preparation

The goal of this study is to probe whether or not lipid acyl chains have greater probability of protrusion into the aqueous phase when Fp is membrane-bound, as has been observed in some simulations. There are several considerations for sample preparation including lipid composition of the membrane and optimal mole% values of membrane-bound Fp and Mn^{2+} , where %Fp or % Mn^{2+} = (mole Fp or Mn^{2+})/(mole lipid) $\times 100$. One consideration for %Fp is simulation data showing that protrusion probability is increased by a factor of 4–20 \times for lipids next to a Fp,

whereas changes are much smaller for more distant lipids.^{49–51} During NMR data collection, a lipid molecule experiences rapid lateral diffusion in the liquid-crystalline phase and will spend time both next to and further from a Fp.^{63–65} Larger %Fp is anticipated to result in greater fraction of time next to Fp but there will also be undesired effects if %Fp is too large. These include oligomeric β sheet rather than monomer helical hairpin Fp structure, with the latter being the likely Fp structure in full-length Ha2.^{15,66} For the 3% Fp of our samples, earlier studies evidence that the Fp adopts monomer helical hairpin structure, and the membrane retains the liquid-crystalline bilayer phase.^{29,37,38,48,67,68} This retention was supported by comparison between lipid samples without vs. with Fp. The two sample types exhibited similar lineshapes, linewidths, and relaxation rates for both their ^{31}P and ^2H static NMR spectra.

We first prepared POPC samples because the earlier simulations had typically used POPC. For sample preparation with an aliquot corresponding to 5% Mn^{2+} , the $[\text{Mn}^{2+}]_{\text{free}} \approx 0$ in the supernatant after centrifugation which correlates with $\sim 5\%$ bound Mn^{2+} , i.e. complete binding to lipid. By contrast, most Mn^{2+} did not bind to POPC with 3% bound Fp. The low binding may be due to electrostatic repulsion between Mn^{2+} and Fp, where the calculated Fp charge is +1.6. We switched to POPC:POPG (4:1), referred to as “PC:PG”, with the choice of POPG based on its -1 charge and on otherwise similar properties to POPC.⁶⁹ For sample preparation with an aliquot corresponding to 5% Mn^{2+} , the $[\text{Mn}^{2+}]_{\text{free}}$ correlated with $\sim 5\%$ bound Mn^{2+} for PC:PG without Fp and $\sim 4\%$ bound Mn^{2+} for PC:PG with 3% Fp. For PC:PG either without or with Mn^{2+} , there was complete binding of Fp to lipid, based on $A_{280} \approx 0$ in the supernatant after centrifugation. This result is consistent with $\text{Fp}_{\text{bound}}:\text{Fp}_{\text{free}} \approx 10^4$ calculated by $K_{\text{bind}} \times [\text{lipid}]$ where $K_{\text{bind}} \approx 10^6 \text{ M}^{-1}$ is the previously-determined binding constant and $[\text{lipid}] \approx 10^{-2} \text{ M}$ in our sample preparation.²⁵

^{13}C CP NMR spectra without Hahn echo are displayed for the lipid samples in Figure 4.3 either (a) without or (c) with Fp. Assignments are shown for the lipid acyl chain peaks use the Figure 3 carbon numbering with prime (') for the palmitoyl chain and no prime for the oleoyl chain.⁶² There typically isn't resolution of individual peaks for POPC vs. POPG. There are resolved peaks for sites with distinctive bonding environments, but with superposition of signals from at least two sites. The 9,10 peak exhibits partial resolution of the 9 and 10 signals at higher and lower shift, respectively. The * peak is a superposition of signals from the 4–7, 12–15, and 4'-13' sites. For the same lipid site peak, there are negligible spectral shift or width differences for samples without vs. with Fp. There aren't peaks that could correspond to Fp signals, with possible explanations being

the: (1) 3% Fp concentration; and (2) broader Fp linewidths because of less motional averaging for Fp vs. lipid.

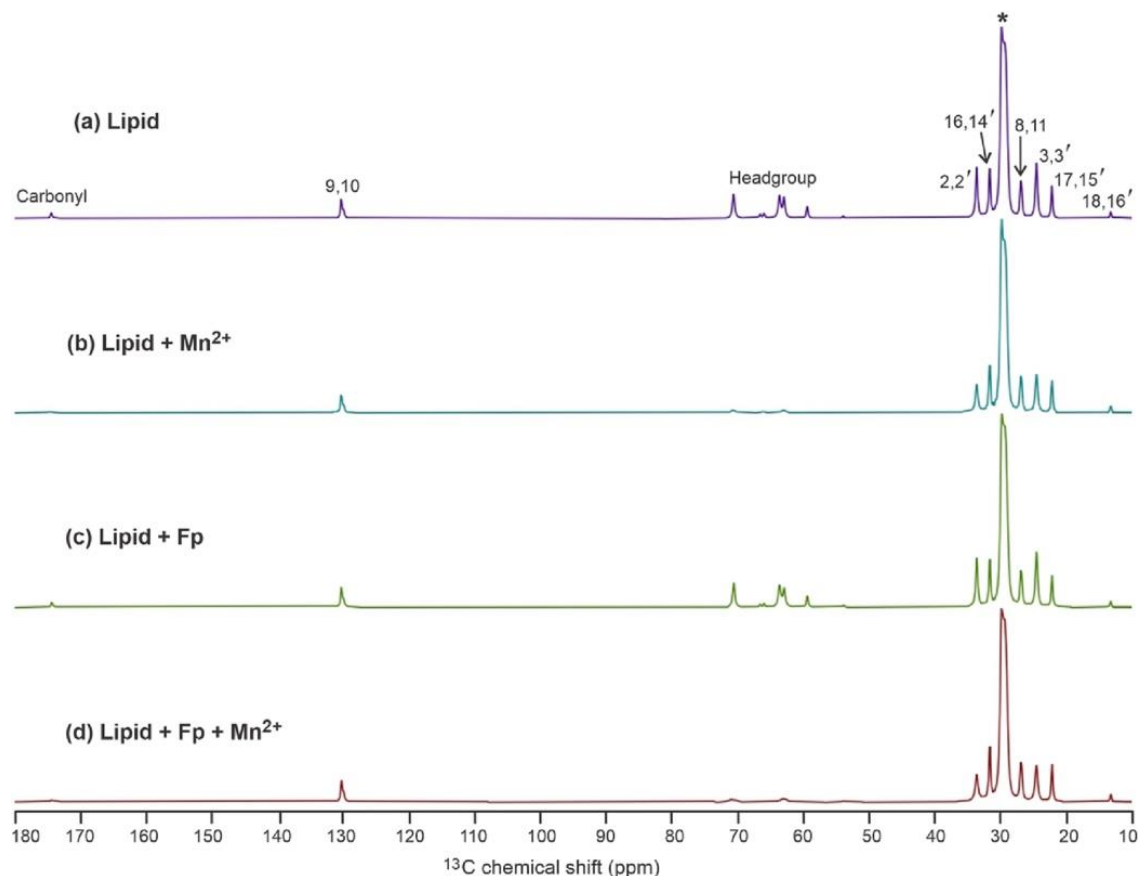


Figure 4.3. ^{13}C NMR spectra of samples containing (a) Lipid, (b) Lipid + Mn^{2+} , (c) Lipid + Fp, and (d) Lipid + Fp + Mn^{2+} , with 0.5% Mn^{2+} and 3% Fp that are calculated as $(\text{mole } \text{Mn}^{2+} \text{ or Fp})/(\text{mole lipid}) \times 100$. There is POPC:POPG (4:1) lipid composition. The data were acquired after CP without the Hahn echo. The vertical scales of the four spectra have been adjusted so that the “*” peaks at 30 ppm have the same height. Assignments are displayed for the lipid acyl chain peaks use the Figure 2 carbon numbering with prime (') for the palmitoyl chain and no prime for the oleoyl chain. There isn't resolution of individual peaks for POPC vs. POPG. There are resolved peaks for sites with distinctive bonding environments, but with superposition of signals from at least two sites. Peaks are assigned for: (1) the 2,2' and 3,3' sites that are one and two bonds from the carbonyl groups; (2) the 9,10 C=C and 8,11 C=C adjacent sites of the oleoyl chain; (3) the 18,16' $-\text{CH}_3$ sites; and (4) the 17,15' and 16,14' sites that are one and two bonds from the $-\text{CH}_3$ groups. The * peak is a superposition of signals from the 4–7, 12–15, and 4'–13' sites. The carbonyl peak and headgroup peak regions are also noted.

We first measured the effect of % Mn^{2+} on ^{13}C R_2 , with the goal of finding the optimal % Mn^{2+} for detection of lipid protrusion using the R_2 difference between samples with vs. without Mn^{2+} , i.e. $\Gamma_2 = R_{2,\text{Mn}} - R_{2,\text{NoMn}}$. The Fp effect on protrusion is assessed by comparison of $\Gamma_{2,\text{Fp}}$ vs. $\Gamma_{2,\text{noFp}}$. The Mn^{2+} is likely close to the negatively-charged lipid phosphate oxygens in a lipid headgroup. The

$\Gamma_2 \propto \langle r^{-6} \rangle$ where r is the ^{13}C - Mn^{2+} distance and $\langle \dots \rangle$ is the average over the ~ 10 ms NMR measurement time.⁵⁹ The $\langle r^{-6} \rangle$ will depend on % Mn^{2+} as well as lipid ^{13}C site (Figures. 6.2, 6.3). We hypothesized that increased lipid protrusion with Fp could be mostly clearly observed if $\Gamma_{2,\text{noFp}}$ were comparable or smaller than $R_{2,\text{noMn,noFp}}$. This is based on the idea that Fp-induced increases in $\langle r^{-6} \rangle$ would be more readily detectable if $\Gamma_{2,\text{noFp}}$ isn't already approaching its maximal value. We first tried 5% Mn^{2+} in lipid without Fp and observed complete loss of the 2,2' and 3,3' signals as well as apparent $\Gamma_2/R_{2,\text{noMn}}$ ratios of ~ 10 for 8,11 and 9,10, ~ 6 for *, and ~ 5 for 16,14'. We then investigated % Mn^{2+} in the 0.5–1.25% range and observed that even at 0.5%, $\Gamma_2 > 0$ for many peaks and $\Gamma_2/R_2 < 1$ for all peaks (Tables 4.1, B1, and B2). We then prepared samples with 0.5% Mn^{2+} for more complete analysis. There was no detectable $[\text{Mn}^{2+}]_{\text{free}}$ for the lipid without Fp sample and small $[\text{Mn}^{2+}]_{\text{free}}$ for the lipid with Fp sample. The % Mn^{2+} bound were 0.50 and 0.48%, respectively.

Table 4.1. Site-specific ^{13}C transverse relaxation rates of acyl chains of POPC:POPG (4:1) membrane and % Mn^{2+} dependence (fitting uncertainties in parentheses)^a.

Mn ²⁺ %	¹³ C R ₂ (s ⁻¹)				¹³ C Γ ₂ (s ⁻¹)			
	2,2'	3,3'	*	16,14'	2,2'	3,3'	*	16,14'
0	28.1(1.2)	20.3(1.0)	15.9(0.2)	13.9(0.9)				
0.5	50.2(1.8)	34.3(2.2)	20.5(0.9)	15.4(1.0)	22.1(2.5)	14.0(2.4)	4.6(0.9)	1.4(1.3)
0.75	81.9(6.3)	59.1(3.4)	26.4(0.9)	13.7(1.9)	53.8(6.5)	38.8(3.5)	10.5(0.9)	0.2(2.1)
1.00	92.1(3.4)	60.8(4.6)	28.4(1.5)	22.0(1.4)	64.0(3.8)	40.5(4.7)	12.5(1.5)	8.1(1.6)
1.25	93.9(7.0)	67.8(5.8)	29.2(1.4)	20.7(1.9)	65.8(7.2)	47.6(5.9)	13.3(1.4)	6.8(2.1)

^a Each ^{13}C transverse relaxation rate (R_2) was determined from best-fitting the integrated NMR peak intensity S vs. delay time τ using $S = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. The fitting uncertainty of R_2 is given in parentheses. The Γ_2 values are the differences between the best-fit R_2 values of samples with vs. without Mn^{2+} . The * peak is the superposition of the 4–7, 12–15, and 4'-13' signals. The typical ppm integration ranges for peaks are: 2,2', 33.00–37.00; 3,3', 24.00–26.30; *, 28.30–31.50; 16,14', 31.50–33.00. The % $\text{Mn}^{2+} = (\text{mole bound Mn}^{2+})/(\text{mole lipid}) \times 100$.

4.2.2 ^{13}C NMR spectra and relaxation

Figure 4.3 displays the ^{13}C CP NMR spectra without Hahn echo of the (b) lipid + Mn^{2+} and (d) lipid + Fp + Mn^{2+} samples with comparison to the (a, c) samples without Mn^{2+} . The vertical scales of the four spectra have been adjusted so that the * peaks at 30 ppm have the same height. For the

same lipid site peak, there is negligible spectral shift without vs. with Mn^{2+} . For both lipid and lipid + Fp, there is attenuation of the lipid 2,2' and 3,3' peak intensities with vs. without Mn^{2+} , whereas there is not obvious attenuation for the 16,14' , 17,15' ,and 18,16' peaks. In addition, the lipid headgroup and CO signals are highly attenuated by Mn^{2+} .

Figure 4.4 displays 2,2' spectral signal intensities vs. $\Delta \tau$ (increment in dephasing time) of the four samples. The spectra were acquired with the CP-Hahn echo sequence (Figure 4.3). The vertical scales of the spectra of each sample were adjusted so that the $\Delta \tau = 0$ spectral peaks of all samples have the same height. The signal intensities of the (a) lipid and (c) lipid + Fp samples exhibit semi-quantitatively similar attenuation of signal intensities with $\Delta \tau$. There is greater attenuation for the (c) lipid + Mn^{2+} sample and even greater attenuation for the (d) lipid + Fp + Mn^{2+} sample. Figure 4.4 provides qualitative spectral evidence which supports the hypothesis that bound Fp induces higher probability of lipid protrusion. Although the shortest τ value was 1.25 ms for panel (a) vs. 2.00 ms for (b-d), the NMR intensity $S(\tau)$ for all data was close to the $S(0)$ intensity prior to relaxation-associated decay, i.e. the $S(\tau)/S(0)$ ratio in (a-d) is ~ 0.96 , 0.90 , 0.95 , and 0.88 , respectively, as calculated using the R_2 rates described in the next paragraph.

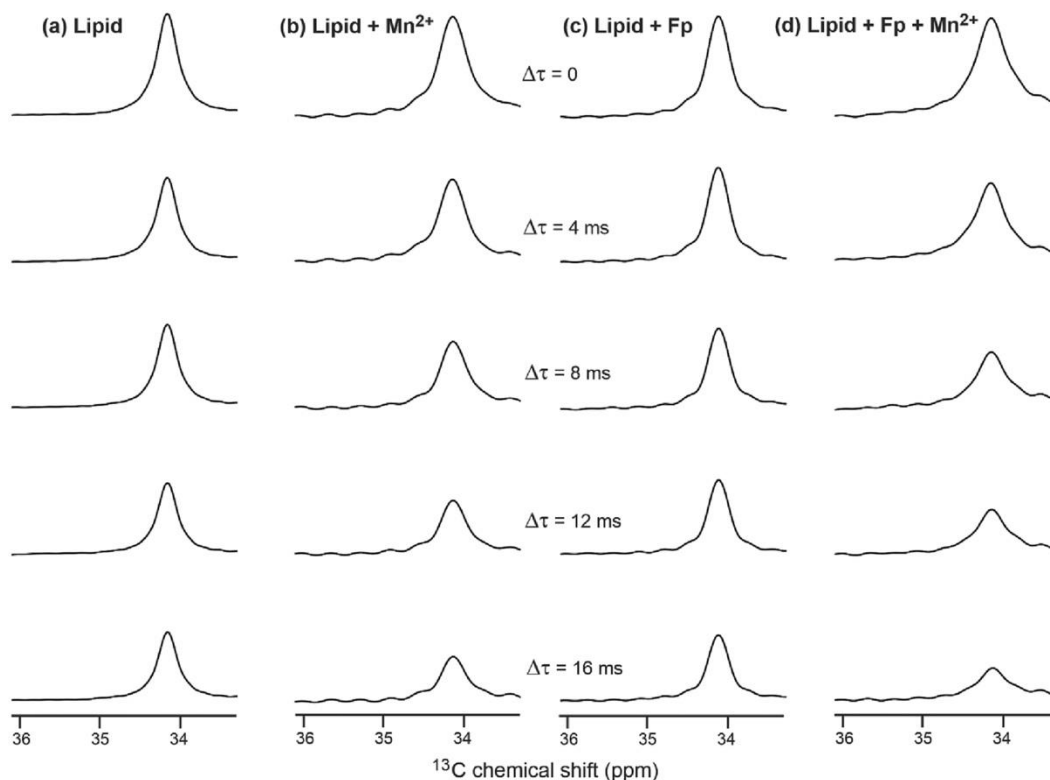


Figure 4.4. The 2,2' spectral signals vs. $\Delta \tau$ (increment in dephasing time) of samples containing (a) Lipid, (b) Lipid + Mn^{2+} , (c) Lipid + Fp, and (d) Lipid + Fp + Mn^{2+} , with 0.5% Mn^{2+} and 3% Fp that are based on (mole Mn^{2+} or Fp)/(mole lipid) \times 100. Spectra were acquired with the CP-Hahn echo sequence. The vertical scales of the spectra of each sample were adjusted so that the $\Delta \tau = 0$ spectral peaks have the same height. For these top spectra, $\tau = 1.25$ ms for (a) and 2.00 ms for (b-d).

The integrated intensities vs. τ of each acyl chain $-\text{CH}_2$ signal were fitted to single exponential decays, i.e. $S(\tau) = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. Figure 4.5 displays plots and fittings of $S(\tau)/A$ vs. τ for some of the signals and Table 4.2 gives the best-fit R_2 's for all fittings as well as the Γ_2 's, the R_2 changes for samples with vs. without Mn^{2+} . Table 4.2 provides the uncertainties in the R_2 and Γ_2 values in parentheses. Fitting is done both for the full integration range of the * peak and for smaller integration ranges of partially-resolved peaks with predominant contributions from a more limited number of sites (Figure B2). The uncertainties of the experimental peak intensities were calculated as the standard deviations of the integrated intensities in noise regions of the spectra. For two-site peaks, these experimental uncertainties were typically between 10^{-3} and 10^{-2} , i.e. smaller than the dimensions of points in the plots in Figure 4.5. Figure B3 displays the data and fittings on a logarithmic scale. Figure 4.6 displays a bar plot of the Γ_2 's of samples without vs. with Fp, with data from each of the resolved peaks that are due to two

^{13}C sites in the acyl chains (Figure 4.3). Table 4.3 lists the experimental linewidths ($\Delta\nu$'s) of the peaks in the four samples as well as differences $\Delta[\Delta\nu]$ with vs. without Mn^{2+} . Table 4.4 provides the inhomogeneous contributions to the linewidths of the resolved peaks, calculated using $\Delta\nu_{\text{inhom}} = \Delta\nu_{\text{exp}} - R_2/\pi$. The $\Delta\nu_{\text{inhom}}$ values are shown for resolved peaks of individual samples, along with average values for the four samples and associated standard deviations. The four samples for Figures 4.3--4.6 were prepared around the same time and as similarly as possible other than the prescribed absence vs. presence of Mn^{2+} and/or Fp. Replicate datasets were acquired and analyzed and Table B3 provides the best-fit R_2 's for the replicate data. The differences in best-fit R_2 's between replicates are typically comparable to the uncertainties for the differences. Table B4 displays best-fit R_2 's for replicate samples and the R_2 's are also similar between samples.

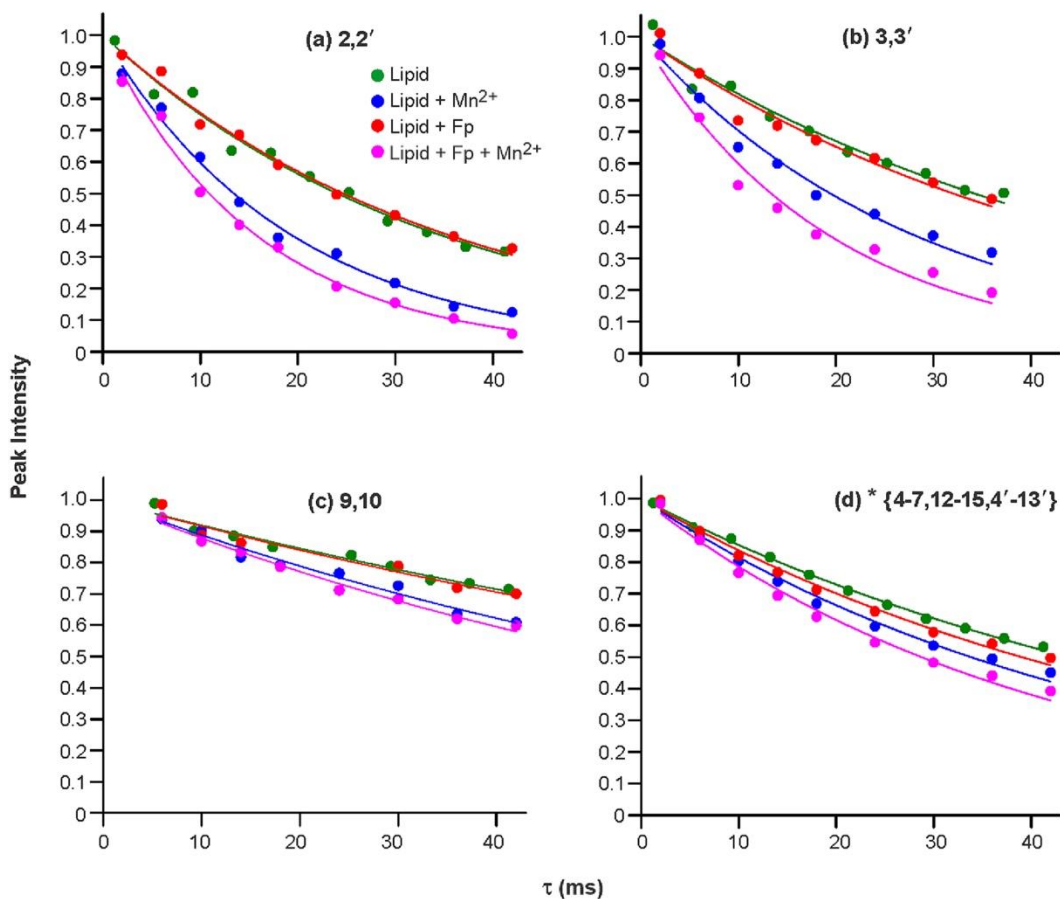


Figure 4.5. Integrated peak intensities vs. dephasing time (τ) and best-fit single exponential decays for the Lipid, Lipid + Mn²⁺, Lipid + Fp, and Lipid + Fp + Mn²⁺ samples. Data and fittings are displayed for the (a) 2,2'; (b) 3,3'; (c) 9,10; and (d) * peaks. The * peak is a superposition of signals from the 4–7, 12–15, and 4'–13' sites. The peak intensities vs. τ were fitted to $A \times \exp(-R_2 \times \tau)$ with A and R_2 as fitting parameters. The displayed intensities have been divided by A so that the best-fit intensity = 1 for all peaks when $\tau = 0$. The best-fit R_2 's and their uncertainties are presented in Table 4.2. The uncertainties in the peak intensities were calculated as the RMSD's of ten different integrals in noise regions of the spectra. For the two-site peaks, these uncertainties were typically between 10^{-3} and 10^{-2} and less than the dimension of the points in the plots.

Table 4.2. Site-specific ^{13}C transverse relaxation rates of acyl chains of POPC:POPG (4:1) membrane and Mn^{2+} and Fp dependences (uncertainties in parentheses)^a.

^{13}C	$R_2(\text{s}^{-1})$				$\Gamma_2(\text{s}^{-1})$	
	w/o Fp		3% Fp		w/o Fp	3% Fp
	w/o Mn^{2+}	0.5% Mn^{2+}	w/o Mn^{2+}	0.5% Mn^{2+}		
2,2'	28.8(1.4)	51.5(2.1)	28.1(1.2)	63.5(3.2)	22.8(2.6)	35.4(3.4)
3,3'	20.0(1.5)	35.2(2.4)	21.3(2.0)	51.2(4.0)	15.2(2.8)	29.8(4.1)
8,11	9.1(0.7)	12.6(0.9)	13.6(1.6)	19.8(1.7)	3.5(1.2)	6.2(2.3)
9,10	8.4(0.6)	11.9(0.8)	8.7(1.0)	13.0(1.2)	3.5(1.0)	4.3(1.2)
16,14'	15.5(0.7)	15.0(0.6)	13.8(0.8)	15.4(0.7)	-0.5(0.9)	1.5(1.0)
17,15'	5.9(0.7)	7.3(0.4)	8.6(0.8)	9.8(0.9)	1.4(0.8)	1.1(1.2)
*[4-7,12-15, 4'-13']	15.8(0.3)	20.5(0.9)	17.8(0.7)	24.1(1.1)	4.7(0.9)	6.4(1.3)
*1[6'-9']	24.8(0.2)	27.2(1.5)	25.4(0.5)	30.3(0.3)	2.5(1.6)	4.9(0.6)
*2[7,10',11']	15.1(0.2)	16.4(0.9)	16.0(0.7)	23.2(1.0)	1.3(0.9)	7.2(1.2)
*3[4-6,12-15,4',5',12',13']	11.7(0.3)	18.8(0.9)	15.8(1.2)	23.1(1.8)	7.1(1.0)	7.4(2.2)

^a Each ^{13}C transverse relaxation rate (R_2) was determined from best-fitting the integrated NMR peak intensity S vs. delay time τ using $S(\tau) = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. The fitting uncertainty of R_2 is given in parentheses. The Γ_2 values are the differences between the best-fit R_2 values of samples with vs. without Mn^{2+} . Typical ppm integration ranges for peaks are: 2,2', 33.00–37.00; 3,3', 24.00–26.30; 8,11, 26.50–28.20; 9,10, 128.00–131.00; 16,14', 31.50–33.00; 17,15', 21.50–23.60; *, 28.30–31.50; *1, 30.24–31.50; *2, 30.04–30.24; *3, 28.30–30.04 (Figure B2). The ^{13}C sites that make the largest contributions to the *1, *2, and *3 integration ranges are listed between the brackets. The % $\text{Mn}^{2+} = (\text{mole bound Mn}^{2+})/(\text{mole lipid}) \times 100$. The % Fp is calculated using the same type of expression.

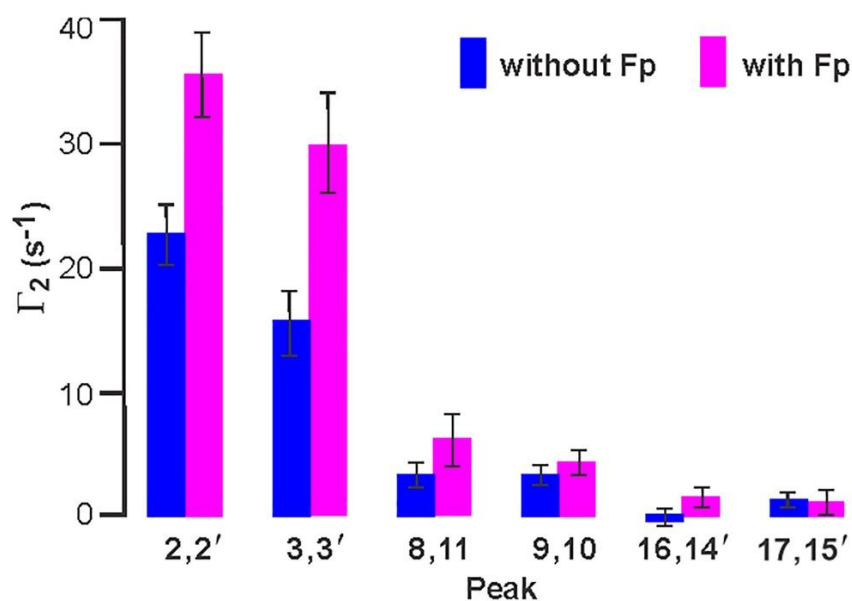


Figure 4.6 Bar plot of the Γ_2 's, i.e. the differences between the R_2 's for samples with vs. without Mn^{2+} . The Γ_2 's are displayed for several peaks. Each peak is due to signals from two ^{13}C sites in the acyl chains. The Γ_2 's are shown for samples without vs. with Fp. The Γ_2 's and their uncertainties are presented numerically in Table 4.2.

Table 4.3. Site-specific ^{13}C FWHM NMR linewidths of acyl chains of POPC:POPG (4:1) membrane and Mn^{2+} and Fp dependences^a.

^{13}C	Δv (Hz)				$\Delta [\Delta v]$ (Hz)	
	w/o Fp		3% Fp		w/o Fp	3% Fp
	w/o Mn^{2+}	0.5% Mn^{2+}	w/o Mn^{2+}	0.5% Mn^{2+}		
2,2'	33.1	43.2	34.8	49.4	10.1	14.6
3,3'	37.3	43.1	38.7	47.8	5.8	9.1
8,11	39.7	42.8	38.8	44.1	3.1	5.3
9,10	33.0	39.4	33.7	36.4	6.4	2.7
16,14'	30.0	34.7	30.0	31.1	4.7	1.1
17,15'	27.5	32.9	28.1	28.9	5.4	0.8
*{4-7,12-15,4'-13'}	110.7	112.4	111.7	116.0	1.7	4.3

^a These are experimental full-width at half-maximum linewidths of the ^{13}C signal of the CP NMR spectrum. The $\Delta[\Delta v]$ is the difference in linewidths between samples with vs. without Mn^{2+} .

Table 4.4. Inhomogeneous contributions to site-specific ^{13}C FWHM NMR linewidths of acyl chains of POPC:POPG (4:1) membrane^a.

^{13}C	Δv_{inhom} (Hz)				Average
	w/o Fp		3% Fp		
	w/o Mn^{2+}	0.5% Mn^{2+}	w/o Mn^{2+}	0.5% Mn^{2+}	
2,2'	23.9	26.8	25.9	29.2	26.4(1.9)
3,3'	30.9	31.9	31.9	31.5	31.6(0.4)
8,11	36.8	38.8	34.5	37.8	37.0(1.6)
9,10	30.3	35.6	30.9	32.6	31.6(2.0)
16,14'	25.1	29.9	25.6	26.2	27.0(1.9)
17,15'	25.6	30.6	25.4	25.8	26.8(2.2)

^a The $\Delta v_{\text{inhom}} = \Delta v_{\text{exp}} - R_2/\pi$ i.e. the difference between the experimental full-width at half-maximum linewidth and the experimental relaxation contribution to the linewidth. The average is for the four samples with the standard deviation in parentheses.

4.3 Discussion

4.3.1 Hydration and location of Mn^{2+} in samples

The lipids in the samples were likely close to full hydration. Water was added to the rotor both before and after the lipid is added, i.e. the water was initially both underneath and on top of the lipid. The rotor was then sealed with the top cap followed by >12 hours prior to acquisition of NMR spectra, and this incubation time was intended to promote water permeation and homogenous hydration of the lipid. Such hydration is evidenced by the: (1) wet appearance of the sample, including after the NMR experiments were completed; and (2) ^{13}C symmetric lineshapes with narrow (~ 0.3 ppm FWHM) linewidths. We also anticipate little evaporation of water because the rotor was sealed, the MAS frequency was moderate (8 kHz), and the ^1H rf circuit of the NMR probe was designed to minimize dielectric heating of the water. For a typical sample, ~ 20 μL water was added to the rotor which has ~ 40 μL total volume rotor, so there is ~ 20 μL lipid in the sample. POPC and water have similar densities, so the lipid: water mass ratio was similar to the volume ratio. By this approach, the lipid: water mass ratio in our samples was comparable to the $\sim 3:2$ ratio for fully-hydrated lipid, which was calculated using 28 water molecules-per-lipid molecule.⁷⁰

The Mn^{2+} likely binds the lipids rather than Fp. As noted in the Results section, membrane with only POPC lipid (with zwitterionic headgroup) quantitatively binds Mn^{2+} whereas POPC

membrane with bound Fp doesn't bind Mn^{2+} . This was the reason that the NMR samples were prepared with 20 mole% POPG lipids with anionic headgroup, with resulting Mn^{2+} binding when peptide was also bound. In addition, the experimentally-determined pK_a 's of the sidechains of the two Glu and one Asp sidechains are > 5 . For the NMR samples at pH 5.0, the Glu and Asp sidechains have only partial negative charge, whereas the lipid headgroups have full negative charge.⁷¹

4.3.2 NMR relaxation data support increased probability of lipid chain protrusion with Fp

There are overall larger Γ_2 's for lipid with vs. without Fp, and the magnitude is most pronounced for the 2,2' and 3,3' signals, with reductions in R_2 and Γ_2 values for $-CH_2$ sites closer to the chain terminus (Figures. 4.4-4.6, Table 4.2). The Γ_2 trend is consistent with the results of the initial molecular dynamics simulation showing increased protrusion probability with Fp and specifically with the result summarized on p. 3 of this article "...our simulations predict the [Fp] effect on tail protrusion to be most profound in the upper region of the acyl chain ..."⁴⁹ (Figure 4.1).

Earlier NMR relaxation data were used to estimate a correlation time of $\sim 10^{-8}$ s for the lipid chain motion that could lead to protrusion in the liquid-crystalline membrane.^{72,73} This time is much smaller than the characteristic $(1/R_2) \approx 10^{-1}$ s for ^{13}C transverse relaxation so the R_2 is most reasonably considered as a weighted average of the chain's high-probability unprotruded state and low-probability protruded state. We hypothesize that a site's Γ_2 is proportional to the probability of chain protrusion (Ω_{prot}) into the headgroup region.⁶¹ This hypothesis is based on: (1) the r^{-6} dependence of Γ_2 where r is the ^{13}C - Mn^{2+} distance; and on (2) smaller r and therefore much larger Γ_2 when the chain is protruded. As described in the Results section, only 0.005 fraction of the lipid headgroups in our samples have bound Mn^{2+} . The Mn^{2+} are likely also exchanging rapidly between headgroups during ^{13}C transverse relaxation. Based on arguments similar to those above for protrusion, we additionally hypothesize that a site's Γ_2 is also proportional to the probability that a Mn^{2+} is bound to a headgroup close to the protruded chain (Ω_{Mn}).

We estimate r for the protruded state with nearby Mn^{2+} by combining our two hypotheses with the known expression for Γ_2 with a nuclear spin I with spin $1/2$ and a nearby paramagnetic species with spin quantum number L :

$$r \approx \{ (1/15) \times (\mu_0/4\pi)^2 \times \gamma_I^2 \times g_e^2 \times \mu_B^2 \times L(L+1) \times (4J_0 + 3J_{\omega I}) \times \Omega_{prot} \times \Omega_{Mn} \} / \Gamma_2^{1/6} \quad (4.1)$$

where μ_0 is the permeability of free space, γ_I is the nuclear spin gyromagnetic ratio, g_e is the electron spin g factor, μ_B is the Bohr magneton, J_ω is the spectral density at angular frequency ω , and ω_I is the angular NMR Larmor frequency of the nucleus.⁵⁹ Using the known values of μ_0 , γ_I for ^{13}C , g_e , and μ_B , and $L = 5/2$ for Mn^{2+} ,

$$r \approx \{ [(9.106 \times 10^{-45} \text{ m}^6\text{-s}^{-2}) \times (4J_0 + 3J_{\omega_I}) \times \Omega_{\text{prot}} \times \Omega_{\text{Mn}}] / \Gamma_2^{1/6} \} \quad (4.2)$$

There are estimates below for the other terms but we note that because of the 6th root dependence, r is fairly insensitive to moderate changes in these estimates, e.g. a 10 \times increase in the expression in braces correlates with a 1.1 \times increase in r . The J_ω is calculated using:

$$J_\omega = \tau_c / [1 + (\tau_c^2 \times \omega^2)] \quad (4.3)$$

where τ_c is the correlation time.⁶¹ The $J_0 = \tau_c$ which is much larger than J_{ω_I} , based on $\tau_c \approx 10^{-8}$ s, the experimentally-based estimate of the correlation time for chain motion associated with protrusion and on $\omega_I \approx 6 \times 10^8 \text{ s}^{-1}$.^{76,77} We estimate $\Omega_{\text{Mn}} \approx 10^{-2}$ based on a protruded ^{13}C being near ~ 2 lipid headgroups and the experimental Mn^{2+} headgroup occupancy ≈ 0.005 . For lipid without Fp, $\Omega_{\text{prot}} \approx 0.01$ in simulations and the $\Gamma_2 \approx 10 \text{ s}^{-1}$ from our experimental data (Table 4.2). The resulting calculated $r \approx 4 \text{ \AA}$ is plausible for the ^{13}C nuclei of a chain that protrudes into the headgroup region of the membrane.

One notable trend of Figure 4.6 and Table 4.2 is the attenuation of Γ_2 as the ^{13}C site moves closer to the chain terminus. This trend holds for samples both without and with Fp. We have hypothesized that $\Gamma_2 \propto \Omega_{\text{prot}}$ and further hypothesize that protrusion of a specific $-\text{CH}_2$ group also means protrusion of the $-\text{CH}_2$ groups that are closer to the lipid glycerol group. This second hypothesis is based on the location of the glycerol group close to the phosphate group and the chemical bonding of the acyl chain (Figure 2.3). For the six ^{13}C NMR signals assigned to two $-\text{CH}_2$ sites with sites numbered x and y , respectively (Figure 4.3), the average number of protruded $-\text{CH}_2$ groups(n) is calculated:

$$n = [(x+y)/2] - 1 \quad (4.4)$$

and protrusion of each $-\text{CH}_2$ group is hypothesized to require free energy ΔG_{prot} so that Γ_2 depends on n as:

$$\Gamma_2(n) = \Gamma_2(0) \times \exp[-(n \times \Delta G_{\text{prot}}) / k_B T] \quad (4.5)$$

where $\Gamma_2(0)$ is the parameter for the rate when no $-\text{CH}_2$ are protruded. Figure 4.7 displays fitting with Equation 4.5 of the experimental $\Gamma_2(n)$ data vs. n , with separate fittings of data without and

with Fp. The best-fit $\Delta G_{\text{prot}}/k_B T$ are 0.249 ± 0.018 without Fp and 0.266 ± 0.016 with Fp. The Equation 4.5 model is supported both by the quantitative similarity of the two values and the semi-quantitative agreement with the 0.25-0.50 $k_B T$ range for $\Delta G_{\text{prot}}/k_B T$ in one of the simulations.⁴⁹ The best-fit $\Gamma_2(0)$ are $27.9 \pm 1.8 \text{ s}^{-1}$ without Fp and $47.3 \pm 2.6 \text{ s}^{-1}$ with Fp.

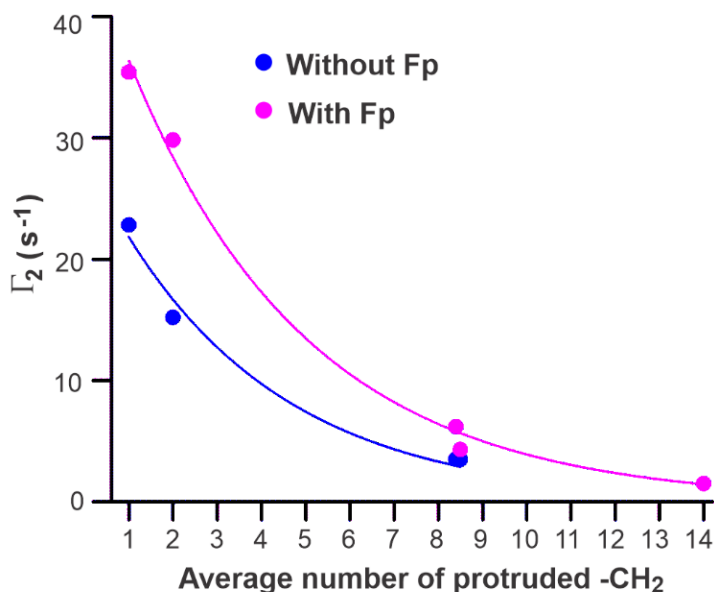


Figure 4.7. Plots and exponential decay fittings of Γ_2 vs. average number (n) of protruded $-\text{CH}_2$. For each peak corresponding to signals from two $-\text{CH}_2$ sites that are numbered x and y (Figs. 3 and 5), this average number is calculated as $[(x+y)/2] - 1$, e.g. 2 for the 3,3' peak and 14 for the 16,14' peak. Data are displayed for samples without and with Fp. Separate fittings are done for each sample type using $\Gamma_2(n) = \Gamma_2(0) \times \exp(-n \times \kappa)$ with $\Gamma_2(0)$ and κ as fitting parameters and $\kappa = \Delta G_{\text{prot}}/k_B T$. Best-fit values with uncertainties in parentheses are: (1) without Fp, $\Gamma_2(0) = 27.9(1.8) \text{ s}^{-1}$ and $\kappa = 0.249(18)$; and (2) with Fp, $\Gamma_2(0) = 47.3(2.6) \text{ s}^{-1}$ and $\kappa = 0.266(16)$.

The molecular dynamics simulations from different groups show that the large enhancement of Ω_{prot} is primarily for lipids next to the Fp whereas the Ω_{prot} of more distant lipids is similar to lipids without Fp. For samples with Fp, the $\sim 10^{-8} \text{ s}$ time for lateral diffusion of a lipid molecule between Fp-neighboring and more distant locations is much more rapid than the $1/R_2$ relaxation time, so the $\Gamma_{2,\text{Fp}}$ will be a weighted average of the larger Fp-adjacent and smaller more distant values, $\Gamma_{2,\text{neighbor}}$ and $\Gamma_{2,\text{distant}}$, respectively. For “q” lipid molecules neighboring a Fp and the ~ 3 mole% Fp of a sample:

$$\Gamma_{2,\text{Fp}} = (0.03 \times q \times \Gamma_{2,\text{neighbor}}) + [1 - (0.03 \times q)] \times \Gamma_{2,\text{distant}} \quad (4.6)$$

which is algebraically rewritten as:

$$\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}} = [\Gamma_{2,\text{Fp}} / \Gamma_{2,\text{distant}} - 1 + (0.03 \times q)] / (0.03 \times q) \quad (4.7)$$

Using the best-fit $\Gamma_{2,\text{Fp}}(0)$ and $\Gamma_{2,\text{noFp}}(0)$ values and the approximation $\Gamma_{2,\text{distant}}(0) \approx \Gamma_{2,\text{noFp}}(0)$:

$$\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}} = [\Gamma_{2,\text{Fp}}(0) / \Gamma_{2,\text{noFp}}(0) - 1 + (0.03 \times q)] / (0.03 \times q) \quad (4.8)$$

The $\Gamma_{2,\text{Fp}}(0) / \Gamma_{2,\text{noFp}}(0) = 1.70 \pm 0.14$. The value of q will depend on the Fp surface area that contacts lipids and this area will vary with location and orientation of the Fp in the membrane. There is also the possibility that q is reduced because of increased protrusion probability is mostly for a subset of neighboring lipids with spatially-specific Fp interactions. A reasonable possible range of q values is 1 to 8, with $q = 8$ based on interfacial Fp location and $\sim 4 \times$ greater cross-sectional area for Fp vs. lipid.^{47,48} From Equation 4.8, $\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}} = 24.3 \pm 4.8$ for $q = 1$ and 3.92 ± 0.60 for $q = 8$. This matches the range of $\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}}$ ratios that were observed in different simulations.⁴⁹⁻⁵¹ The inverse correlation between the experimentally-derived $\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}}$ ratio and q is consistent with the larger simulation-derived $\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}}$ ratios for transmembrane vs. membrane surface location of the Fp, and the likely larger Fp lipid-contacting area and q value for the surface location.^{49,51} In addition, large $\Gamma_{2,\text{neighbor}} / \Gamma_{2,\text{distant}}$ is observed in a different simulation in which the effective $q \approx 1$ because of the strong correlation between protrusion and a hydrogen bond between one of the four N-terminal residues of the Fp and a lipid phosphate oxygen, with consequent headgroup intrusion into the membrane interior.⁵⁰

Besides the aforementioned simulation observations that increased chain protrusion may be associated with Fp/phosphate hydrogen bonding and headgroup intrusion, protrusion may also be augmented by solvation of lipid chains by hydrophobic Fp sidechains at the membrane surface. This could be part of the basis for much greater Ha2-induced intervesicle lipid mixing at endosomal pH 5.0 vs. physiologic pH 7.4.¹⁵ This effect is observed with POPC vesicles for which there isn't bulk Ha2/vesicle electrostatic energy. At both pH's, Fp is a mixture of closed and semi-closed helical hairpin structures, and semi-closed has significantly greater hydrophobic surface area.⁴⁸ There is higher semi-closed population at pH 5 vs. 7 and therefore larger Fp hydrophobic surface area.

4.3.3 Linewidth and * peak analysis also support protrusion

For the 2,2' and 3,3' signals, there are larger Mn^{2+} -associated increases in linewidth, $\Delta[\Delta\nu]$, for

samples with vs. without Fp (Table 4.3). The difference is less apparent for two-site signals from ^{13}C nuclei closer to the chain termini, and these observations correlate with the trend of $\Gamma_{2,\text{Fp}} - \Gamma_{2,\text{NoFp}}$ (Figure 4.6 and Table 4.2). For the two-site signals, Table 4.4 provides the inhomogeneous contributions to linewidths, which are calculated by $\Delta\nu_{\text{inhom}} = \Delta\nu - (R_2 / \pi)$. Table 4.4 also provides for each signal the RMSD and average value calculated from the data of the four samples. The typical RMSD is ± 2 Hz and the average values are in the 26-37 Hz range. For our spectra, the individual contributions to the signal from each site are typically unresolved, except for 9,10 which has partially-resolved C9 and C10 contributions with respective higher and lower shifts. For earlier ^{13}C NMR spectra of POPC with somewhat narrower linewidths than our spectra, there aren't resolved shift differences ($\Delta\delta$) between sites for the 2,2', 16,14', and 17,15' signals, and the $\Delta\delta$ are $\sim 0.09, 0.13,$ and 0.36 ppm for the 3,3', 8,11, and 9,10 signals, respectively.⁶² For the present study, the 2,2', 16,14', and 17,15' signals have $\Delta\nu_{\text{inhom}} \approx 27$ Hz which is likely due to 20 Hz exponential line broadening and shimming. There are larger $\Delta\nu_{\text{inhom}}$ of ~ 32 and 37 Hz for the 3,3' and 8,11 signals, respectively, and the ~ 5 and 10 Hz increases over the $\Delta\nu_{\text{inhom}} \approx 27$ Hz baseline value correlate semi-quantitatively with the $\Delta\delta$ values of ~ 9 and ~ 13 Hz, respectively. For the 9,10 signal, the $\Delta\nu_{\text{inhom}} \approx 32$ Hz is smaller than would be expected from the $\Delta\delta \approx 36$ Hz. This anomaly is likely a consequence of the larger C9 vs. C10 contribution to the 9,10 signal. There is greater ^1H - ^{13}C cross-polarization for C9 vs. C10 because of the smaller vs. larger site mobility that was previously described by differences in site order parameters.⁶²

The * peak is a superposition of signals from eighteen ^{13}C sites in the middle of the two chains, 4-7, 12-15, and 4'-13'. Table 4.2 lists the best-fit R_2 and Γ_2 values both for full integration of * intensities and for integration ranges denoted *1, *2, and *3 for which a subset of the ^{13}C sites make the largest contributions, respectively 6'-9'; 7,10',11'; and 4-6,12-15,4',5',12',13'.⁶² The Γ_2 values from the * fittings are similar to those of 8,11 and 9,10, and are intermediate between the larger 2,2' and 3,3' Γ_2 values and smaller 16,14' and 17,15' values. The *1 and *2 integrations are dominated by signals from ^{13}C sites closer to CO whereas the *3 integration has large contributions from ^{13}C sites closer to the chain terminus. These different locations correlate with the $\Gamma_{2,\text{Fp}} > \Gamma_{2,\text{NoFp}}$ for *1 and *2 and for 2,2'; 3,3'; 8,11; and 9,10 fittings, whereas $\Gamma_{2,\text{Fp}} \approx \Gamma_{2,\text{NoFp}}$ for *3; 16,14'; and 17,15' fittings. The Mn^{2+} -associated increases of the linewidths of the * peaks are similar to those of resolved ^{13}C sites in the middle and terminal regions of the chains, and the increase for *

is a little larger with vs. without Fp (Table 4.3). The * linewidth has substantial inhomogeneous contribution from the superposition of unresolved signals from many ^{13}C sites.

4.3.4 Comparison between PRE and other experimental approaches to probe chain motions relevant to fusion

The hypothesis of increased protrusion induced by Fp was initially proposed based on molecular dynamics simulations.^{49–51} The typical $\Omega_{\text{prot,NoFp}} \approx 0.01$ and the $\Omega_{\text{prot,Fp}}$ was larger but still < 0.15 , i.e. protrusion was always a low-probability state. It is therefore anticipated that the observables from some experimental approaches have large contributions from the high-probability unprotruded chains. The Mn^{2+} paramagnetic relaxation enhancement (PRE) approach of the present study has the advantage that the Γ_2 observable is dominated by the small population of protruded chains, based on: (1) Mn^{2+} is predominantly bound to the lipid headgroup phosphate; (2) the $\langle r^{-6} \rangle$ dependence of Γ_2 ; and (3) the larger τ_c for protrusion vs. other chain motions.⁷³ Some other approaches wouldn't have this advantage. As one example, protrusion would also affect lipid ^{13}C - ^{31}P dipolar coupling that can be measured by NMR.²⁶ The coupling is proportional to $\langle \rho^{-3} \rangle$ where ρ is the internuclear distance. For the 2,2' and 3,3' sites, the $\langle \rho^{-3} \rangle$ for membrane without vs. with Fp are estimated to be $\sim 2.01 \times 10^{-3}$ vs. $\sim 2.26 \times 10^{-3} \text{ \AA}^{-3}$, i.e. only $\sim 12\%$ increase, as calculated from $\rho_{\text{C-P,unprot}} \approx 8 \text{ \AA}$, $\rho_{\text{C-P,unprot}} \approx 5 \text{ \AA}$, $\Omega_{\text{prot,NoFp}} \approx 0.01$, and $\Omega_{\text{prot,Fp}} \approx 0.05$. Another consideration is that investigation of protrusion by simulation has been done in fluid rather than gel membrane phases, in part because fluid phases are similar to those of membranes in viral fusion. Lipids in fluid phases experience rapid lateral diffusion and also other large-amplitude motions.^{64,73} These motions are usually advantageous for the PRE approach because they result in smaller R_2 's that can generally be measured more accurately than larger R_2 's. The motions also reduce dipolar couplings; however, the NMR spectra often have lower signal-to-noise when measuring smaller vs. larger couplings, so smaller couplings are less accurately-determined.⁷⁴

Splay is the term used to describe the large-amplitude movement by the terminal region of the lipid chain into the headgroup region. Splay may be relevant to fusion and has been detected using the ^1H - ^1H NOESY NMR cross-relaxation rate between the terminal methyl and the headgroup nuclei.⁷⁵ The NOESY rates have been positively-correlated with the extents of fusion between vesicles with transmembrane peptides.⁷⁶ However, these vesicle fusion rates are $\sim 3 \times 10^{-4} \text{ s}^{-1}$ which are $\sim 1000 \times$ smaller than rates of vesicle fusion induced by Fp's.⁴⁸ These results suggest that splay is a less important motion for fusion than the protrusion of chain $-\text{CH}_2$ groups closest to the

glycerol linkage, i.e. the motion probed in the present study. This conclusion is supported by the statement on p. 5 of Ref. 49, "...our simulations predict the effect on tail protrusion to be most profound in the upper region of the acyl chain, so a difference in tail exposure might be suboptimal probe for fusion peptide activity."⁴⁹

4.4 Conclusions

The present study presents convincing experimental data that support a large increase in lipid acyl chain protrusion caused by the membrane-bound Fp domain of the influenza virus Ha2 protein. Increased protrusion had previously been observed in computational simulations and may play an important role in fusion between the viral and the endosome membranes. In particular, protrusion may accelerate the transition from the initial separate apposed membranes to the stalk intermediate that connects and is contiguous with the outer leaflets of the two bodies. For the present study, protrusion was detected by larger Mn^{2+} -associated increases in transverse relaxation rates of lipid chain ^{13}C nuclei for samples with vs. without Fp. Analysis of the $\Gamma_{2,\text{Fp}}$ vs. $\Gamma_{2,\text{NoFp}}$ rate increases resulted in a calculated ratio $\Gamma_{2,\text{neighbor}}/\Gamma_{2,\text{distant}}$ in the range of 4-24 where the ratio is for lipids neighboring vs. more distance from Fp. The ratio values within this range are inversely-correlated with the number of neighboring lipids. The experimental range is similar to the range in simulations for increased protrusion probability of a lipid neighboring vs. more distant from the Fp. For samples either with or without Fp, the Γ_2 values are well-fitted by an exponential decay as the ^{13}C site moves closer to the chain terminus. The decay correlates with a positive free-energy of protrusion that is proportional to the number of protruded $-\text{CH}_2$ groups. The experimentally-determined free energy per $-\text{CH}_2$ is ~ 0.25 k_BT which matches the value in one of the simulations. Overall, the NMR data support one major fusion role of the Fp to be much greater chain protrusion with highest probability for chain regions closest to the headgroups.

REFERENCES

- (1) White, J. M.; Delos, S. E.; Brecher, M.; Schornberg, K. Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Crit Rev Biochem Mol Biol* 2008, *43* (3), 189–219. DOI: 10.1080/10409230802058320.
- (2) Kielian, M. Mechanisms of Virus Membrane Fusion Proteins. *Annu Rev Virol* 2014, *1* (1), 171–189. DOI: 10.1146/annurev-virology-031413-085521.
- (3) Harrison, S. C. Viral Membrane Fusion. *Virology* 2015, *479–480*, 498–507. DOI: 10.1016/j.virol.2015.03.043.
- (4) Boonstra, S.; Blijleven, J. S.; Roos, W. H.; Onck, P. R.; Van Der Giessen, E.; Van Oijen, A. M. Hemagglutinin-Mediated Membrane Fusion: A Biophysical Perspective. *Annu Rev Biophys* 2018, *47*, 153–173. DOI: 10.1146/annurev-biophys-070317-033018.
- (5) Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S. Coronavirus Membrane Fusion Mechanism Offers a Potential Target for Antiviral Development. *Antiviral Res* 2020, *178* (March), 104792. DOI: 10.1016/j.antiviral.2020.104792.
- (6) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the Haemagglutinin Membrane Glycoprotein of Influenza Virus at 3 Å Resolution. *Nature* 1981, *289*, 366–373.
- (7) Pancera, M.; Zhou, T.; Druz, A.; Georgiev, I. S.; Soto, C.; Gorman, J.; Huang, J.; Acharya, P.; Chuang, G. Y.; Ofek, G.; Stewart-Jones, G. B. E.; Stuckey, J.; Bailer, R. T.; Joyce, M. G.; Louder, M. K.; Tumba, N.; Yang, Y.; Zhang, B.; Cohen, M. S.; Haynes, B. F.; Mascola, J. R.; Morris, L.; Munro, J. B.; Blanchard, S. C.; Mothes, W.; Connors, M.; Kwong, P. D. Structure and Immune Recognition of Trimeric Pre-Fusion HIV-1 Env. *Nature* 2014, *514* (7253), 455–461. DOI: 10.1038/nature13808.
- (8) Ward, A. B.; Wilson, I. A. The HIV-1 Envelope Glycoprotein Structure: Nailing down a Moving Target. *Immunol Rev* 2017, *275* (1), 21–32. DOI: 10.1111/imr.12507.
- (9) Kirchdoerfer, R. N.; Cottrell, C. A.; Wang, N.; Pallesen, J.; Yassine, H. M.; Turner, H. L.; Corbett, K. S.; Graham, B. S.; McLellan, J. S.; Ward, A. B. Pre-Fusion Structure of a Human Coronavirus Spike Protein. *Nature* 2016, *531* (7592), 118–121. DOI: 10.1038/nature17200.
- (10) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* (1979) 2020, *367* (6483), 1260–1263. DOI: 10.1126/science.aax0902.
- (11) Chen, J.; Skehel, J. J.; Wiley, D. C. A Polar Octapeptide Fused to the N-Terminal Fusion Peptide Solubilizes the Influenza Virus HA2 Subunit Ectodomain. *Biochemistry* 1998, *37* (39), 13643–13649. DOI: 10.1021/bi981098l.
- (12) Lev, N.; Fridmann-Sirkis, Y.; Blank, L.; Bitler, A.; Epand, R. F.; Epand, R. M.; Shai, Y. Conformational Stability and Membrane Interaction of the Full-Length Ectodomain of HIV-1 Gp41: Implication for Mode of Action. *Biochemistry* 2009, *48* (14), 3166–3175. DOI: 10.1021/bi802243j.

- (13) Sackett, K.; Nethercott, M. J.; Epand, R. F.; Epand, R. M.; Kindra, D. R.; Shai, Y.; Weliky, D. P. Comparative Analysis of Membrane-Associated Fusion Peptide Secondary Structure and Lipid Mixing Function of HIV Gp41 Constructs That Model the Early Pre-Hairpin Intermediate and Final Hairpin Conformations. *J Mol Biol* 2010, 397 (1), 301–315. DOI: 10.1016/j.jmb.2010.01.018.
- (14) Aydin, H.; Al-Khooly, D.; Lee, J. E. Influence of Hydrophobic and Electrostatic Residues on SARS-Coronavirus S2 Protein Stability: Insights into Mechanisms of General Viral Fusion and Inhibitor Design. *Protein Science* 2014, 23 (5), 603–617. DOI: 10.1002/pro.2442.
- (15) Ranaweera, A.; Ratnayake, P. U.; Weliky, D. P. The Stabilities of the Soluble Ectodomain and Fusion Peptide Hairpins of the Influenza Virus Hemagglutinin Subunit II Protein Are Positively Correlated with Membrane Fusion. *Biochemistry* 2018, 57 (37), 5480–5493. DOI: 10.1021/acs.biochem.8b00764.
- (16) Caffrey, M.; Cai, M.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Covell, D. G.; Gronenborn, A. M.; Clore, G. M. Three-Dimensional Solution Structure of the 44 KDa Ectodomain of SIV Gp41. *EMBO Journal* 1998, 17 (16), 4572–4584. DOI: 10.1093/emboj/17.16.4572.
- (17) Yang, Z. N.; Mueser, T. C.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Hyde, C. C. The Crystal Structure of the SIV Gp41 Ectodomain at 1.47 Å Resolution. *J Struct Biol* 1999, 126 (2), 131–144. DOI: 10.1006/jsbi.1999.4116.
- (18) Chen, J.; Skehel, J. J.; Wiley, D. C. N- and C-Terminal Residues Combine in the Fusion-PH Influenza Hemagglutinin HA2 Subunit to Form an N Cap That Terminates the Triple-Stranded Coiled Coil. *Proc Natl Acad Sci U S A* 1999, 96 (16), 8967–8972. DOI: 10.1073/pnas.96.16.8967.
- (19) Duquerroy, S.; Vigouroux, A.; Rottier, P. J. M.; Rey, F. A.; Jan Bosch, B. Central Ions and Lateral Asparagine/Glutamine Zippers Stabilize the Post-Fusion Hairpin Conformation of the SARS Coronavirus Spike Glycoprotein. *Virology* 2005, 335 (2), 276–285. DOI: 10.1016/j.virol.2005.02.022.
- (20) Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Velesler, D. Tectonic Conformational Changes of a Coronavirus Spike Glycoprotein Promote Membrane Fusion. *Proc Natl Acad Sci U S A* 2017, 114 (42), 11157–11162. DOI: 10.1073/pnas.1708727114.
- (21) Chernomordik, L. V.; Kozlov, M. M. Mechanics of Membrane Fusion. *Nat Struct Mol Biol* 2008, 15 (7), 675–683. DOI: 10.1038/nsmb.1455.
- (22) Blumenthal, R.; Durell, S.; Viard, M. HIV Entry and Envelope Glycoprotein-Mediated Fusion. *Journal of Biological Chemistry* 2012, 287 (49), 40841–40849. DOI: 10.1074/jbc.R112.406272.
- (23) Durrer, P.; Galli, C.; Hoenke, S.; Corti, C.; Glück, R.; Vorherr, T.; Brunner, J. H⁺-Induced Membrane Insertion of Influenza Virus Hemagglutinin Involves the HA2 Amino-Terminal Fusion Peptide but Not the Coiled Coil Region. *Journal of Biological Chemistry* 1996, 271 (23), 13417–13421. DOI: 10.1074/jbc.271.23.13417.

- (24) Durell, S. R.; Martin, I.; Ruysschaert, J. M.; Shai, Y.; Blumenthal, R. What Studies of Fusion Peptides Tell Us about Viral Envelope Glycoprotein-Mediated Membrane Fusion. *Mol Membr Biol* 1997, 14 (3), 97–112. DOI: 10.3109/09687689709048170.
- (25) Han, X.; Tamm, L. K. A Host-Guest System to Study Structure-Function Relationships of Membrane Fusion Peptides. *Proc Natl Acad Sci U S A* 2000, 97 (24), 13097–13102. DOI: 10.1073/pnas.230212097.
- (26) Jia, L.; Liang, S.; Sackett, K.; Xie, L.; Ghosh, U.; Weliky, D. P. REDOR Solid-State NMR as a Probe of the Membrane Locations of Membrane-Associated Peptides and Proteins. *Journal of Magnetic Resonance* 2015, 253, 154–165. DOI: 10.1016/j.jmr.2014.12.020
- (27) Epand, R. M. Fusion Peptides and the Mechanism of Viral Fusion. *Biochimica et Biophysica Acta - Biomembranes*. Elsevier July 11, 2003, pp 116–121. DOI: 10.1016/S0005-2736(03)00169-X.
- (28) Ge, M.; Freed, J. H. Fusion Peptide from Influenza Hemagglutinin Increases Membrane Surface Order: An Electron-Spin Resonance Study. *Biophys J* 2009, 96 (12), 4925–4934. DOI: 10.1016/j.bpj.2009.04.015.
- (29) Gabrys, C. M.; Yang, R.; Wasniewski, C. M.; Yang, J.; Canlas, C. G.; Qiang, W.; Sun, Y.; Weliky, D. P. Nuclear Magnetic Resonance Evidence for Retention of a Lamellar Membrane Phase with Curvature in the Presence of Large Quantities of the HIV Fusion Peptide. *Biochim Biophys Acta Biomembr* 2010, 1798 (2), 194–201. DOI: 10.1016/j.bbamem.2009.07.007.
- (30) Tristram-Nagle Stephanie, S.; Chan, R.; Kooijman, E.; Uppamoochikkal, P.; Qiang, W.; Weliky, D. P.; Nagle, J. F. HIV Fusion Peptide Penetrates, Disorders, and Softens T-Cell Membrane Mimics. *J Mol Biol* 2010, 402 (1), 139–153. DOI: 10.1016/j.jmb.2010.07.026.
- (31) Yao, H.; Hong, M. Membrane-Dependent Conformation, Dynamics, and Lipid Interactions of the Fusion Peptide of the Paramyxovirus Piv5 from Solid-State NMR. *J Mol Biol* 2013, 425 (3), 563–576. DOI: 10.1016/j.jmb.2012.11.027.
- (32) Smrt, S. T.; Draney, A. W.; Lorieau, J. L. The Influenza Hemagglutinin Fusion Domain Is an Amphipathic Helical Hairpin That Functions by Inducing Membrane Curvature. *Journal of Biological Chemistry* 2015, 290 (1), 228–238. DOI: 10.1074/jbc.M114.611657.
- (33) Chakraborty, H.; Lentz, B. R.; Kombrabail, M.; Krishnamoorthy, G.; Chattopadhyay, A. Depth-Dependent Membrane Ordering by Hemagglutinin Fusion Peptide Promotes Fusion. *Journal of Physical Chemistry B* 2017, 121 (7), 1640–1648. DOI: 10.1021/acs.jpcc.7b00684.
- (34) Lai, A. L.; Freed, J. H. HIV Gp41 Fusion Peptide Increases Membrane Ordering in a Cholesterol-Dependent Fashion. *Biophys J* 2014, 106 (1), 172–181. DOI: 10.1016/j.bpj.2013.11.027.
- (35) Lai, A. L.; Millet, J. K.; Daniel, S.; Freed, J. H.; Whittaker, G. R. The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform That Perturbs Membrane Order in a

- Calcium-Dependent Manner. *J Mol Biol* 2017, 429 (24), 3875–3892. DOI: 10.1016/j.jmb.2017.10.017.
- (36) Heller, W. T.; Zolnierczuk, P. A. The Helix-to-Sheet Transition of an HIV-1 Fusion Peptide Derivative Changes the Mechanical Properties of Lipid Bilayer Membranes. *Biochim Biophys Acta Biomembr* 2019, 1861 (3), 565–572. DOI: 10.1016/j.bbamem.2018.12.004.
- (37) Ghosh, U.; Weliky, D. P. 2H Nuclear Magnetic Resonance Spectroscopy Supports Larger Amplitude Fast Motion and Interference with Lipid Chain Ordering for Membrane That Contains β Sheet Human Immunodeficiency Virus Gp41 Fusion Peptide or Helical Hairpin Influenza Virus Hemagglutini. *Biochim Biophys Acta Biomembr* 2020, 1862 (10), 183404. DOI: 10.1016/j.bbamem.2020.183404.
- (38) Ghosh, U.; Weliky, D. P. Rapid 2H NMR Transverse Relaxation of Perdeuterated Lipid Acyl Chains of Membrane with Bound Viral Fusion Peptide Supports Large-Amplitude Motions of These Chains That Can Catalyze Membrane Fusion. *Biochemistry* 2021, 60 (35), 2637–2651. DOI: 10.1021/acs.biochem.1c00316.
- (39) Freed, E. O.; Delwart, E. L.; Buchschacher, G. L.; Panganiban, A. T. A Mutation in the Human Immunodeficiency Virus Type 1 Transmembrane Glycoprotein Gp41 Dominantly Interferes with Fusion and Infectivity. *Proc Natl Acad Sci U S A* 1992, 89 (1), 70–74. DOI: 10.1073/pnas.89.1.70.
- (40) Qiao, H.; Armstrong, R. T.; Melikyan, G. B.; Cohen, F. S.; White, J. M. A Specific Point Mutant at Position 1 of the Influenza Hemagglutinin Fusion Peptide Displays a Hemifusion Phenotype. *Mol Biol Cell* 1999, 10 (8), 2759–2769. DOI: 10.1091/mbc.10.8.2759.
- (41) Madu, I. G.; Roth, S. L.; Belouzard, S.; Whittaker, G. R. Characterization of a Highly Conserved Domain within the Severe Acute Respiratory Syndrome Coronavirus Spike Protein S2 Domain with Characteristics of a Viral Fusion Peptide. *J Virol* 2009, 83 (15), 7411–7421. DOI: 10.1128/jvi.00079-09
- (42) Madu, I. G.; Belouzard, S.; Whittaker, G. R. SARS-Coronavirus Spike S2 Domain Flanked by Cysteine Residues C822 and C833 Is Important for Activation of Membrane Fusion. *Virology* 2009, 393 (2), 265–271. DOI: 10.1016/j.virol.2009.07.038.
- (43) Rokonujjaman, M.; Sahyouni, A.; Wolfe, R.; Jia, L.; Ghosh, U.; Weliky, D. P. A Large HIV Gp41 Construct with Trimer-of-Hairpins Structure Exhibits V2E Mutation-Dominant Attenuation of Vesicle Fusion and Helicity Very Similar to V2E Attenuation of HIV Fusion and Infection and Supports: (1) Hairpin Stabilization of Membrane Appositi. *Biophys Chem* 2023, 293 (November 2022), 106933. DOI: 10.1016/j.bpc.2022.106933.
- (44) Sup Kim, C.; Epand, R. F.; Leikina, E.; Epand, R. M.; Chernomordik, L. V. The Final Conformation of the Complete Ectodomain of the HA2 Subunit of Influenza Hemagglutinin Can by Itself Drive Low PH-Dependent Fusion. *Journal of Biological Chemistry* 2011, 286 (15), 13226–13234. DOI: 10.1074/jbc.M110.181297.
- (45) Gallaher, W. R. Detection of a Fusion Peptide Sequence in the Transmembrane Protein of Human Immunodeficiency Virus. *Cell* 1987, 50 (3), 327–328. DOI: 10.1016/0092-8674(87)90485-5.

- (46) Nobusawa, E.; Aoyama, T.; Kato, H.; Suzuki, Y.; Tateno, Y.; Nakajima, K. Comparison of Complete Amino Acid Sequences and Receptor-Binding Properties among 13 Serotypes of Hemagglutinins of Influenza A Viruses. *Virology* 1991, *182* (2), 475–485. DOI: 10.1016/0042-6822(91)90588-3.
- (47) Lorieau, J. L.; Louis, J. M.; Bax, A. The Complete Influenza Hemagglutinin Fusion Domain Adopts a Tight Helical Hairpin Arrangement at the Lipid: Water Interface. *Proc Natl Acad Sci U S A* 2010, *107* (25), 11341–11346. DOI: 10.1073/pnas.1006142107.
- (48) Ghosh, U.; Xie, L.; Jia, L.; Liang, S.; Weliky, D. P. Closed and Semiclosed Interhelical Structures in Membrane vs Closed and Open Structures in Detergent for the Influenza Virus Hemagglutinin Fusion Peptide and Correlation of Hydrophobic Surface Area with Fusion Catalysis. *J Am Chem Soc* 2015, *137* (24), 7548–7551. DOI: 10.1021/jacs.5b04578.
- (49) Larsson, P.; Kasson, P. M. Lipid Tail Protrusion in Simulations Predicts Fusogenic Activity of Influenza Fusion Peptide Mutants and Conformational Models. *PLoS Comput Biol* 2013, *9* (3). DOI: 10.1371/journal.pcbi.1002950.
- (50) Légaré, S.; Lagüe, P. The Influenza Fusion Peptide Promotes Lipid Polar Head Intrusion through Hydrogen Bonding with Phosphates and N-Terminal Membrane Insertion Depth. *Proteins: Structure, Function and Bioinformatics* 2014, *82* (9), 2118–2127. DOI: 10.1002/prot.24568.
- (51) Victor, B. L.; Lousa, D.; Antunes, J. M.; Soares, C. M. Self-Assembly Molecular Dynamics Simulations Shed Light into the Interaction of the Influenza Fusion Peptide with a Membrane Bilayer. *J Chem Inf Model* 2015, *55* (4), 795–805. DOI: 10.1021/ci500756v.
- (52) Pabis, A.; Rawle, R. J.; Kasson, P. M. Influenza Hemagglutinin Drives Viral Entry via Two Sequential Intramembrane Mechanisms. *Proc Natl Acad Sci U S A* 2020, *117* (13), 7200–7207. DOI: 10.1073/pnas.1914188117
- (53) Worch, R.; Krupa, J.; Filipek, A.; Szymaniec, A.; Setny, P. Three Conserved C-Terminal Residues of Influenza Fusion Peptide Alter Its Behavior at the Membrane Interface. *Biochim Biophys Acta Gen Subj* 2017, *1861* (2), 97–105. DOI: 10.1016/j.bbagen.2016.11.004.
- (54) Jaroniec, C. P.; Kaufman, J. D.; Stahl, S. J.; Viard, M.; Blumenthal, R.; Wingfield, P. T.; Bax, A. Structure and Dynamics of Micelle-Associated Human Immunodeficiency Virus Gp41 Fusion Domain. *Biochemistry* 2005, *44* (49), 16167–16180. DOI: 10.1021/bi051672a.
- (55) Gabrys, C. M.; Weliky, D. P. Chemical Shift Assignment and Structural Plasticity of a HIV Fusion Peptide Derivative in Dodecylphosphocholine Micelles. *Biochim Biophys Acta Biomembr* 2007, *1768* (12), 3225–3234. DOI: 10.1016/j.bbamem.2007.07.028.
- (56) Qiang, W.; Bodner, M. L.; Weliky, D. P. Solid-State NMR Spectroscopy of Human Immunodeficiency Virus Fusion Peptides Associated with Host-Cell-like Membranes: 2D Correlation Spectra and Distance Measurements Support a Fully Extended Conformation and Models for Specific Antiparallel Strand Regis. *J Am Chem Soc* 2008, *130* (16), 5459–5471. DOI: 10.1021/ja077302m.

- (57) Schmick, S. D.; Weliky, D. P. Major Antiparallel and Minor Parallel β Populations Detected in the Membrane-Associated Human Immunodeficiency Virus Fusion Peptide. *Biochemistry* 2010, 49 (50), 10623–10635. DOI: 10.1021/bi101389r.
- (58) Gabrys, C. M.; Qiang, W.; Sun, Y.; Xie, L.; Schmick, S. D.; Weliky, D. P. Solid-State Nuclear Magnetic Resonance Measurements of HIV Fusion Peptide ^{13}C to Lipid ^{31}P Proximities Support Similar Partially Inserted Membrane Locations of the α Helical and β Sheet Peptide Structures. *Journal of Physical Chemistry A* 2013, 117 (39), 9848–9859. DOI: 10.1021/jp312845w.
- (59) Buffy, J. J.; Hong, T.; Yamaguchi, S.; Waring, A. J.; Lehrer, R. I.; Hong, M. Solid-State NMR Investigation of the Depth of Insertion of Protegrin-1 in Lipid Bilayers Using Paramagnetic Mn^{2+} . *Biophys J* 2003, 85 (4), 2363–2373. DOI: 10.1016/S0006-3495(03)74660-8.
- (60) Su, Y.; Mani, R.; Hong, M. Asymmetric Insertion of Membrane Proteins in Lipid Bilayers by Solid-State NMR Paramagnetic Relaxation Enhancement: A Cell-Penetrating Peptide Example. *J Am Chem Soc* 2008, 130 (27), 8856–8864. DOI: 10.1021/ja802383t.
- (61) Marius Clore, G.; Iwahara, J. Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem Rev* 2009, 109 (9), 4108–4139. DOI: 10.1021/cr900033p.
- (62) Ferreira, T. M.; Coreta-Gomes, F.; Samuli Ollila, O. H.; Moreno, M. J.; Vaz, W. L. C.; Topgaard, D. Cholesterol and POPC Segmental Order Parameters in Lipid Membranes: Solid State ^1H - ^{13}C NMR and MD Simulation Studies. *Physical Chemistry Chemical Physics* 2013, 15 (6), 1976–1989. DOI: 10.1039/c2cp42738a.
- (63) Vaz, W. L. C.; Clegg, R. M.; Hallmann, D. Translational Diffusion of Lipids in Liquid Crystalline Phase Phosphatidylcholine Multibilayers. A Comparison of Experiment with Theory. *Biochemistry* 1985, 24 (3), 781–786. DOI: 10.1021/bi00324a037.
- (64) Orädd, G.; Lindblom, G. NMR Studies of Lipid Lateral Diffusion in the DMPC/Gramicidin D/Water System: Peptide Aggregation and Obstruction Effects. *Biophys J* 2004, 87 (2), 980–987. DOI: 10.1529/biophysj.103.038828.
- (65) Lindblom, G.; Orädd, G. Lipid Lateral Diffusion and Membrane Heterogeneity. *Biochim Biophys Acta Biomembr* 2009, 1788 (1), 234–244. DOI: 10.1016/j.bbmem.2008.08.016.
- (66) Yang, J.; Parkanzky, P. D.; Bodner, M. L.; Duskin, C. A.; Weliky, D. P. Application of REDOR Subtraction for Filtered MAS Observation of Labeled Backbone Carbons of Membrane-Bound Fusion Peptides. *Journal of Magnetic Resonance* 2002, 159 (2), 101–110. DOI: 10.1016/S1090-7807(02)00033-2.
- (67) Wasniewski, C. M.; Parkanzky, P. D.; Bodner, M. L.; Weliky, D. P. Solid-State Nuclear Magnetic Resonance Studies of HIV and Influenza Fusion Peptide Orientations in Membrane Bilayers Using Stacked Glass Plate Samples. In *Chemistry and Physics of Lipids*; 2004; Vol. 132, pp 89–100. DOI: 10.1016/j.chemphyslip.2004.09.008.

- (68) Sun, Y. Secondary Structure and Membrane Insertion of the Membrane-Associated Influenza Fusion Peptide Probed by Solid-State Nuclear Magnetic Resonance, Michigan state university, 2009. <http://dx.doi.org/10.1016/j.jaci.2012.05.050>.
- (69) Janosi, L.; Gorfe, A. A. Simulating POPC and POPC/POPG Bilayers: Conserved Packing and Altered Surface Reactivity. *J Chem Theory Comput* 2010, 6 (10), 3267–3273. DOI: 10.1021/ct100381g.
- (70) Golovina, E. A.; Golovin, A. V.; Hoekstra, F. A.; Faller, R. Water Replacement Hypothesis in Atomic Detail - Factors Determining the Structure of Dehydrated Bilayer Stacks. *Biophys J* 2009, 97 (2), 490–499. DOI: 10.1016/j.bpj.2009.05.007.
- (71) Chang, D. K.; Cheng, S. F.; Lin, C. H.; Kantchev, E. A. B.; Wu, C. W. Self-Association of Glutamic Acid-Rich Fusion Peptide Analogs of Influenza Hemagglutinin in the Membrane-Mimic Environments: Effects of Positional Difference of Glutamic Acids on Side Chain Ionization Constant and Intra- and Inter-Peptide Interactions. *Biochim Biophys Acta Biomembr* 2005, 1712 (1), 37–51. DOI: 10.1016/j.bbamem.2005.04.003.
- (72) Meier, P.; Ohmes, E.; Kothe, G. Multipulse Dynamic Nuclear Magnetic Resonance of Phospholipid Membranes. *J Chem Phys* 1986, 85 (6), 3598–3614. DOI: 10.1063/1.450931.
- (73) Scott Prosser, R.; Davis, J. H.; Mayer, C.; Weisz, K.; Kothe, G. Deuterium NMR Relaxation Studies of Peptide-Lipid Interactions. *Biochemistry* 1992, 31 (39), 9355–9363. DOI: 10.1021/bi00154a005.
- (74) Bodner, M. L.; Gabrys, C. M.; Parkanzky, P. D.; Yang, J.; Duskin, C. A.; Weliky, D. P. Temperature Dependence and Resonance Assignment of ¹³C NMR Spectra of Selectively and Uniformly Labeled Fusion Peptides Associated with Membranes. *Magnetic Resonance in Chemistry* 2004, 42 (2), 187–194. DOI: 10.1002/mrc.1331
- (75) Huster, D.; Arnold, K.; Gawrisch, K. Investigation of Lipid Organization in Biological Membranes by Two-Dimensional Nuclear Overhauser Enhancement Spectroscopy. *Journal of Physical Chemistry B* 1999, 103 (1), 243–251. DOI: 10.1021/jp983428h.
- (76) Scheidt, H. A.; Kolocaj, K.; Veje Kristensen, J.; Huster, D.; Langosch, D. Transmembrane Helix Induces Membrane Fusion through Lipid Binding and Splay. *Journal of Physical Chemistry Letters* 2018, 9 (12), 3181–3186. DOI: 10.1021/acs.jpcllett.8b00859.

Chapter 5 NMR assignment and structural probing of a small protein HM in bacterial inclusion bodies

5.1 Introduction

Escherichia coli (*E. coli*) is the most widely used bacterial host for producing recombinant proteins. It has a fast growth rate and *E. coli* culture related reagents are usually not expensive. With these advantages, lots of efforts have been taken to make well-developed tools of molecular manipulations as well as to explore its biology.¹

Many recombinant proteins produced in *E. coli* undergo irregular or incomplete folding processes resulting in insoluble aggregates, known as inclusion bodies (IBs).² They are highly dense refractile aggregates under electron microscope (Figure 5.1). Recombinant proteins (Rp) expressed by *E. coli* is the desired product and its biological function can be explored by subsequent studies. Overexpression of non-native proteins and highly hydrophobic protein often causes insoluble inclusion bodies aggregates.^{3,4} Although some Rp are expressed in both soluble and insoluble fractions, many other proteins can only be produced as IBs. The low yields of producing recombinant protein from *E. coli* brought by IBs formation makes express some biologically functional proteins in bacterial system less practical. As many recombinant proteins of commercial interest are expressed in IBs formation, a complete understanding of structure of IBs proteins may provide critical information about protein interaction to form aggregates. It may suggest new methods to improve active protein yields.⁵

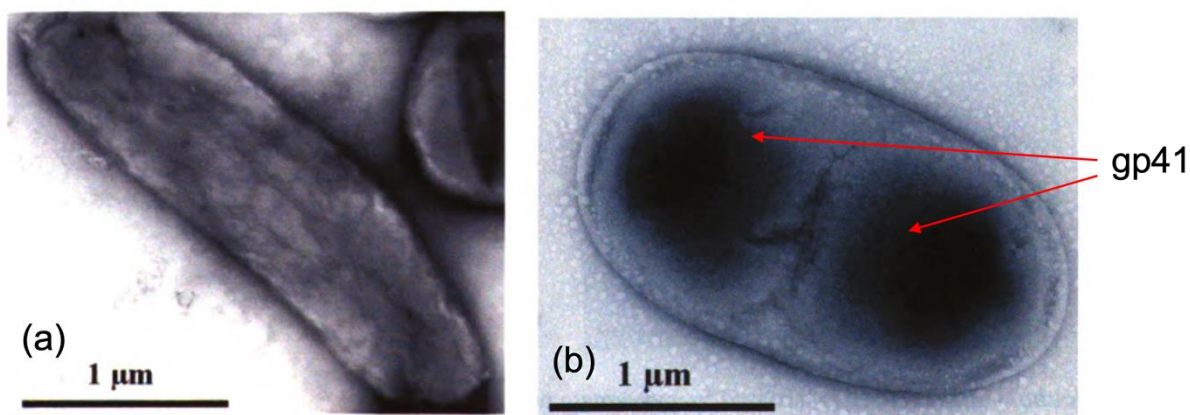


Figure 5.1. Electron micrograph of (a) un-induced *E. coli* cell, (b) induced bacterial cell producing gp41 of HIV.⁶

Figure 5.2 illustrates how IBs can be triggered to form. Post translational modifications (PTMs) are covalent modifications to some proteins after biosynthesized, such as acetyl, phosphoryl, and methyl to one or more amino acids.^{7,8} PTMs were thought to be restricted to eukaryotic systems but recent studies imply PTMs are important in prokaryotes. The most common PTMs to amino acid side modification in bacterial is shown in Figure 5.3.^{9,10} The PTMs could affect structures and functions of the proteins to generate bioactive proteins, which is usually seen in membrane proteins.⁷ If the unfolded peptide chains are expressed at the rate exceeding the ability that the host cells are capable of to manage protein PTMs and folding, the increasingly misfolded proteins tend to aggregate into IBs with hydrophobic residues exposed to exterior environment. The expressing rate of *E. coli* is influenced by environmental conditions, such as culture temperature and pH, since culture conditions control the partition of the recombinant protein into soluble and IB fractions.¹¹ Tailoring culture properties is a common strategy to minimize IBs formation of recombinant proteins in *E. coli*. The mechanism of how temperature and expression time influence the IBs formation is not well understood and should be treated case by case.

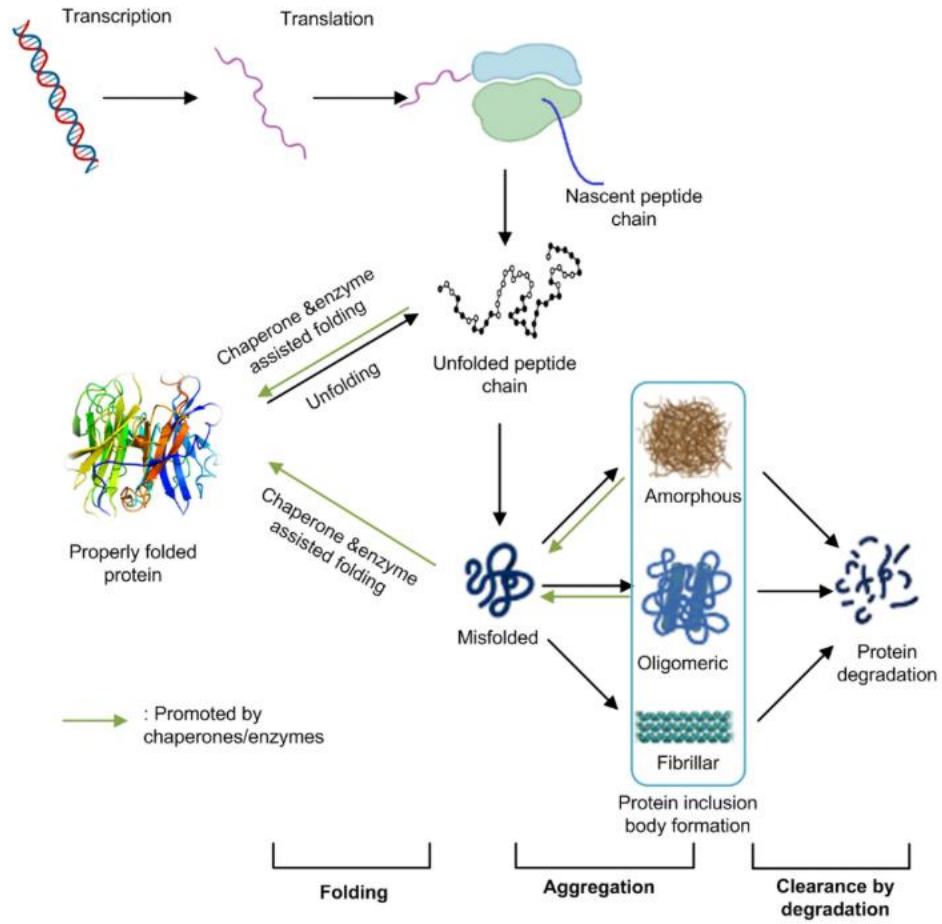


Figure 5.2. The protein homeostasis network in *E. coli* cells.¹

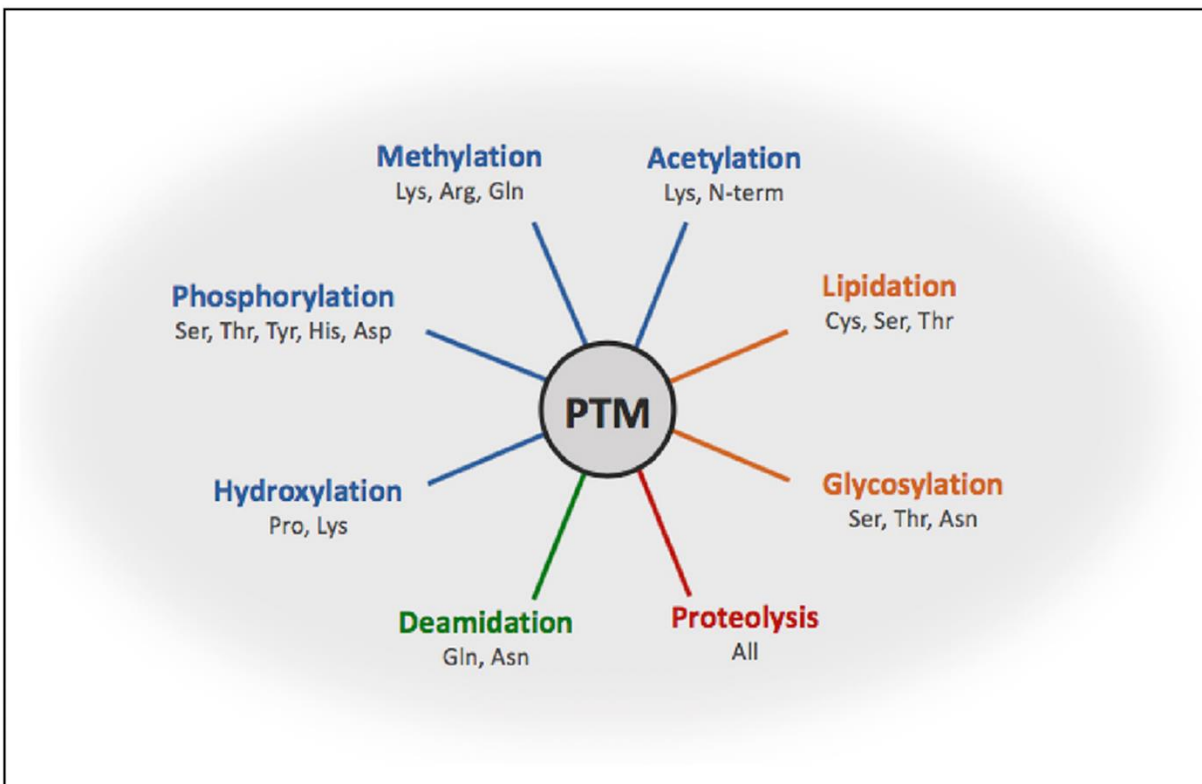


Figure 5.3. The amino acid side chains are most frequently modified in bacteria, making the corresponding PTMs the most common post-translational modifications for proteins expressed in bacterial system.¹⁰

To recover the soluble Rp from inclusion bodies, the general protocol contains solubilization of inclusion bodies by denaturant followed by removal of the denaturant, and then refolding with assistance of small molecule additives. The re-solubilization proteins of IBs formation can be achieved by strong denaturants, where the proteins are kept unfolded. The unfolded proteins are expected to refold properly to have an operative structure and function. Refolding is initiated by reducing the concentration of denaturant for solubilization. The denaturants help to maintain the protein's unfolded structure. When transferring from denaturant solution to aqueous solvent, hydrophobic residues of protein is exposed to water and tend to collapse into a compact structure. In the meantime, refolding competes with other side reactions, such as misfolding and aggregation. If the concentration of denaturant is too low, proteins lose flexibility to reorganize their structures. Consequently, misfolding and aggregation is very likely to happen and there is a kinetic competition between folding and aggregation. Therefore, there should be a balance that the unfolded protein can be compacted to a rigid structure while its structural flexibility is retained to

form bioactive folded structures.⁴ The refolding protocol is an attempt to maximize the ratio of folding-to-aggregation rates.

The overall yield of bioactive proteins recovered from inclusion bodies can vary significantly and depends on several factors including the nature of the protein, expression system, induction and harvesting condition, etc. A large volumes of biological fluid is needed to purify the recombinant protein and optimizing each factors through experimental process is crucial for maximizing the overall yeild.¹² Sometimes the recovery yield is very low, Voegl et al. reported the final Rp yield of FP-containing gp41 ectodomain (Fgp41) was only 0.1 mg of Rp/L of culture.^{13,14} Even though IBs formation of recombinant proteins is undesirable, there are several advantages of IBs. IBs are typically larger and more dense than the rest of the cellular components and the insolubility make them can be separated from the soluble cellular components by centrifugation. The homogeneity of IBs proteins reduces the number of steps to purify the protein as the result of very little contaminations existing in Ibs. The possible contaminations could be host membrane protein, plasmid DNA, and ribosomal RNA.¹¹ The purity of a plasmid-encoded v-galactosidase fusion protein, VPILAC, expressed in inclusion bodies is determined as 80-100% of the total protein by Western-Blot analysis.¹⁵

The IBs proteins was originally considered as aggregates lacking biological activity. However, biological activity and presence of native-like secondary structure of protein in IBs was confirmed three decades ago.¹⁶ Green Fluorescent Protein in IBs is highly fluorescent; the cytokine human granulocyte colony-stimulating factor in IBs adopts native structure and is fully active.^{17,18} Proteins aggregates can be either ordered amyloid fibrils or amorphous structure, such as bacterial IBs. IBs proteins are not just unstructured aggregates consisting of misfolded proteins associating by non-specific hydrophobic interactions. Studies indicates that IBs are often enriched in β sheet secondary structure and ssNMR studies supports α helical structure also exists in some IBs proteins.¹⁹

There is very little information about proteins structure as IBs formation, but the structural information may be expanding our knowledge of recombinant proteins structure and then helpful to develop a solubilization and refold approach to reach a high purified yield. FTIR, XRD, CD, cryo-electron microscopy and NMR are common methods to evaluate the structure of IBs.^{19,20} FTIR and CD can provide overall secondary structural information of the sample. It is also possible to know the relative amount of different secondary structure by curve-fitting of CD data. On the

contrary, NMR detect more detailed information regarding to proteins, but a large quantity of sample is needed. Curtis-Fisk et al. have used REDOR to explore native conformation of FHA2 IBs by site-specific labelling. FHA2 is a membrane protein including fusion peptide and soluble ectodomain of HA2 subunit of influenza virus. Chemical shifts of labeled residues are more consistent with characteristic α -helical structure rather than β -sheet secondary structure, proving the FHA2 IBs retains native α -helical conformation at least for a significant fraction as of IBs.²¹ This project is focused on the insoluble fraction of HM protein. HM stands for hairpin consisting of N- and C-helix connecting by a non-native loop of gp41 along with membrane proximal external region (MPER). It has 109 residues with molecular weight as 12961 Dalton (Figure 5.4). The gp41 is critical for membrane fusion which serves as catalyst to merge the viral and host cell's membranes.

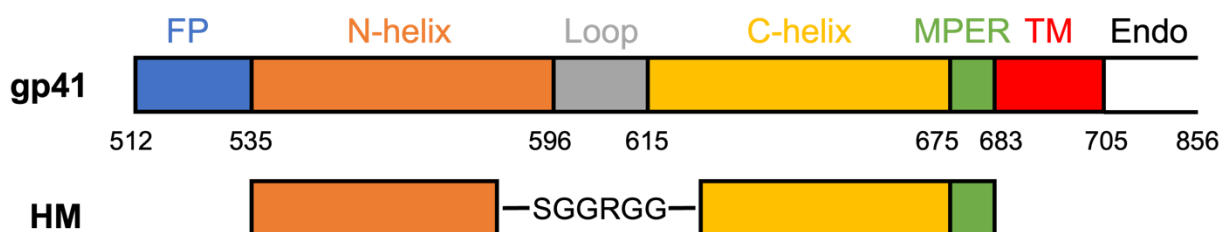


Figure 5.4. Schematic representation of HM.

Tran et al. found that N-helix and C-helix regions of gp41 at pre- and post-fusion state are both helical, shown in Figure 5.5 and Figure 5.6.^{22,23} The three N-helices of the trimeric envelope glycoprotein (Env) are less compactly packed than in the post-fusion. The crystal structure of N- and C-helices indicates helical conformation at both pre- and post-fusion state.

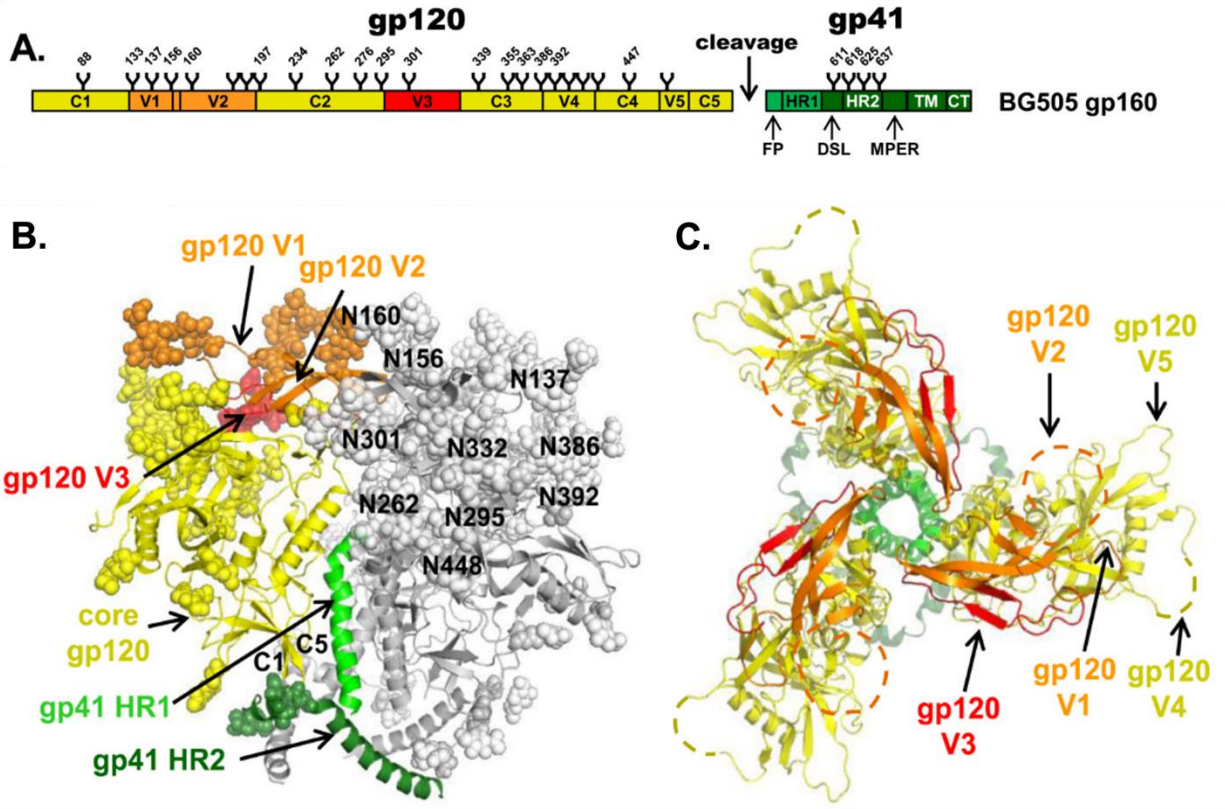


Figure 5.5. (A) Schematic construct of full-length gp160. N-linked glycans are shown and numbered on their respective Asn residues. C and V represents constant and variable regions of gp120. (B) Side view of the Env trimer. (C) View of Env down the trimer axis.

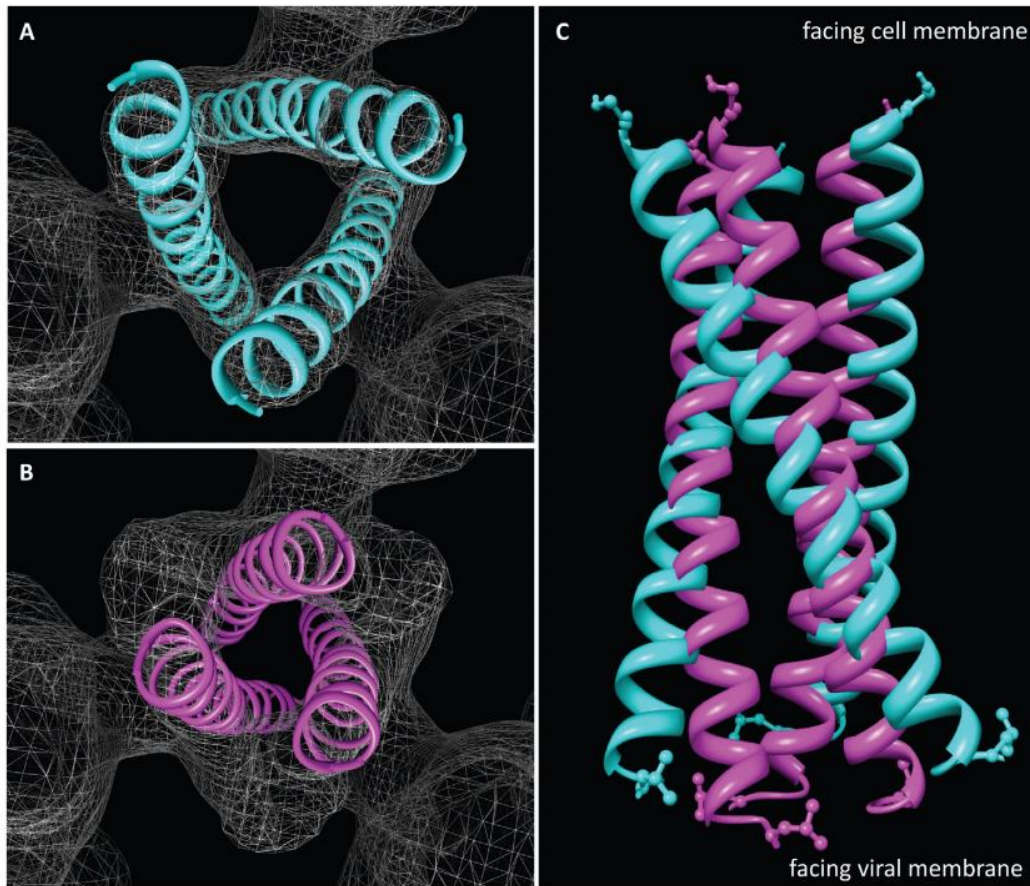


Figure 5.6. (a) Top view of the three N-helices in the pre-fusion state. (b) Top view of the same helices at post-fusion state. (c) Superposition of the N-helices in the pre- and post-fusion state.

Tan et al. reported the crystal structure of N34(L6)C28, which is N-34 and C-28 residues of N- and C- helices region connect by a hydrophilic linker (Figure 5.8). The N34(L6)C28 trimer folds into a six-helical bundle with three N-terminal helices forming a central, parallel, trimeric coiled coil whereas three C-terminal helices folding into the opposite direction into three hydrophobic grooves on the surface of N-terminal trimer.²⁴

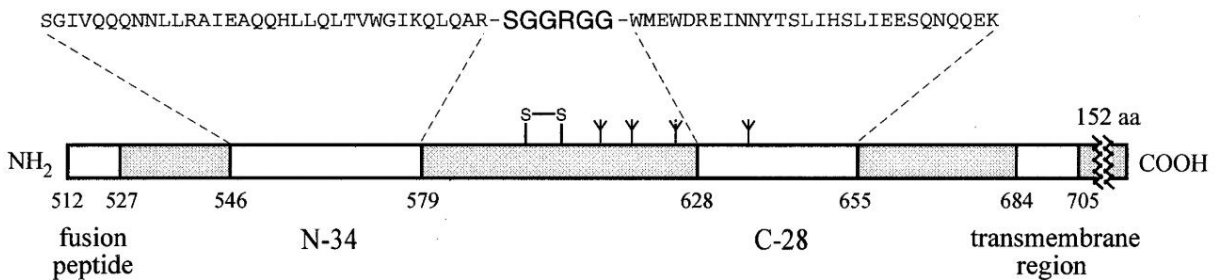


Figure 5.7. Schematic representation of N34(L6)C28.

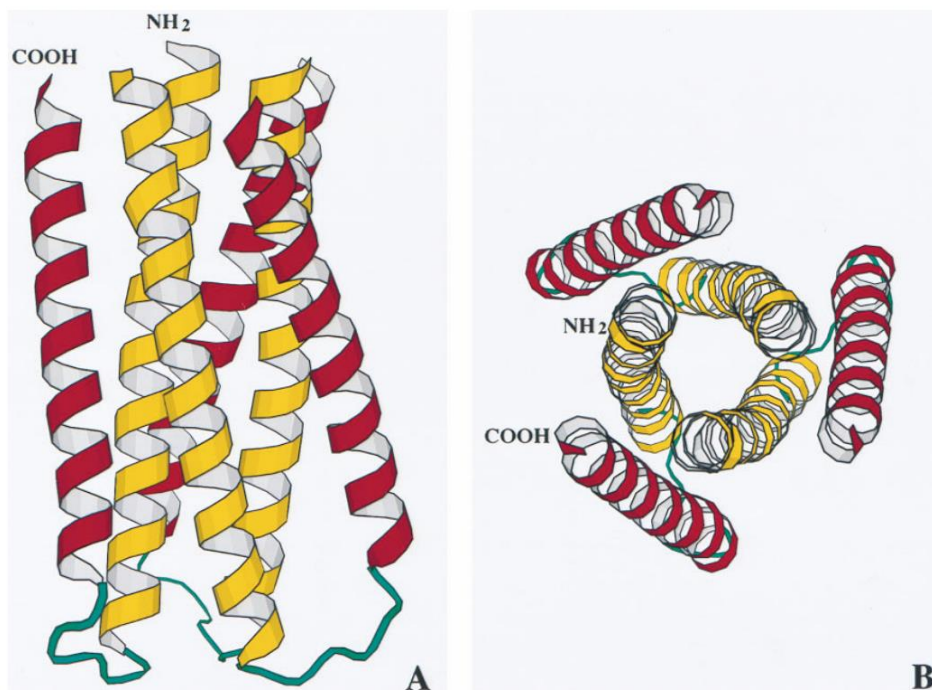


Figure 5.8. Structure of the N34(L6)C28 trimer. (A) Side view of the N34(L6)C28 trimer where N-helices are colored yellow, and the C-helices are purple. (B) End-on view of the trimer looking down the three-fold axis of the trimer.

Compared with N34(L6)C28, HM contains MPER region and longer N- and C- helical region. Purified soluble HM retains α -helical structure evidenced by CD data.²⁵ However, there is very limited information about HM structure in IBs. Explore HM structure in IBs may help us create a more efficient protocol to recover HM from IBs formation. Perhaps it will shed lights to better understand the folding mechanism of HM. ssNMR is the major characterization to probe the structure of HM. The advantage of ssNMR is that the insoluble HM IBs is directly lyophilized so the native structure of HM will not be destroyed. The assignment of multidimensional correlation spectra demonstrates the native structure of the HM as insoluble formation.

5.2 Results

5.2.1 Protein purification

Figure 5.11 is the MALDI mass spectrum of HM inclusion bodies protein, showing the molecular weight of HM is 13kDa. Figure 5.9 is the SDS-PAGE gel of pellet after lysis with PBS buffer only and PBS buffer followed by wash buffer (50 mM Na_3PO_4 , 300 mM NaCl, 1% w/w Triton X-100, pH 8.0). The wash buffer successfully purified HM inclusion bodies protein from soluble protein, lipids and membrane proteins, which is supported by the protomeric analysis (Figure 5.10). Figure

5.11 is the SDS-PAGE of different labeling HM inclusion bodies proteins: uniformly ^{13}C , ^{15}N labeled HM; 1,3- ^{13}C -glycerol, ^{15}N labeled HM; 2- ^{13}C -glycerol, ^{15}N labeled HM and Leucine reversed labeled, ^{13}C , ^{15}N labeled HM. Samples and their labeling materials can be found at Section 2.2.1. The major band is located at the position corresponding to 39 kDa and there also exist bands at 74 kDa and 13 kDa. The bands can be interpreted that the HM trimer is dominated and the monomer and hexamer probably exist.

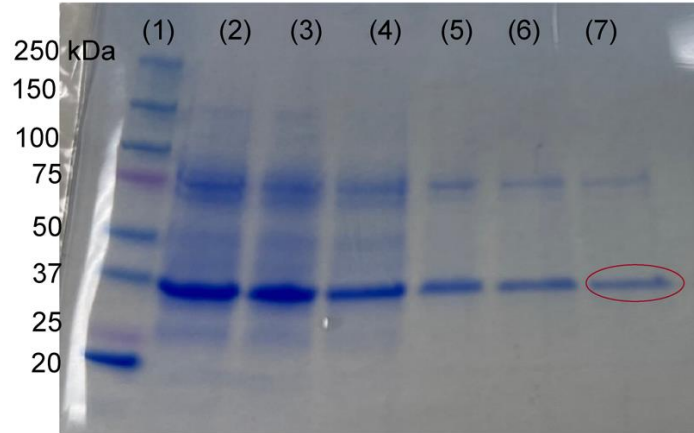


Figure 5.9. SDS-PAGE gel of HM inclusion bodies protein. (2)-(4) Pellet of cell lysis with PBS buffer; (5)-(7) Pellet of cell lysis with PBS buffer followed by wash buffer.

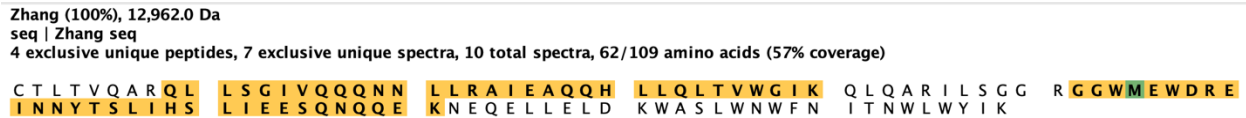


Figure 5.10. Proteomic analysis for the protein which is located at ~37 kDa in SDS-PAGE gel (circled in Figure 5.9).

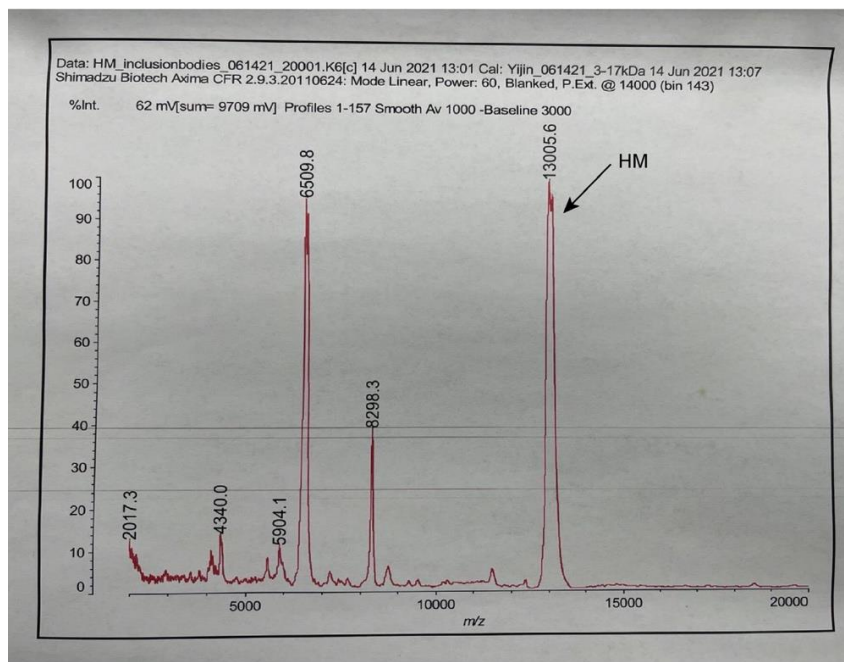


Figure 5.11. MALDI-mass spectrum of HM after lysis with PBS and wash buffers. The peak at 13005.6 represents HM^+ ion and 6509.8 peak is assigned as HM^{2+} .

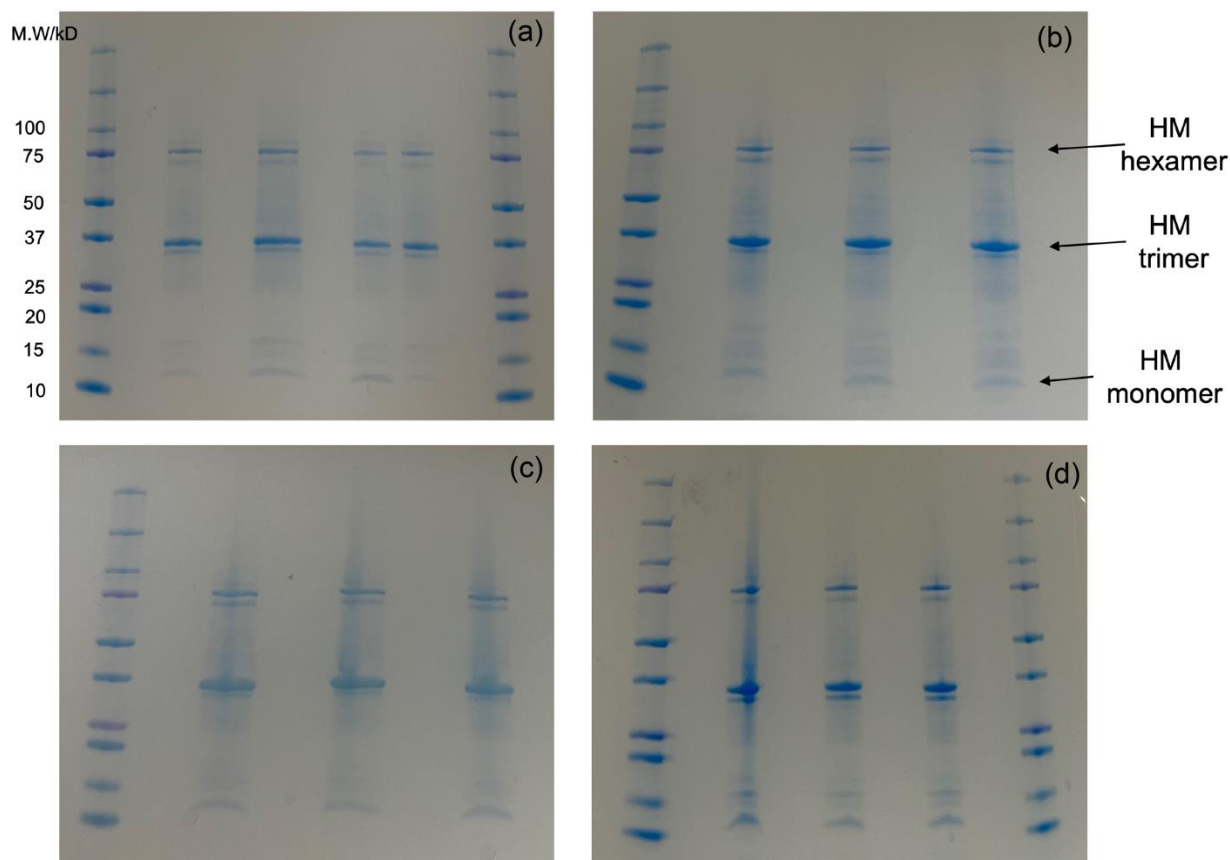


Figure 5.12. SDS-PAGE of HM inclusion bodies proteins. (a) uniformly ^{13}C , ^{15}N labeled HM; (b) $2\text{-}^{13}\text{C}$ -glycerol, ^{15}N labeled HM; (c) $1,3\text{-}^{13}\text{C}$ -glycerol, ^{15}N labeled HM; (d) Leucine reversely labeled, ^{13}C , ^{15}N labeled HM.

5.2.2 Microscopy images

The light microscopy images (Figure 5.13) support different morphologies of the two samples. Bigger clusters are observed in HM purified with wash buffer (50 mM Na_3PO_4 , 300 mM NaCl, 1% w/w Triton X-100, pH 8.0). Only one cluster appears in the sample washed by PBS buffer while there are more clusters in the sample washed by PBS and wash buffer. The latter has cluster of bigger size. Transmission electron microscopy (TEM) images (Figure 5.14) of HM reflected that the wash buffer had a big impact to the sample's morphology as well. The sample purified by PBS and wash buffer reveals characteristic feature of filament. The width of each fibril is ~ 5 nm which matches the typical diameter of amyloid fibril.²⁶

To further verify the existence of β sheet in HM IBs, the sample were stained by Congo red to test for birefringence (Figure 5.15). If the protein containing amyloid-like structure is stained by Congo red, the birefringence will present. The color is anomalous under polarized field, which means it is different from the color of Congo red in ordinary illumination. The green, apple-green and

yellow are the traditional colors reported birefringence.^{27,28} Although there are one or two spots (circled in Figure 5.15) under polarized field microscopy in green/yellow, it is not solid evidence for amyloid-like structure.

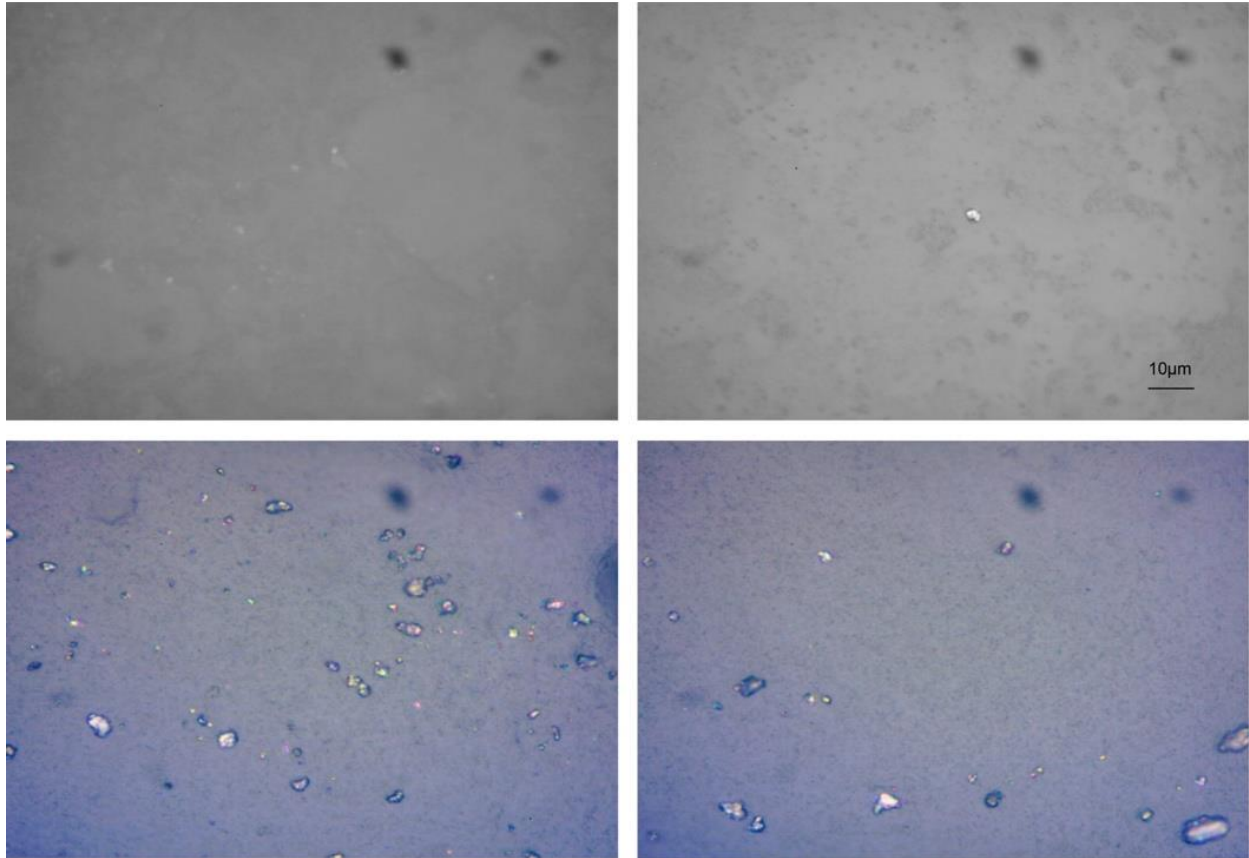


Figure 5.13. Light microscopy images for HM. Top: Pellet after PBS (3×) lysis. Bottom: Pellet lysis in PBS (3×) followed by lysis in wash buffer (50 mM Na₃PO₄, 300 mM NaCl, 1% w/w Triton X-100, pH 8.0).

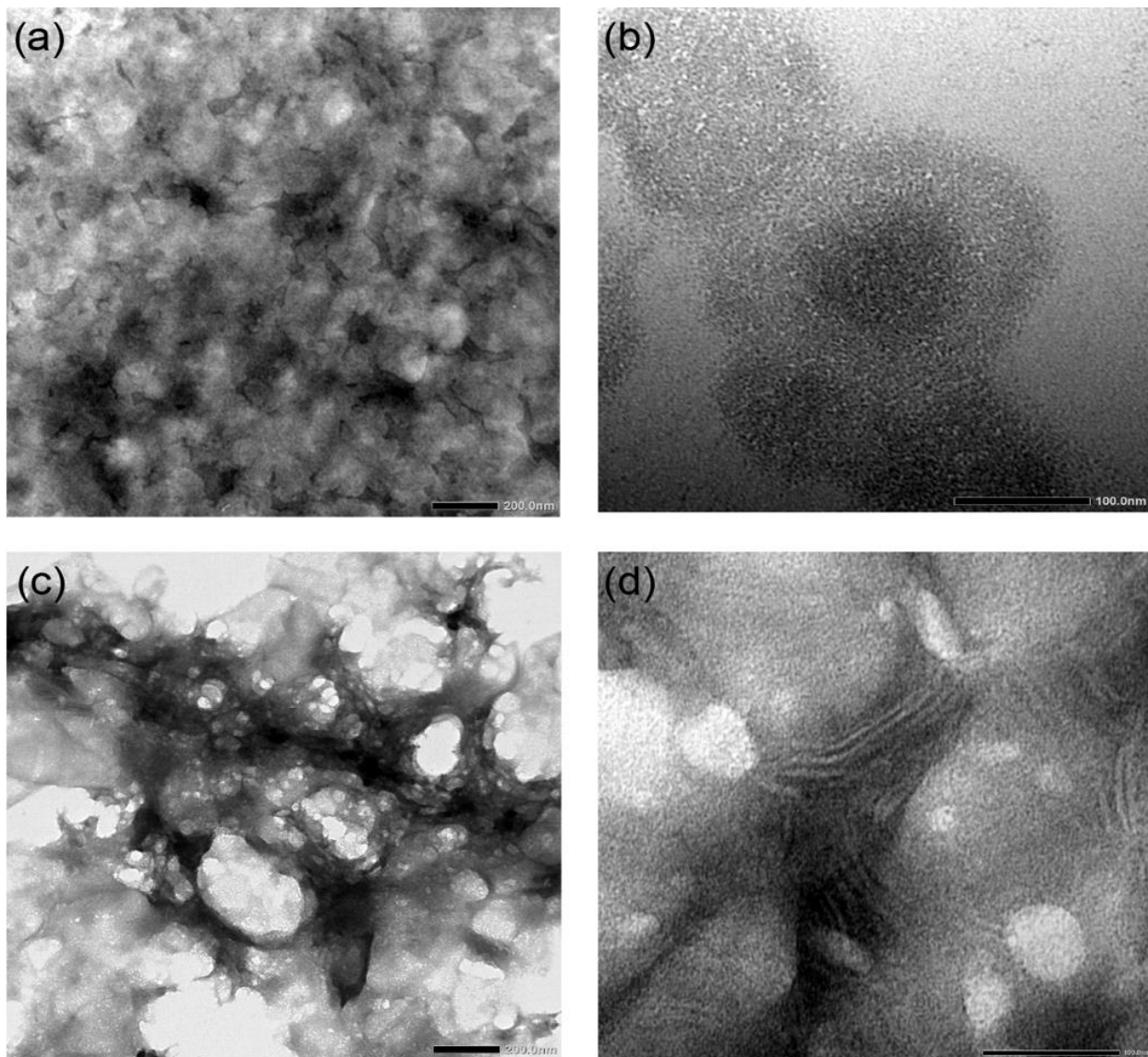


Figure 5.14. TEM images for unlabeled HM lysis with PBS wash (3 \times) (a and b) and PBS wash (3 \times) followed by wash buffer. The magnifications are $\times 20k$ (a and c), $\times 100k$ (b and d).

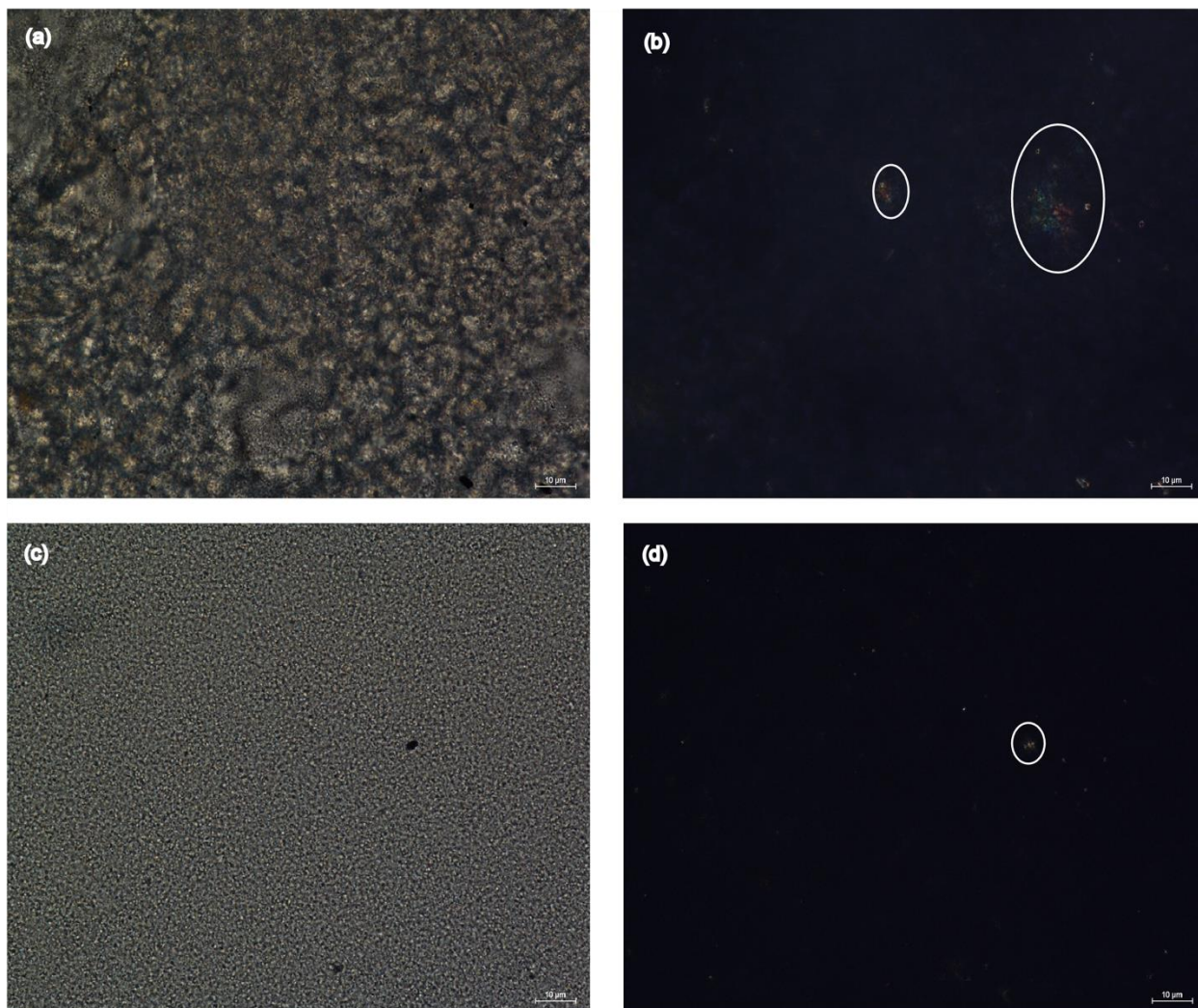


Figure 5.15. Congo red-stained HM under bright field microscopy (left column) and polarized filed microscopy (right column). (a) and (b) are images of unlabeled HM lysis with PBS wash (3×). (c) and (d) are images of unlabeled HM lysis with PBS wash (3×) followed by wash buffer.

5.2.3 ssNMR spectra

Figure 5.16 is the 1D INEPT and $^1\text{H}\rightarrow^{13}\text{C}$ CP spectrum of the HM sample. INPET is the magnetization transfer driven by scalar coupling. It is an effective measurement for highly mobile region of the protein samples. The ratio of integrated CA peak (48-65 ppm) between INEPT and CP is ~ 0.24 and 0.35 for sidechain carbons (10-45 ppm). Substantial signals in INPET indicates that part of the protein sample has disordered dynamic region.²⁹ The INEPT signal is superimposed with ^1H - ^{13}C CP signal suggesting that the HM sample is a mixture of dynamic and immobile region.

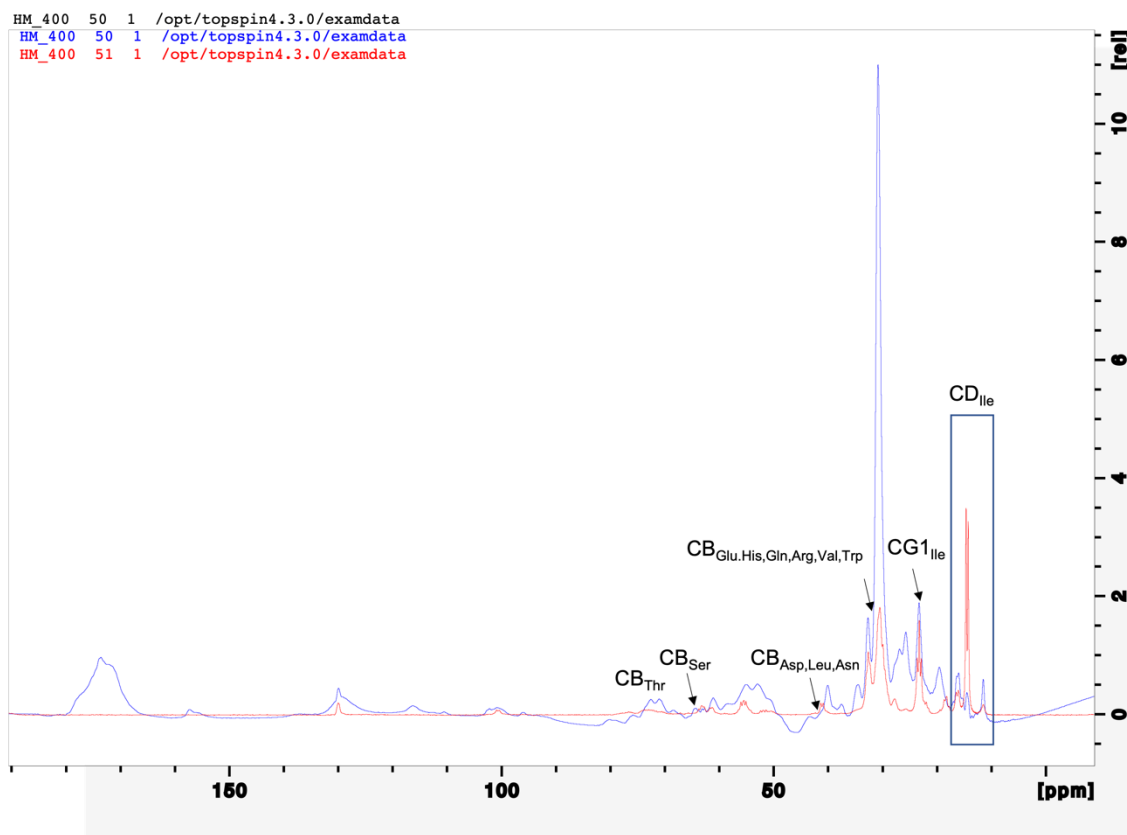


Figure 5.16. Comparison of INEPT (red) and ^1H - ^{13}C CP (blue) spectra of HM. Both were collected in a 9.4 T magnet at 298 K with 8 kHz spinning rate, 2048 scans.

The ^1H - ^{13}C CP spectra of U-HM, Leu-Rev-HM, 1,3- ^{13}C -Glyc-HM, and 2- ^{13}C -Glyc-HM is shown in Figure 5.17. The spectrum of Leu-Rev-HM and U-HM have a similar pattern and the difference is supposed to be due to unlabeled leucines. It is expectable that there exist strong signals at ~ 43 - 45 ppm and ~ 27 ppm matching CB and CG of leucine respectively in CP spectrum of U-HM. Compare with extensive labeling, selectively labeling is a labeling strategy for resolution enhancement relying on the amino acid biosynthetic pathways in bacteria.^{30,31} Following the

labeling pattern from 1,3- ^{13}C -glycreol and 2- ^{13}C -glycerol as sole carbon source, some of the peaks at aliphatic region have been assigned.(Figure 5.16)³⁰⁻³²

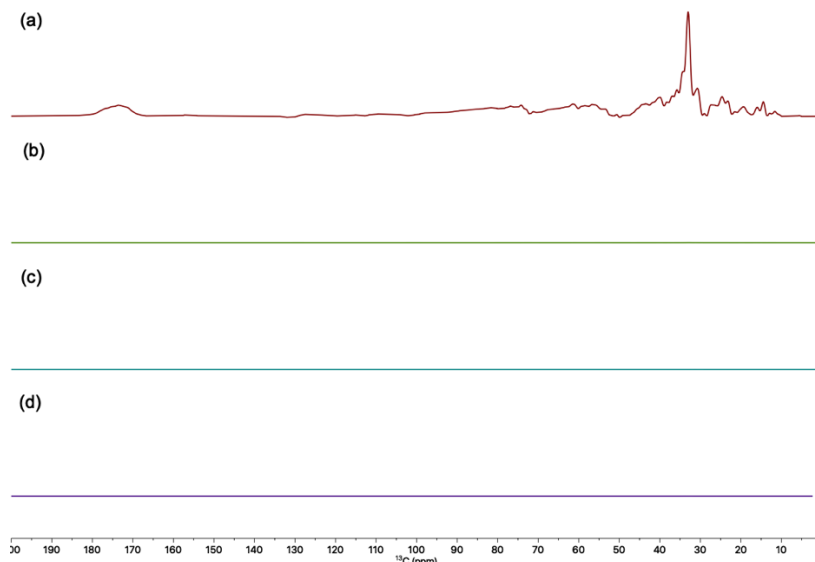


Figure 5.17. ^1H - ^{13}C CP spectra of (a) U-HM, (b) Leu-Rev-HM, (c) 1,3- ^{13}C -Glyc-HM, (d) 2- ^{13}C -Glyc-HM recorded on an 18.8 T magnet at 253 K with spinning frequency of 16 kHz. Typical parameters are: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H , ~ 80 kHz decoupling, and 50 ms acquisition time. There are 128, 64, 16, and 16 scans for U-HM, Leu-Rev-HM, 1,3- ^{13}C -Glyc-HM, and 2- ^{13}C -Glyc-HM respectively. Top: the original full spectra of all samples. Middle: the full spectra of all samples but (b) has been scaled up by 60 folds and (c) and (d) has been scaled up by 120 folds. The weak intensity of 1,3- ^{13}C -Glyc-HM, and 2- ^{13}C -Glyc-HM could be because of the low abundance of ^{13}C compared with U-HM. Bottom: the expanded region of 10-90 ppm of the spectra.

Figure 5.17 (cont'd)

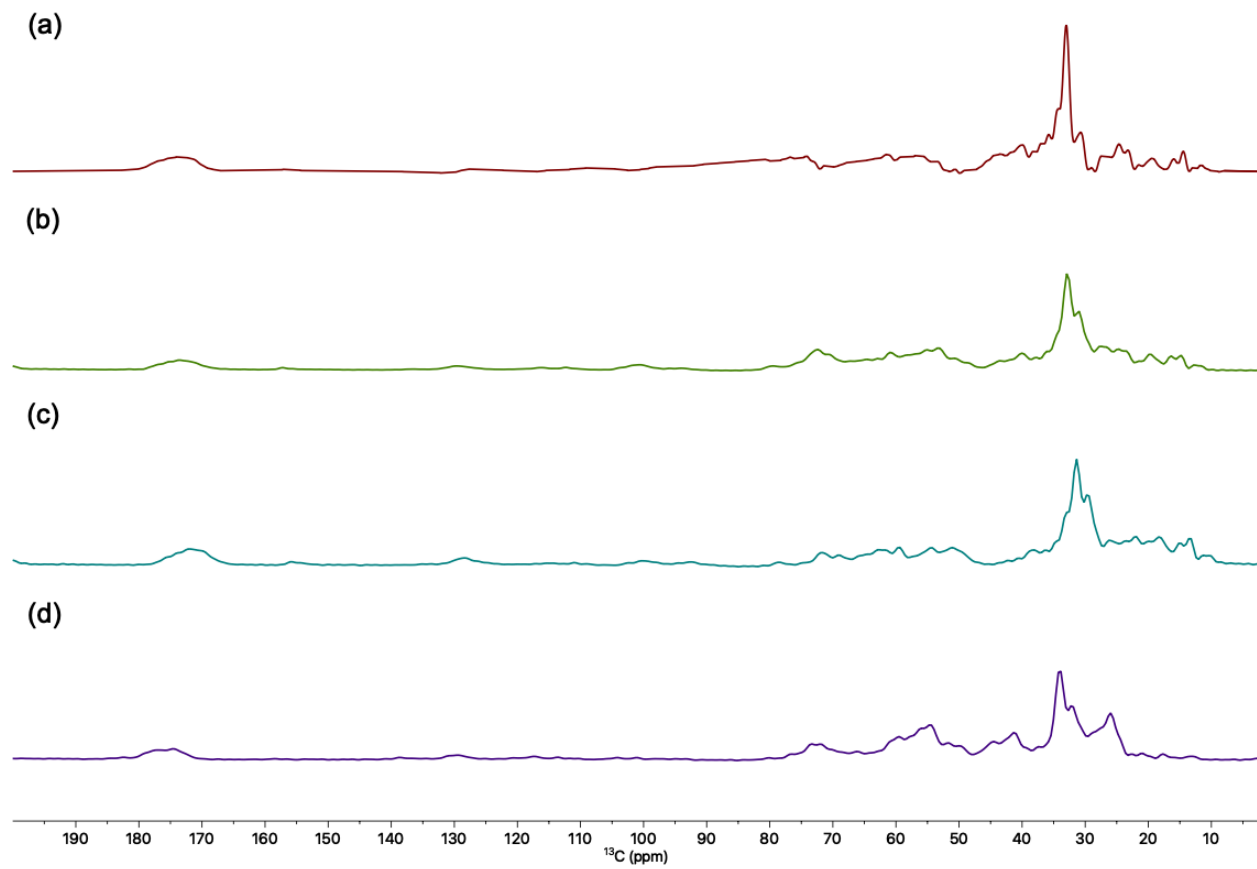
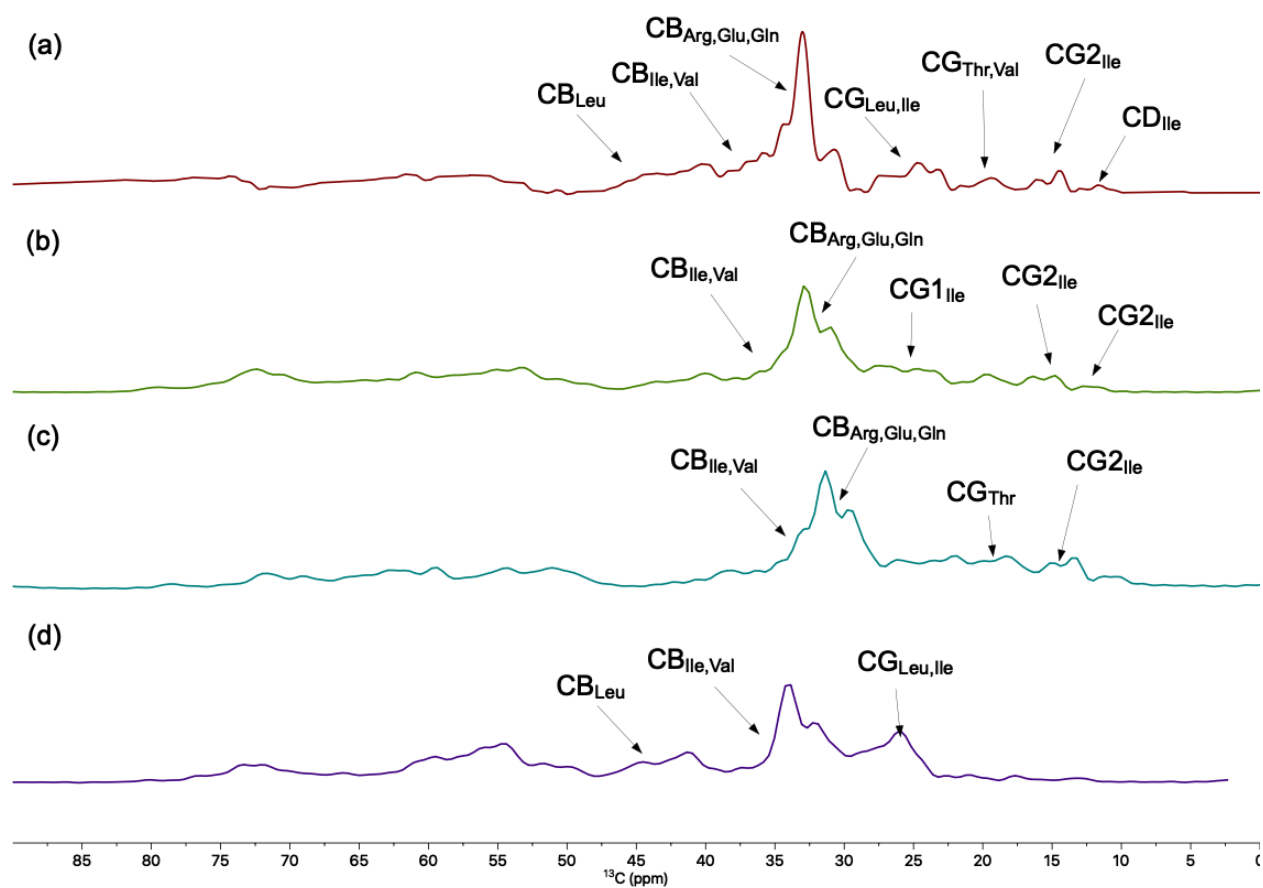


Figure 5.17 (cont'd)



Samples of all multidimensional ssNMR measurements were prepared following the protocol described in Chapter 2. The purified HM Ibs was typically lyophilized for two hours. An example of assignment based on NCACX is shown in Figure 5.17. The spectra are viewed by ^{15}N slices at 121 ppm and 125 ppm. The cross peaks at (55 ppm, 18 ppm) ($\text{C}\alpha$, $\text{C}\beta$) and (52 ppm, 21 ppm) ($\text{C}\alpha$, $\text{C}\beta$) are corresponding to Alanine adopting α -helical and β -sheet structure respectively by comparing with statistical chemical shifts of different secondary structures. ^{13}C - ^{13}C homonuclear correlation and NCOCA spectra and NCACX spectra of 1,3- ^{13}C -Glyc-HM and 2- ^{13}C -Glyc-HM are provided (Figures 5.19-5.22). All spectra were recorded at an 18.8 T Bruker spectrometer.

The last step of multidimensional measurement before data acquisition in my project is spin diffusion. The magnetization transfers driven by dipolar couplings between carbons and spatially transfer to proximity carbons. It is very possible that magnetization from source $\text{C}\alpha$ can transfer to its nearby $\text{C}\alpha$. In this scenario, there should be a cross peak at 50~60 ppm ($\text{C}\alpha$ region) next to a diagonal peak at $\text{C}\alpha$ region, which has been observed in some 3D spectra. Those peaks can be helpful for backbone assignment when the connecting $\text{C}\alpha$ have well separated chemical shifts. For example, glycine $\text{C}\alpha$ at 45-50ppm and isoleucine $\text{C}\alpha$ at 60-70 ppm. It is possible to sequentially assign those cross peaks combining with the protein sequence.

Most amino acids have $^{15}\text{N}_i$ - $^{13}\text{C}\alpha_i$ resonance at the region (^{15}N , ^{13}C) = (110-120 ppm, 52-60 ppm) except for Ala (^{15}N , ^{13}C) = (121 ppm, 55 ppm), Val (^{15}N , ^{13}C) = (120 ppm, 66 ppm), Gly (^{15}N , ^{13}C) = (106 ppm, 47 ppm), Pro (^{15}N , ^{13}C) = (N/A, 66 ppm), Ser (^{15}N , ^{13}C) = (115 ppm, 61 ppm) and Thr (^{15}N , ^{13}C) = (115 ppm, 66 ppm) having relatively distinct ^{15}N or $^{13}\text{C}\alpha$ chemical shifts.³³ For HM sequence, 77% of the residues fall at the 52-60ppm region and the 17 leucine, 9 isoleucine, 14 glutamine, and 9 glutamic acid of HM limited the spectral resolution. The acquired 3D correlation spectra have not solved this problem completely. In order to finish sequential assignment, it might be necessary to sparsely label the protein utilizing ^{15}N , ^{13}C enriched amino acid to label the protein.³⁴ Commonly, only one amino acid should be labeled. But label multiple amino acids simultaneously could be useful. If alanine and arginine are labeled at the same time, the sequential number of alanine (R23, R45) can be identified with either NCOCX or CONCACX experiment. The only observed arginine signal in NCOCX spectrum would be R23 transferred from A24 and the only observed arginine signal in CONCACX experiment would be R45 transferred from A44. It might also be feasible to label multiple amino acids with relatively distinctive resonance to reduce workload. For example, glycine and isoleucine can be labeled at

the same time since glycine C α locates at 45-47 ppm while isoleucine C α locates at 60-66 ppm. It is easily to distinguish them from one another.

The peak assignment done with NCACX is summarized as Table 5.1. The secondary shift δ_{sec} is defined as deviation from chemical shift (δ) of random-coil structure, $\delta_{\text{sec}} = (\delta_{\text{C}\alpha,\text{sample}} - \delta_{\text{C}\beta,\text{sample}}) - (\delta_{\text{C}\alpha,\text{coil}} - \delta_{\text{C}\beta,\text{coil}})$. The sign of secondary shift is predictive for secondary structure of the sample. The positive sign indicates the sample adopts α helical structure while negative sign is an indicator for β sheet.^{35,36} The secondary shifts of HM IBs proteins compared with 44 kDa gp41 ectodomain of Simian immunodeficiency virus (SIV) is presented in Figure 5.22. Each dot represents an occurrence of resonance assignment. The solution NMR data of gp41 of SIV indicated that the ectodomain of gp41 consists of a N-terminal helix (residues 29 to 77) and a C-terminal helix (residue 108 to 147) connecting by a 30 residue loop. It is believed that SIV results can be directly transferred to ectodomain of gp41 of HIV as the result of the high degree of sequence identity (~55%). Consequently, it is convincing to compare HM IBs protein with SIV gp41.³⁷ The secondary shifts of gp41 of SIV are almost all positive, representing a helical structure. On the contrary, there are some distributions which the secondary shifts are negative for HM IBs proteins.

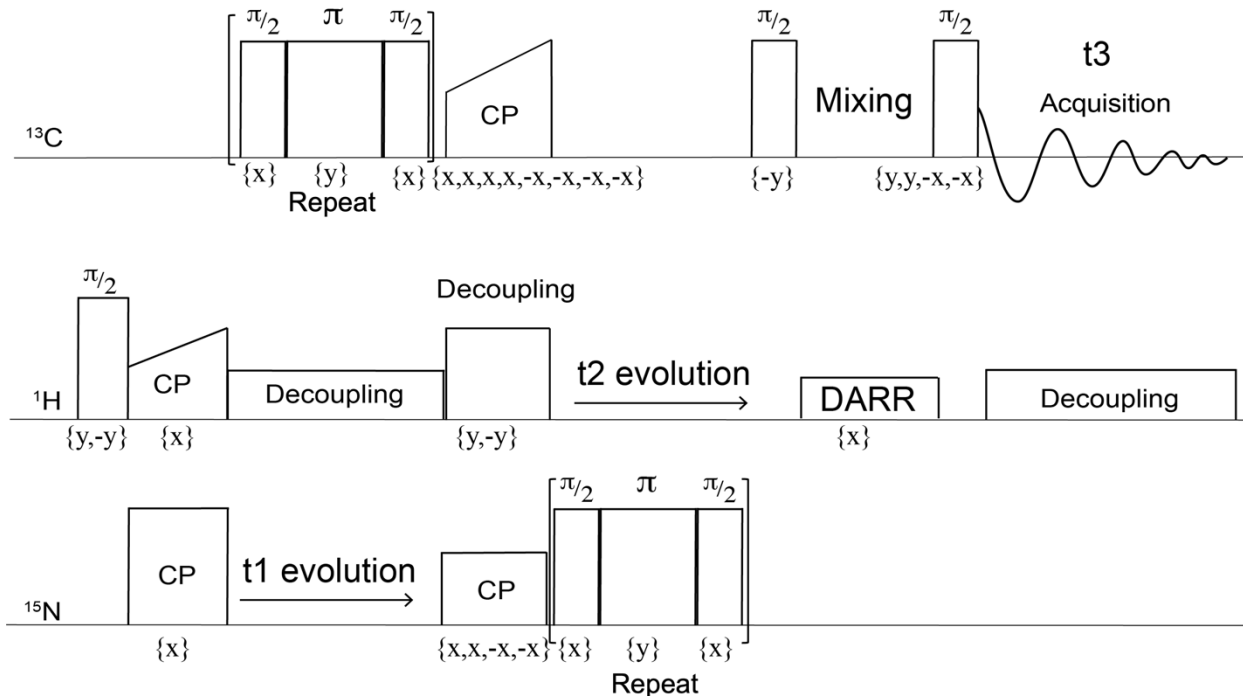
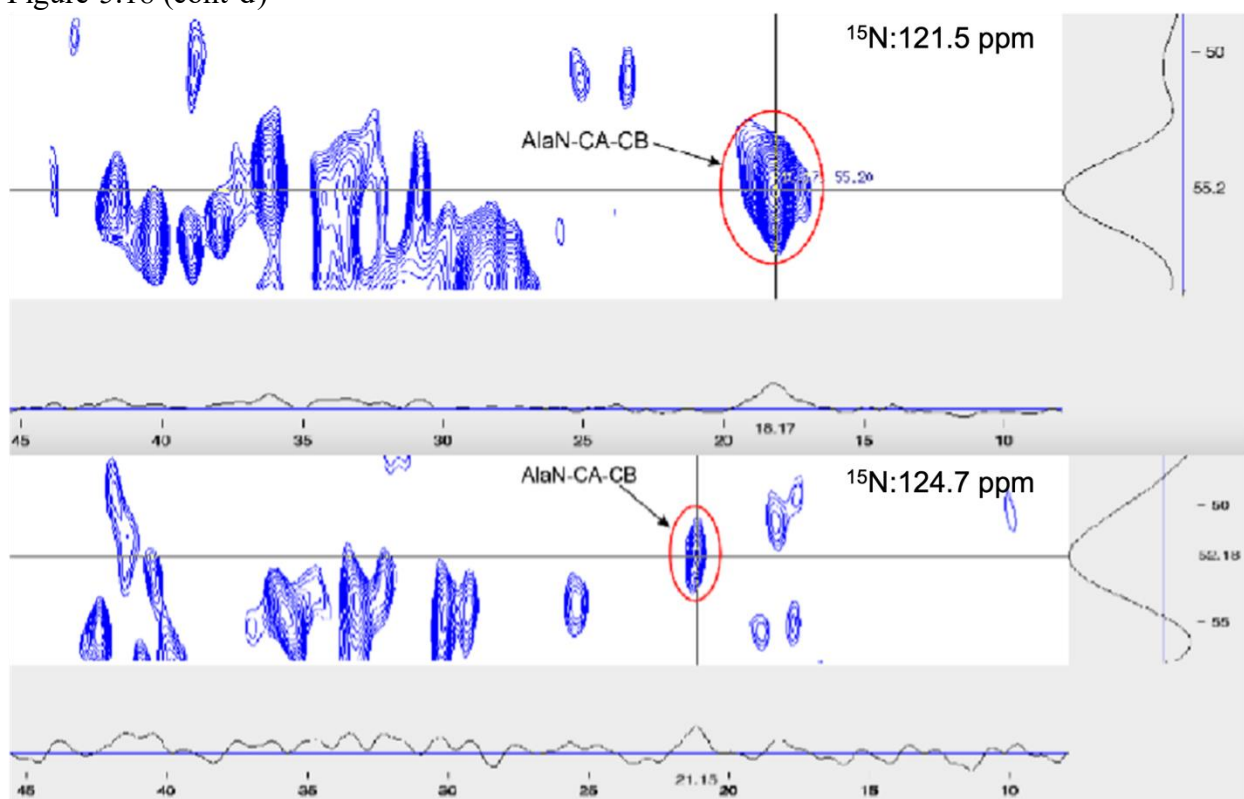


Figure 5.18. 3D NCACX spectra of Leu-Rev-HM edited by ^{15}N slices, both corresponding to Alanine resonance. The pulse sequence is shown on the top of the figure. The spectrum was processed in Topspin and viewed in Pooky for assignment. Top: Ala with a α -helical structure; Bottom: Ala with a β -sheet structure. 1D slice of AlaN-CA-CB peak in CA and CX dimension are shown in the right side and the bottom of each spectrum. Pulse width are: 2.5 μ , 4.5 μ and 7 μ for ^1H , ^{13}C and ^{15}N $\pi/2$ pulse respectively. Other parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{15}\text{N}$ CP contact time 1.1 ms with 25 kHz on ^{15}N and 63-75 kHz linear CP ramp on ^1H ; $^{15}\text{N} \rightarrow ^{13}\text{C}$ CP: ^{13}C offset was set to 55 ppm and 120 ppm for ^{15}N , contact time 4ms, 35 kHz on ^{15}N and 16-22 kHz on ^{13}C , 87 kHz of ^1H decoupling applied at the same time; DARR mixing: 56 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 75 kHz. Size of FID (number of points) are 56 for ^{15}N and 32 for ^{13}C ; spectral width was set 80 ppm for ^{15}N and 40 ppm for ^{13}C and increment for delay were 0.308 ms and 0.062 ms for ^{15}N and ^{13}C respectively. (The t1 and t2 values are incremented by increment of delay and the sequence is repeated for each points indirect dimension to create an FID array with three dimension $S(t_1, t_2, t_3)$.) The data was collected at 253 K with 16 kHz spinning rate and 96 scans. The spectrum was processed with the widow function QSINE, SSB = 2 for all three dimensions. QSINE (Quadratic Sine Bell) window function is defined as: $\omega(n) = \sin\left(\frac{\pi n}{N}\right)^2$, where $\omega(n)$ is the value of the window function at sample point n and N is the total number of samples in the window.

Figure 5.18 (cont'd)



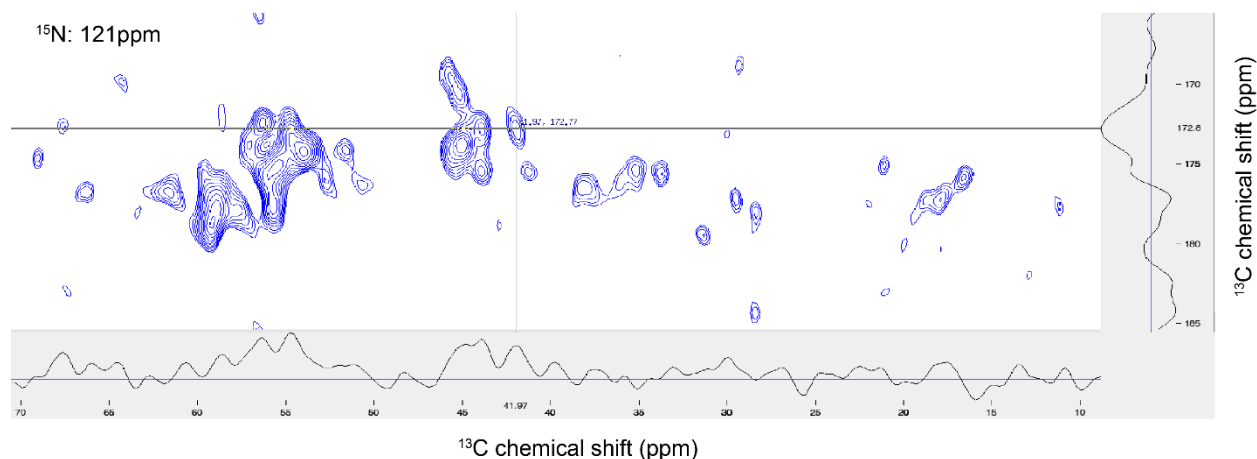


Figure 5.19. 3D NCOCX spectrum of Leu-Rev-HM edited by ^{15}N slices. The spectrum was displayed with vertical alignment for clarity. The CX region of 70-130 ppm was omitted for the reason that no peak was observed in the region. The peak with crosshairs were randomly chosen to exhibit the 1D slice of a CO and CB peak where readers can find the signal-to-noise ratio. Most experimental parameters were the same to NCACX experiment shown in Figure 6-15 except for $^{15}\text{N} \rightarrow ^{13}\text{CO}$ CP. To match $^{15}\text{N} \rightarrow ^{13}\text{CO}$, ^{13}C offset was set to 175 ppm and ^{13}C field was 6.6-7.3 kHz. There are 32 number of points in CO dimension with spectral width 20 ppm and 36 number of points in CX dimension with spectral width 40 ppm. The spectrum was collected at 253 K with 16 kHz spinning rate and 192 scans. The spectrum was processed with the widow function QSINE, SSB = 2 for all three dimensions. It is obvious that signal-to-noise ratio of CX region of NCOCX is worse than that of NCACX.

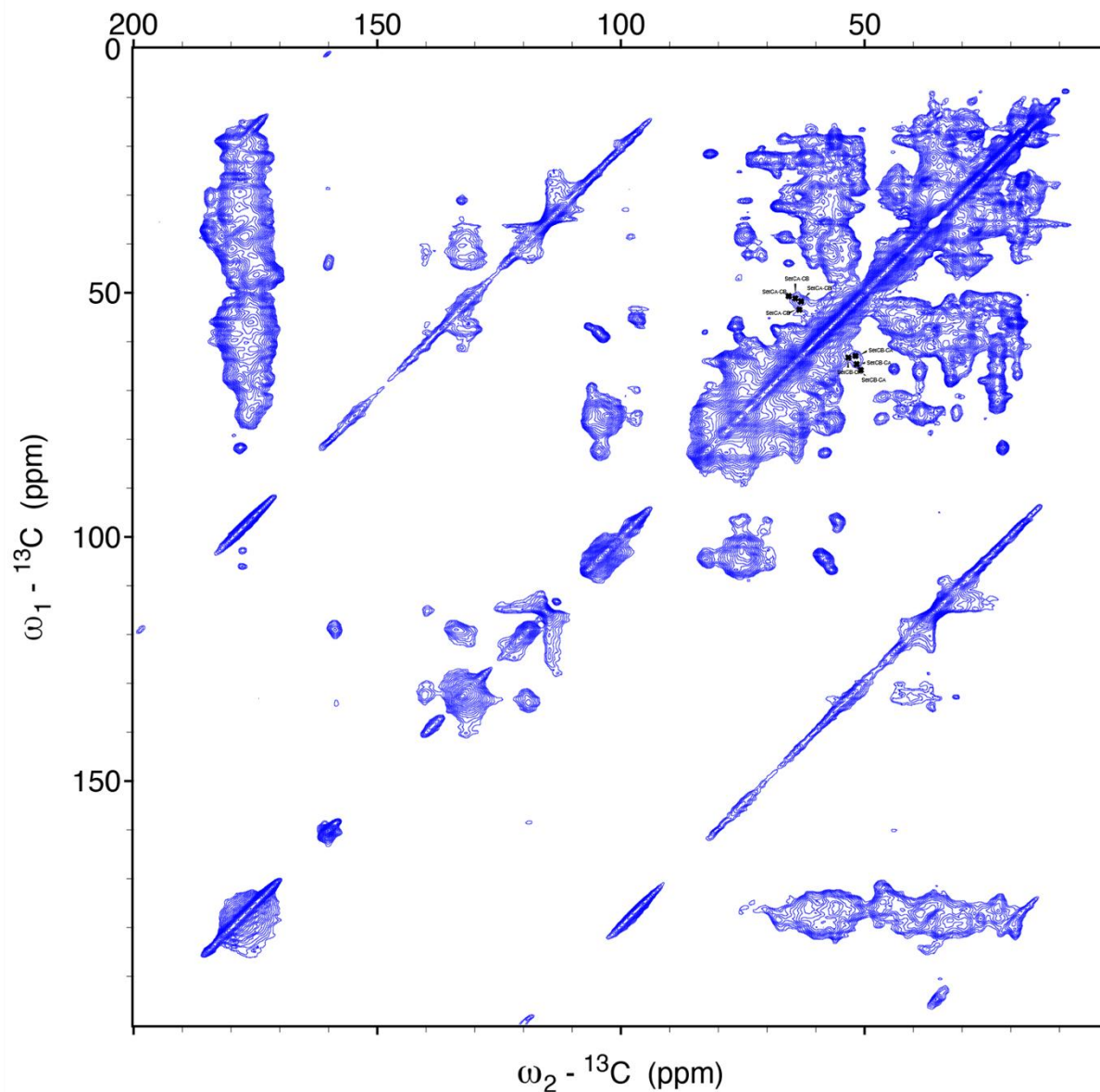
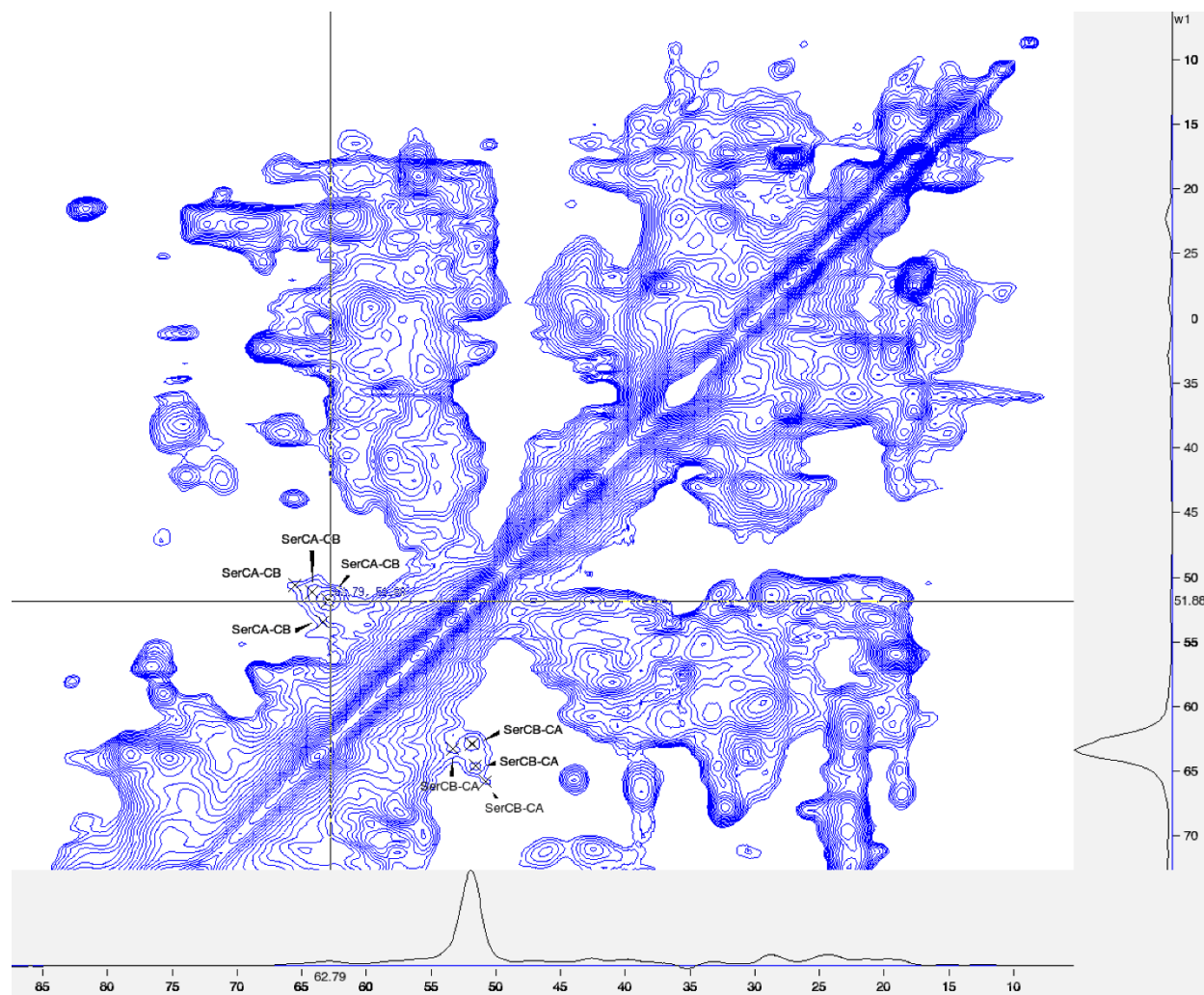


Figure 5.20. Homonuclear ^{13}C - ^{13}C correlation spectrum of Leu-Rev-HM using DARR with mixing time 30 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The numbers of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate and 128 scans. The spectrum was processed with the window function QSINE, SSB = 2 for both dimensions.

Figure 5.20 (cont'd)



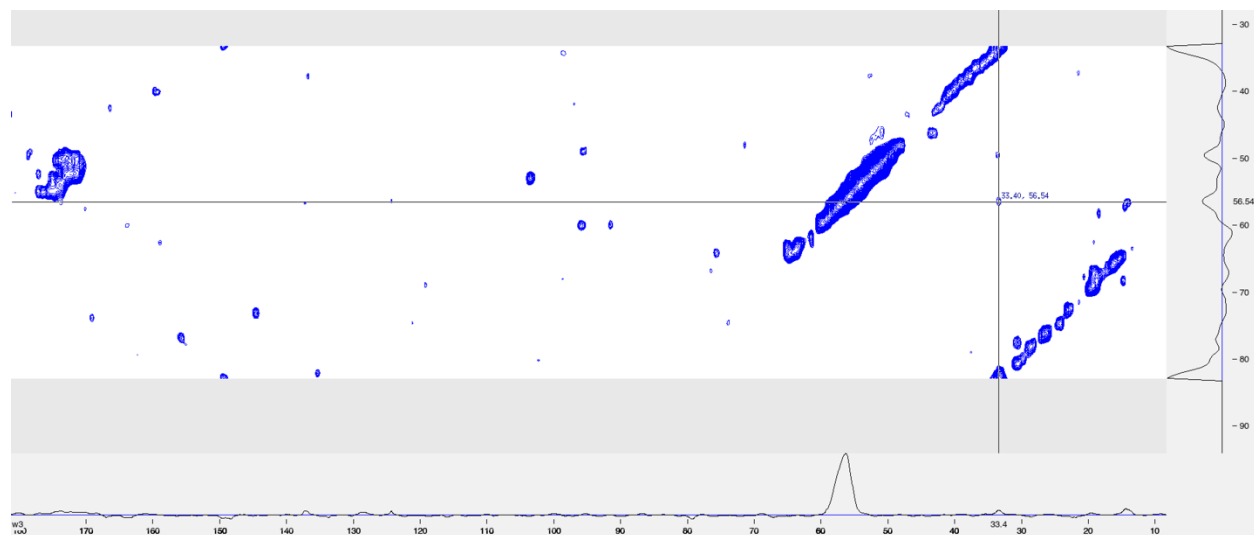


Figure 5.21. 3D NCACX spectrum of 1,3- ^{13}C -Glyc-HM edited by ^{15}N slices (^{15}N chemical shift is 121 ppm). The peak with crosshairs were randomly chosen to exhibit the 1D slice of a CB peak where readers can find the signal-to-noise ratio. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{15}\text{N}$ CP contact time 0.5 ms with 29 kHz on ^{15}N and 75-63 kHz linear CP ramp on ^1H ; $^{15}\text{N} \rightarrow ^{13}\text{C}$ CP: ^{13}C offset was set to 55 ppm and 120 ppm for ^{15}N , contact time 5ms, 35 kHz on ^{15}N and 22-24 kHz on ^{13}C , 87 kHz of ^1H decoupling applied at the same time; DARR mixing: 58 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was 81 kHz. Increment of delay (number of points) are 54 for ^{15}N and 36 for ^{13}C ; spectral width was set 50 ppm for ^{15}N and 40 ppm for ^{13}C . The data was collected at 253 K with 16 kHz spinning rate and 512 scans. The data was collected by non-uniform sampling method. The amount of sparse sampling is 25% which means that 50% data of each indirect dimension was recorded, and the data acquisition time is only a quarter of uniform sampling would cost. The T2 relaxation time of the indirect dimension was set to 0.002 s. Compressed sensing was used to replenishing missing data points of the recorded data followed by regular Fourier transform processing of the complete data set. The spectrum was processed with the widow function QSINE, SSB = 2 for direct observed dimensions CX and the indirect dimensions were processed with Gaussian Multiplication, LB = -30, GB = 0.07.

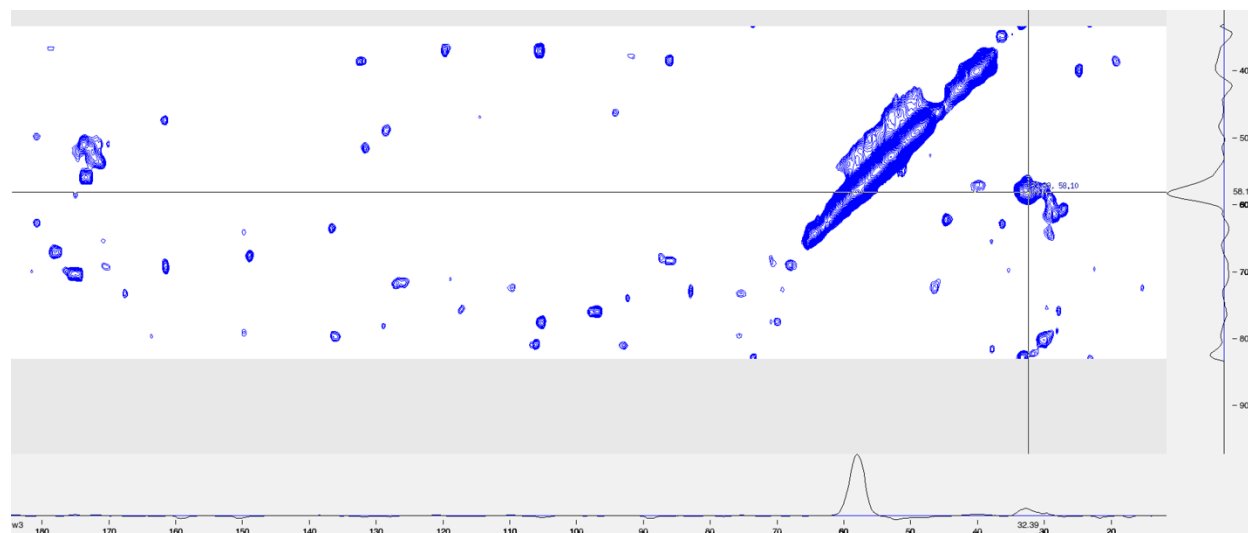


Figure 5.22. 3D NCACX spectrum of 2- ^{13}C -Glyc-HM edited by ^{15}N slices (^{15}N chemical shift is 121 ppm). The peak with crosshairs were randomly chosen to exhibit the 1D slice of a CB peak where readers can find the signal-to-noise ratio. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{15}\text{N}$ CP contact time 0.5 ms with 29 kHz on ^{15}N and 57-68 kHz linear CP ramp on ^1H ; $^{15}\text{N} \rightarrow ^{13}\text{C}$ CA CP: ^{13}C offset was set to 55 ppm and 120 ppm for ^{15}N , contact time 4.5ms, 35 kHz on ^{15}N and 21.5-23.8 kHz on ^{13}C , 90 kHz of ^1H decoupling applied at the same time; DARR mixing: 58 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was 80 kHz. Increment of delay (number of points) are 54 for ^{15}N and 36 for ^{13}C ; spectral width was set 50 ppm for ^{15}N and 40 ppm for ^{13}C . The data was collected at 253 K with 16 kHz spinning rate and 256 scans. The data was collected by 25% non-uniform sampling method. The T_2 relaxation time of the indirect dimension was set to 0.002 s. Compressed sensing was used to replenishing missing data points of the recorded data followed by regular Fourier transform processing of the complete data set. The spectrum was processed with the widow function QSINE, SSB = 2 for all dimensions.

Table 5.1. Chemical shift (ppm) of HM IBs proteins with relative peak intensity and linewidth assigned from NCACX spectra^a.

Residue Type	¹⁵ N Chemical shift/ppm	CA			CB			CG/CG1			CG2			CD/CD1			CO		
		Chemical shift/ppm	Relative Intensity	Linewidth/ppm	Chemical shift/ppm	Relative Intensity	Linewidth/ppm	Chemical shift/ppm	Relative Intensity	Linewidth/ppm	Chemical shift/ppm	Relative Intensity	Linewidth/ppm	Chemical shift/ppm	Relative Intensity	Linewidth/ppm	Chemical shift/ppm	Relative Intensity	Linewidth/ppm
Ala (5) Helix: 121.4 Sheet 124.5	121.5	55.39 (0.15)			18.18 (0.07)	1.00	1.31												
	121.5	55.25 (0.37)			17.3 (0.26)	0.95	1.31												
	123.3	55.39 (0.15)			17.08 (0.15)	0.88	1.40												
	123.3	54.36 (0.15)			17.53 (0.20)	0.42	ND												
	124.7	52.17 (0.15)			21.21 (0.07)	0.19	1.10												
	126.5	51.84 (0.15)			18.11 (0.07)	0.19	ND												
Glu (9) Helix: 119.0 Sheet 122.1	117.6	59.23 (0.07)			31.93 (0.15)	0.34	ND												
	118.7	56.41 (0.07)			29.78 (0.07)	0.69	ND												
	118.7	58.08 (0.07)			29.23 (0.15)	0.58	ND												
	121.8	54.71 (0.15)			33.83 (0.07)	0.65	ND												
	123.3	59.64 (0.11)			32.35 (0.45)	0.58	ND												
	124.6	54.32 (0.37)			33.45 (0.07)	0.56	ND												
	124.1	55.21 (0.22)			35.92 (0.15)	1.00	ND												
Gly (6) Helix: 107.5 Sheet 109.3	105	43.67 (0.22)	0.21	1.12													175.8 (0.15)	0.26	ND
	105	46.28 (0.11)	0.73	2.10													175.7 (0.08)	0.49	ND
	106	45.8 (0.18)	0.79	2.26													174.1 (0.24)	0.70	ND
	106.8	46.25 (0.18)	0.88	2.20													171.3 (0.10)	0.86	ND
	106.8	47.52 (0.09)	0.41	ND													176.3 (0.18)	0.52	ND
	106.8	47.69 (0.18)	0.39	ND													178.1 (0.10)	0.20	ND
	108.2	43.86 (0.09)	0.20	1.40													173.3 (0.15)	0.77	ND
	109.8	43.99 (0.09)	0.40	1.76													173.3 (0.15)	0.81	ND
	111.4	44.14 (0.09)	0.44	1.58													173.2 (0.18)	0.81	ND
	111.4	45.85 (0.09)	0.33	1.72													174.5 (0.09)	0.81	ND
112	45.58 (0.18)	1.00	2.05													171.9 (0.18)	1.00	ND	
Ile (9) Helix: 119.7 Sheet 122.8	118.3	59.89 (0.11)			36.83 (0.33)	0.67	ND	30.57 (0.22)	1.00	ND	22.16 (0.33)	0.81	ND						
	118.3	60.18 (0.36)			35.90 (0.33)	0.97	ND				21.36 (0.22)	0.82	1.14						
	118.3	61.39 (0.11)			36.80 (0.22)	0.87	ND				22.08 (0.33)	1.00	1.28						
	118.3	62.02 (0.22)			35.46 (0.33)	0.68	ND				21.32 (0.11)	0.97	ND						
	119.5	64.79 (0.67)			38.45 (0.11)	0.56	ND				18.05 (0.11)	0.36	ND						

Table 5.1 (cont'd)

	119.5	65.64 (0.67)	38.50 (0.22)	0.67	ND				14.38 (0.11)	0.36	0.80
	119.5	65.69 (0.09)	38.01 (0.11)	0.61	ND				17.94 (0.11)	0.35	ND
	122	59.72 (0.30)	36.06 (0.15)	1.00	ND				22.08 (0.22)	0.80	1.02
	122.2	62.58 (0.15)	37.89 (0.15)	0.45	ND				19.83 (0.22)	0.37	0.91
	122.2	60.39 (0.05)	37.50 (0.15)	0.55	ND				19.08 (0.30)	0.40	ND
	122.5	64.83 (0.05)	38.54 (0.10)	0.56	ND	29.86 (0.20)	0.35	1.40	18.13 (0.20)	0.37	0.94
	122.5	66.38 (0.40)	38.42 (0.10)	0.65	ND	29.45 (0.10)	0.35	0.80	16.74 (0.10)	0.37	0.83
	123.7	60.75 (0.44)	36.96 (0.07)	0.59	ND				20.62 (0.15)	ND	ND
	124.3	60.97 (0.44)	36.45 (0.48)	0.84	ND						
	118.7	52.65 (0.30)	42.33 (0.15)	0.58	1.09						
Leu (14)	118.7	52.5 (0.30)	39.56 (0.15)	0.37	ND						
Helix:	118.9	56.74 (0.37)	41.47 (0.30)	1.00	ND						
Sheet	119.6	54.30 (0.18)	43.71 (0.07)	0.53	ND	31.60 (0.30)	1.00	ND			
124.1	119.7	52.57 (0.52)	42.41 (0.35)	0.64	1.25	22.92 (0.22)	0.64	1.34			
	125.3	52.71 (0.37)	42.55 (0.22)	0.49	ND						
	115.4	55.71 (0.22)	42.76 (0.22)	0.77	1.16						
	115.9	57.77 (0.22)	40.89 (0.11)	0.67	1.12						
	116.5	55.24 (0.45)	40.44 (0.07)	0.34	ND						
	116.6	56.45 (0.22)	41.63 (0.11)	0.85	ND						
	116.8	56.00 (0.45)	38.52 (0.45)	1.00	1.28						
Asn (9)	117.3	55.71 (0.45)	37.60 (0.67)	0.91	ND						
Helix:	117.3	54.69 (0.22)	40.29 (0.67)	0.29	ND						
Sheet	121.5	52.72 (0.11)	42.34 (0.44)	0.71	1.18						
121.6	121.5	53.75 (0.22)	44.47 (0.11)	0.56	1.64						
	121.5	54.13 (0.11)	41.23 (0.11)	0.67	1.06						
	121.5	54.39 (0.22)	39.35 (0.67)	0.42	ND						
	122	54.22 (0.22)	43.75 (0.56)	0.69	ND						
	122.2	52.46 (0.22)	39.48 (0.45)	0.52	1.06						
Gln (14)	117.5	59.41 (0.22)	28.58 (0.22)	0.83	1.88						
Helix:	118.4	59.51 (0.22)	29.42 (0.22)	1.00	2.32	33.06 (1.25)	0.50	ND			
Sheet	121.1	59.27 (0.22)	28.65 (0.22)	0.99	2.00	32.12 (0.11)	0.53	ND			

Table 5.1 (cont'd)

Ser (6)	113.6	57.25 (0.20)	65.23 (0.10)	1.00	0.86		
Helix: 114.9	113.6	58.19 (0.10)	64.74 (0.40)	0.86	1.53		
Sheet 116.9	116.4	58.02 (0.40)	64.98 (0.30)	0.86	1.27		
Thr (5)	114.8	56.92 (0.20)	68.04 (0.10)	1.00	0.71		
Helix: 114.6	114.8	60.71 (0.10)	71.26 (0.21)	0.77	ND		
Sheet 116.5	114.8 116.7	59.65 (0.17) 64.09 (0.35)	71.58 (0.17) 69.50 (0.10)	0.87 0.67	1.27 0.67		
	119.3	65.73 (0.22)	32.25 (0.22)	0.71	1.03	22.97 (0.88)	0.59 1.52
Val (3)	119.3	67.85 (0.11)	30.83 (0.33)	0.53	ND	24.03 (0.22)	0.38 1.08
Helix: 119.2	119.3	67.85 (0.11)	31.82 (0.11)	1.00	1.16	22.34 (0.11)	1.00 1.65
Sheet 121.9	123.3 123.3	65.41 (0.22) 67.86 (0.11)	32.07 (0.22) 31.88 (0.22)	0.58 0.91	1.42 1.24	23.04 (0.22) 21.10 (0.05)	0.45 1.59 0.44 0.80
	123.3	66.75 (0.11)	31.82 (0.22)	0.96	1.24	21.10 (0.10)	0.38 1.02
	119	56.26 (0.09)	39.64 (0.32)	0.78	ND	129.4 (0.16)	0.73 ND
	119	57.01 (0.48)	39.74 (0.64)	0.85	ND	129.7 (0.08)	0.74 ND
Tyr (2)	123.1	57.39 (0.48)	41.02 (0.32)	0.99	ND	131.4 (0.16)	1.00 1.05
Helix: 119.2	123.1	57.67 (0.16)	40.34 (0.16)	1.00	ND	131.3 (0.08)	1.00 0.93
Sheet 121.4	126.1	55.69 (0.05)	42.63 (0.57)	0.52	ND	131.0 (0.09)	0.41 ND
	126.1	56.61 (0.16)	41.92 (0.43)	0.59	ND	131.8 (0.16)	0.74 ND
	126.1	57.56 (0.09)	42.57 (0.57)	0.41	ND	131.0 (0.05)	0.94 ND

a. Available uncertainties are shown in parenthesis. The uncertainty was determined by the range of the chemical shift where the peak locates in the 1D slice of the cross peak. Linewidth is determined by full width at half maximum of a peak.

b. Residue numbers of each amino acid type is provided in parentheses under each residue type.

c. The statistically derived reference of ^{15}N chemical shift for helix and sheet is provided as helix and sheet in the first column.³⁸

Table 5.2. Secondary shift table of HM^a.

Residue Type	¹⁵ N/ppm	CA/ppm	CB/ppm	CA-CB/ppm	(CA- CB) _{ref,coil}	Secondary shift/ppm
Ala	121.5	55.25(0.37)	17.30(0.26)	37.95(0.45)	33.7	4.25
	121.5	55.39(0.15)	18.18(0.07)	37.21(0.17)		3.51
	123.3	54.36(0.15)	17.53(0.20)	36.83(0.25)		3.13
	123.3	55.39(0.15)	17.08(0.15)	38.31(0.21)		4.61
	124.7	52.17(0.15)	21.21(0.07)	30.96(0.17)		-2.74
	126.5	51.84(0.15)	18.11(0.07)	33.73(0.17)		0.03
Glu	117.6	59.23(0.07)	31.93(0.15)	27.30(0.17)	26.7	0.6
	118.7	56.41(0.07)	29.78(0.07)	26.63(0.10)		-0.07
	118.7	58.08(0.07)	29.23(0.15)	28.85(0.17)		2.15
	121.8	54.71(0.15)	33.83(0.07)	20.88(0.17)		-5.82
	123.3	59.64(0.11)	32.35(0.45)	27.29(0.46)		0.59
	124.1	55.21(0.22)	35.92(0.15)	19.29(0.27)		-7.41
Ile	124.6	54.32(0.37)	33.45(0.07)	20.87(0.38)	-5.83	
	118.3	59.89(0.11)	36.83(0.33)	23.06(0.35)	22.3	0.76
	118.3	60.18(0.36)	35.90(0.33)	24.28(0.49)		1.98
	118.3	61.39(0.11)	36.80(0.22)	24.59(0.25)		2.29
	118.3	62.02(0.22)	35.46(0.33)	26.56(0.4)		4.26
	119.5	64.79(0.67)	38.45(0.11)	26.34(0.68)		4.04
	119.5	65.64(0.67)	38.50(0.22)	27.14(0.71)		4.84
	119.5	65.69(0.09)	38.01(0.11)	27.68(0.14)		5.38
	122.0	59.72(0.30)	36.06(0.15)	23.66(0.34)		1.36
	122.2	60.39(0.05)	37.50(0.15)	22.89(0.16)		0.59
	122.2	62.58(0.15)	37.89(0.15)	24.69(0.21)		2.39
	122.5	64.83(0.05)	38.54(0.10)	26.29(0.11)		3.99
	122.5	66.38(0.40)	38.42(0.10)	27.96(0.41)		5.66
	123.7	60.75(0.22)	36.96(0.07)	23.79(0.23)		1.49
124.3	60.97(0.44)	36.45(0.48)	24.52(0.65)	2.22		

Table 5.2 (cont'd)

	118.7	52.50(0.30)	39.56(0.15)	12.94(0.34)		0.44
	118.7	52.65(0.30)	42.33(0.15)	10.32(0.34)		-2.18
Leu	118.9	56.74(0.37)	41.47(0.30)	15.27(0.48)	12.5	2.77
	119.0	54.30(0.18)	43.71(0.07)	10.59(0.19)		-1.91
	119.7	52.57(0.52)	42.41(0.35)	10.16(0.63)		-2.34
	125.3	52.71(0.37)	42.55(0.22)	10.16(0.43)		-2.34
	115.4	55.71(0.22)	42.76(0.22)	12.95(0.31)		-1.65
	115.9	57.77(0.22)	40.89(0.11)	16.88(0.25)		2.28
	116.5	55.24(0.45)	40.44(0.07)	14.80(0.46)		0.2
	116.6	56.45(0.22)	41.63(0.11)	14.82(0.25)		0.22
	116.8	56.00(0.45)	38.52(0.45)	17.48(0.64)		2.88
	117.3	55.71(0.45)	37.60(0.67)	18.11(0.81)		3.51
Asn	117.5	54.69(0.22)	40.29(0.67)	14.40(0.71)	14.6	-0.2
	121.5	52.72(0.11)	42.34(0.44)	10.38(0.45)		-4.22
	121.5	53.75(0.22)	44.47(0.11)	9.28(0.25)		-5.32
	121.5	54.13(0.11)	41.23(0.11)	12.90(0.16)		-1.7
	121.5	54.39(0.22)	39.35(0.67)	15.04(0.71)		0.44
	122.0	54.22(0.22)	43.75(0.56)	10.47(0.60)		-4.13
	122.2	52.46(0.22)	39.48(0.45)	12.98(0.50)		-1.62
	117.5	59.41(0.22)	28.58(0.22)	30.83(0.31)		
Gln	120.9	59.27(0.22)	28.65(0.22)	30.62(0.31)	27	3.62
	120.9	59.51(0.22)	29.42(0.22)	30.09(0.31)		3.09
	113.6	57.25(0.20)	65.23(0.10)	-7.98(0.22)		
Ser	113.6	58.19(0.10)	64.74(0.40)	-6.55(0.41)	-5.6	-0.95
	116.4	58.02(0.40)	64.98(0.30)	-6.96(0.50)		-1.36
	114.8	56.92(0.20)	68.04(0.10)	11.12(0.22)		
Thr	114.8	59.65(0.17)	71.58(0.17)	11.93(0.24)	-8.5	-3.43
	114.8	60.71(0.10)	71.26(0.21)	10.55(0.23)		-2.05
	116.7	64.09(0.35)	69.5(0.10)	-5.41(0.36)		3.09

Table 5.2 (cont'd)

	119.3	65.73(0.22)	32.25(0.22)	33.48(0.31)		4.08
	119.3	67.85(0.11)	30.83(0.33)	37.02(0.35)		7.62
Val	119.3	67.85(0.11)	31.82(0.11)	36.03(0.16)	29.4	6.63
	123.3	65.41(0.22)	32.07(0.22)	33.34(0.31)		3.94
	123.3	66.75(0.11)	31.82(0.22)	34.93(0.25)		5.53
	123.3	67.86(0.11)	31.88(0.22)	35.98(0.25)		6.58
	119.0	56.26(0.09)	39.64(0.32)	16.62(0.33)		-2.38
	119.0	57.01(0.48)	39.74(0.64)	17.27(0.80)		-1.73
	123.1	57.39(0.48)	41.02(0.32)	16.37(0.58)		-2.63
Tyr	123.1	57.67(0.16)	40.34(0.16)	17.33(0.23)	19.0	-1.67
	126.1	55.69(0.05)	42.63(0.57)	13.06(0.57)		-5.94
	126.1	56.61(0.16)	41.92(0.43)	14.69(0.46)		-4.31
	126.1	57.56(0.09)	42.57(0.57)	14.99(0.58)		-4.01

^a Available uncertainties are shown in parentheses.

Table 5.3. The deviation of (CA+CO) chemical shift of Gly of HM from the referenced Gly in a random coil structure.

	¹⁵ N/ ppm	CA/ ppm	CO/ ppm	CA _{ref.coil} / ppm	CO _{ref.coil} / ppm	(CA+CO)- (CA _{ref.coil} +CO _{ref.coil})/ppm
	105.0	43.67(0.22)	175.8(0.15)			0.07
	108.2	43.86(0.09)	173.3(0.15)			-2.24
	109.8	43.99(0.09)	173.3(0.15)			-2.11
	111.4	44.14(0.09)	173.2(0.18)			-2.06
	112.0	45.58(0.18)	171.9(0.18)			-1.92
Gly	106.0	45.80(0.18)	174.1(0.24)	45.5	173.9	0.50
	111.4	45.85(0.09)	174.5(0.09)			0.95
	106.8	46.25(0.18)	171.3(0.10)			-1.85
	105.0	46.28(0.11)	175.7(0.08)			2.58
	106.8	47.52(0.09)	176.3(0.18)			4.42
	106.8	47.69(0.18)	178.1(0.10)			6.39

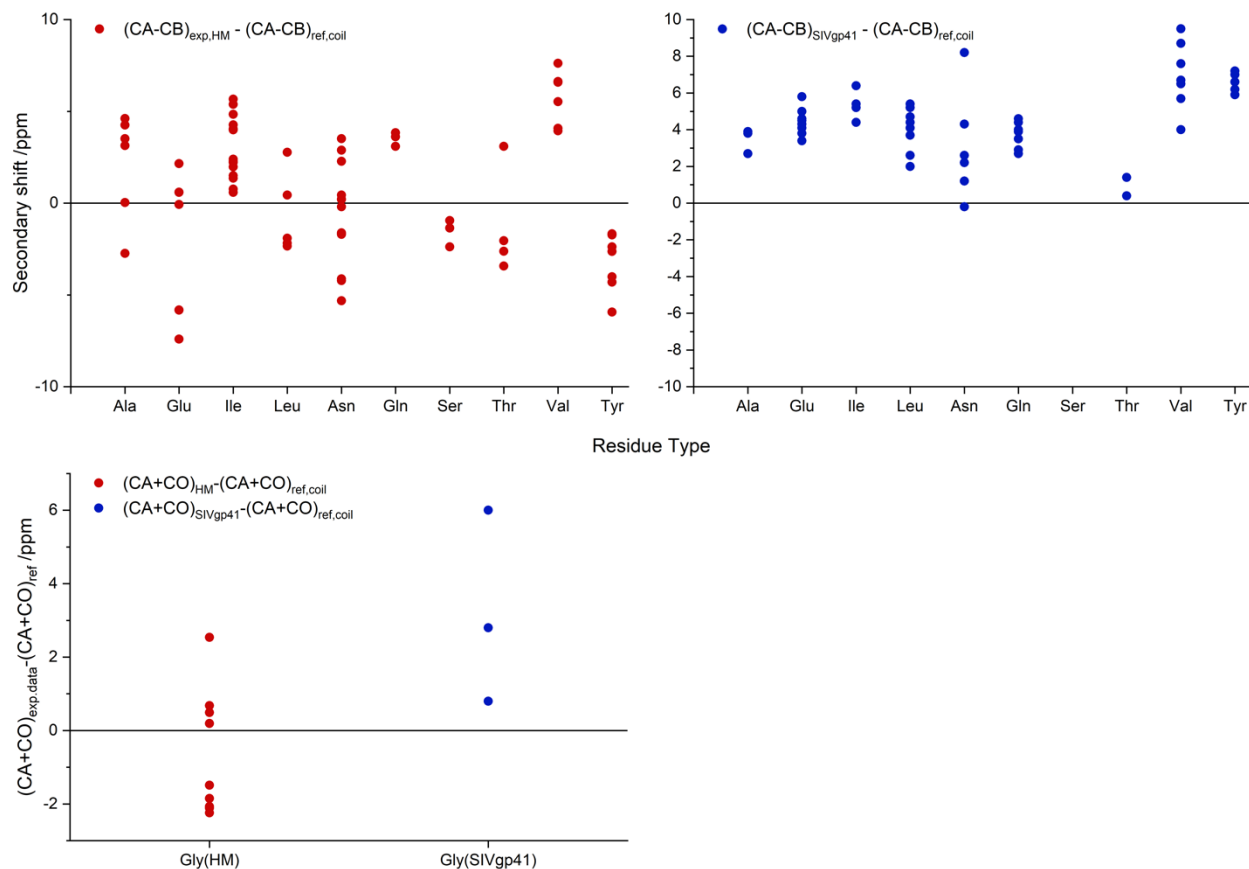


Figure 5.23. Top: comparison of the secondary shift between HM IBs proteins and ectodomain of gp41 of the Simian immunodeficiency virus evaluated from solution NMR. Bottom: The deviation of (CA+CO) of HM and SIV gp41 from that of referenced random coil chemical shift.

5.3 Discussion

5.3.1 Feasibility of purify protocol

It is clear that the wash buffer separated the HM IBs from undesired components, supported by MALDI-mass spectrum and 57% sequence coverage from proteomic analysis. Even though there exist some impurities in the pellet after lysis with wash buffer, the majority product is HM IBs, supported by the dominated HM trimer band of SDS-PAGE gel. Given the fact that unlabeled medium was switched to labeled ones before E.coli started expressing proteins, the ^{13}C and ^{15}N labeled source were used to express HM. Thus, only HM proteins should be labeled, and impurities should not have a substantial signal in ssNMR. Based on light microscopy and TEM images, HM samples subjected to wash buffer sonication render distinctive morphologies. The amyloid-like structure is observed for the sample with wash buffer purification. The major functional reagents in wash buffer are Triton X-100, a detergent used for solubilizing membrane proteins. The SDS-

PAGE gel implies that the composition of HM samples with and without wash buffer purification are similar and the wash buffer make HM component more dominated. Consequently, the wash buffer removed soluble membrane proteins, so the HM sample was left to be easily observed.

5.3.2 Structural information based on ssNMR data analysis

From secondary shift analysis, HM IBs protein consists of major positive secondary shift mixed with minor negative secondary shift. It illustrates that the large fraction of the protein is α -helical whereas there is a small fraction is β -sheet due to the fact that for each residue type except Tyrosine, there are relative more populations of α -helical conformation. Compared with SIV gp41 which is pure α -helices evidenced by its almost completely positive secondary shift plot, the secondary shift analysis of HM IBs may support that IBs protein of HM have more than one secondary structure. What worth to be mentioned is that the cross peak intensity of alanine in β -sheet is weaker than that of α -helix. The α helix: β -sheet ratio is estimated to be 9:1 based on ratio between the integral of AlaN-CA-CB peak in α helix structure ($^{15}\text{N}=121.5\text{ppm}, 123.3\text{ppm}$ and the peak in β -sheet conformation ($^{15}\text{N}=124.7\text{ppm}$) (Figure 5.18). Additionally, there exist four cross peaks of alanine matching α -helix and two for β -sheet. The current peak summary listed in Table 5.1 is the assignment from NCACX spectra without sequential assignment from NCOCX and CONCACX spectra. Since only the amino acid type has been determined and no information about sequential assignment can be extracted from NCACX data, we cannot confirm whether a peak in NCACX spectrum is from the same amino acid type at a different position in the same protein chain or the same residue of a protein chain with a different secondary structure. Overall, the assignment implies that the major fraction of HM IBs proteins matches α helix, and the minor fraction is β sheet.

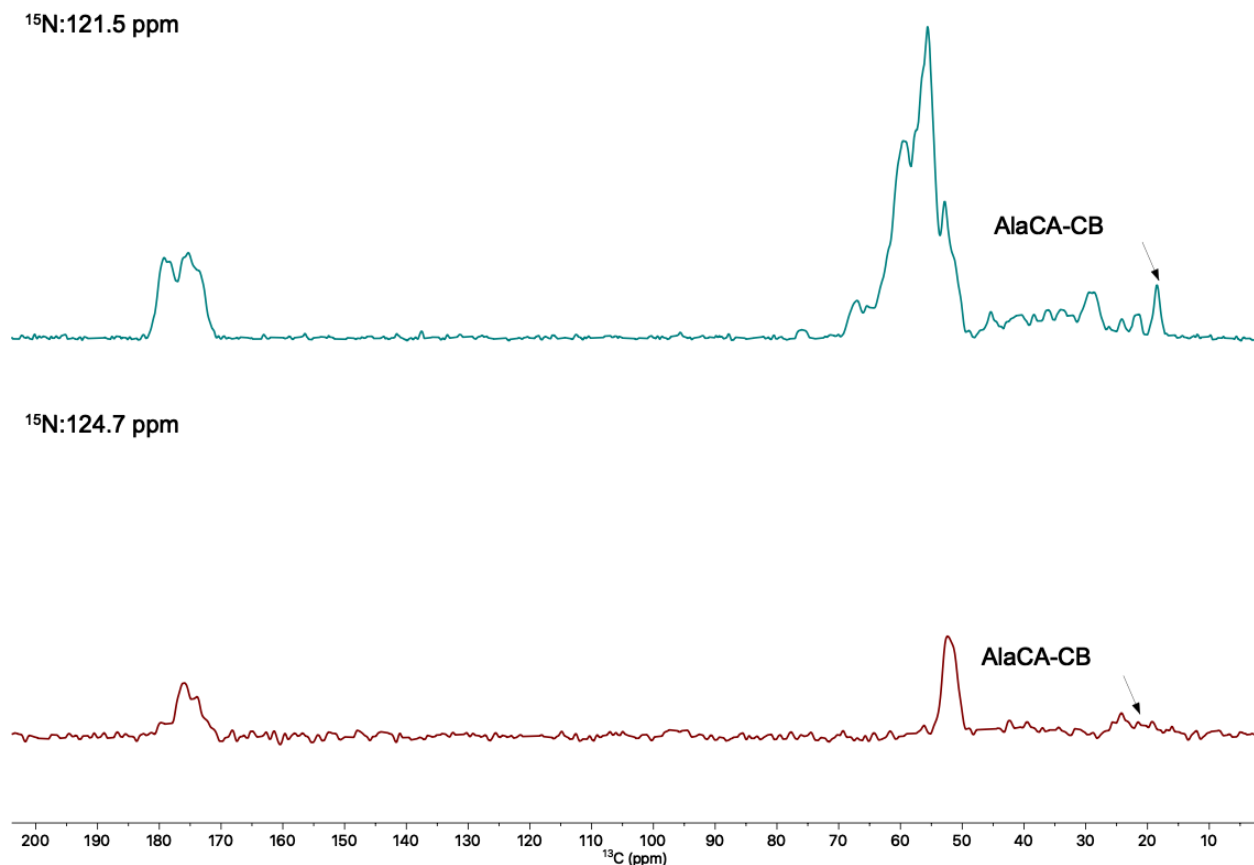


Figure 5.24. The extract 1D slice of Figure 5.15 corresponding to AlaCA-CB peak at 121.5 ppm and 124.7 ppm. The peak has been integrated to estimate the ratio between α helical and β sheet conformation. The integrated regions are 15.42-20.17 ppm and 19.85-22.27 ppm for the top and bottom spectrum respectively.

Due to the extensive peak overlap in spectra, only the well-resolved resonance can be assigned unambiguously. There are many residues assigned without a ^{13}CO chemical shift because of the extensively spectral overlap in carbonyl region. Even though the residue type can be determined from sidechain carbons without carbonyl carbon, the ^{13}CO chemical shift is critical for sequential assignment.

Previous studies from Weliky group indicated that the Fgp41 IBs (Fgp41 = Fp+gp41) adopts α helical structure and has a distribution of α helix. The Fgp41 IBs was site-specially labeled, and the expected residues were labeled by feeding the E.coli with labeled amino acid before expression. By comparing the chemical shift of ^{13}CO and the peak width of the sample with the statistical data of a typical α -helical protein, the author concluded that Fgp41 IBs adopt native secondary structure in inclusion bodies formation with a distribution of α -helical conformation. On the contrary, our data supports that the HM as IBs formation consists of a big fraction of α helix with a small fraction

of β sheet, which is partly consistent with the previous findings. The reason for the difference could be that β sheet made a contribution but the signal could not be resolved. The author pointed out that Gly linewidth is relatively broad (~6 ppm) probably due to a distribution of conformation at these positions. Besides, the assigned chemical shifts only represent the local secondary structure. It is not surprising that the conformation of Fgp41 could be a mixture of different secondary structures. In the present project, the full region of gp41 was either uniformly labeled or sparsely labeled so each residue has a contribution to the spectrum theoretically. The interpreted structural information should reflect the structure of the full gp41 region. What should be mentioned is that only the chemical shift of Gly, Leu and Ser of gp41 region were reported by Curtis-Fisk, and Gly-38, Ser-66 and Ser-70 have two distinctive ^{13}C O chemical shifts. For the two chemical shifts of Gly ^{13}C O, 173.2 ppm matches β sheet statistical data and 177.7 ppm matches α -helix.⁶ Vogel et al did the REDOR measurements for lyophilized 1- ^{13}C , ^{15}N -Leu labeled cells induced to produce Fgp41 which most Fgp41 should be inclusion bodies proteins. The ^{13}C O peak was consisted of contribution from α helix and β -sheet with a ratio ~0.85:0.15.¹⁴

The chemical shift consistent with β sheet statistical data is evidenced by the present data as well. We found there are several residues type (Ala, Glu, Gly, Ile, Asn, Tyr) having one or more chemical shift distribution associated with β -sheet. Due to the lack of resolution of ^{13}C O region, we were not able to extract the chemical shift about the ^{13}C O for residue type other than Gly. Interestingly, the ^{13}C O chemical shift of Gly not only supports β sheet but also supports the random coil structure whose typical chemical shift is 173.9 ppm. Additionally, Ala exhibits some ^{15}N peaks matching the chemical shift of random coil secondary structure with a narrow linewidth around 1 ppm. However, the linewidth of random coil structure is usually broader than that of α helix and β sheet for the reason that disordered protein has a broad conformational distribution giving rise to multiple isotropic chemical shifts.³⁹⁻⁴¹ Based on the HM assignment, the linewidth of the cross peaks of either α helix or β sheet is around 1 ppm. The narrow linewidth of the cross peaks whose ^{15}N chemical shift matches random coil structure indicates that those peaks are not from random coil and Ala is very likely to adopt another conformation which is neither α helix nor β sheet. We also noticed that each residue type has more than one chemical shift consistent with α -helical structure and could be attributed two possibilities: (1) there exists a mixture of α helical HM molecules and some of the HM molecules have a distinguished chemical shift; (2) the assigned

chemical shift reflects that a HM molecule is a mixture structure of major α helix and minor β sheet. (3) Consider the HM proteins are packed in IBs, it is possible that the region packed in the core of the aggregated proteins has different chemical shift than the region which is at the surface and close to the water.

The ratio between α helix and β sheet is estimated from Table 5.1. The product of relative peak intensity and linewidth of N-CA-CB peak is calculated, and its fraction is used to estimate the ratio of the population between α helix and β sheet conformation.(shown in Table 5.4) The reasons for choosing N-CA-CB peak for estimation are: (1) almost every amino acid type has CB except for glycine; (2) CB is close to CA so there should be relative stronger intensity from CA spin diffusion for CB peaks compared with other side chain carbons. Table 5.4 is a summary of ratio between helix and sheet when assigned helical and sheet's linewidth are both available. The population fraction of helix is underestimated due to there exist several assigned helical N-CA-CX peaks with a not-determined linewidth as the peak locates at the superpositioned region. From Table 5.4, there are β sheet population fraction for Ala and Asn. It is likely that the C-helices adopt β sheet conformation in IBs. It's evidenced that N-helices is stable, and the N-helical coiled coil can be formed at suboptimal temperature.^{42,43} On the contrary, the heptad repeat region at C-terminus will no longer be helical conformation without the presence of N-helices.

Table 5.4. The estimated ratio between α helix and β sheet conformation^a.

Residue Type	¹⁵ N Chemical shift (ppm)	CA Chemical shift (ppm)	CB Chemical shift (ppm)	Relative Peak Intensity	Linewidth (ppm)	Helix:Sheet	
Ala	Helix	121.5	18.18	1.00	1.31	0.93:0.07	
		121.5	17.3	0.95	1.31		
		123.3	17.08	0.88	1.4		
	Sheet	124.7	21.21	0.31	0.95		
Leu	Helix	118.7	42.33	0.58	1.09	1:0	
		119.7	42.41	0.64	1.25		
Asn	Helix	115.4	42.76	0.77	1.16	0.91:0.09	
			115.9	40.89	0.67		1.12
			116.8	38.52	1.00		1.28
	Sheet		121.5	42.34	0.71		1.18
			121.5	44.47	0.56		1.64
			121.5	41.23	0.67		1.06
			122.2	39.48	0.45		1.14
Gln	Helix	117.5	28.58	1.16	1.36	1:0	
		120.9	29.42	1.00	2.32		
		120.9	28.65	0.99	2.00		
Ser	Helix	113.6	65.23	1.00	0.86	1:0	
		113.6	64.74	0.86	1.53		
		116.4	64.98	0.86	1.27		
Thr	Helix	114.8	68.04	1.00	0.71	0.8:0.2	
		114.8	71.58	0.87	1.27		
	Sheet	116.7	69.5	0.67	0.67		
Val ^b	Helix	119.3	32.25	0.71	1.03	1:0	
			119.3	31.82	1.00		1.16
			123.3	32.07	0.58		1.42
			123.3	31.88	0.91		1.24
			123.3	31.82	0.96		1.24

^a All CB peaks with a determined linewidth in Table 5.1 has been included in this table.

^b The statistical ¹⁵N chemical shift data of valine is 119.77 (5.45) for α helix and 121.90 (5.05) for β sheet. Given the large deviation and the secondary shift of valine are positive in Figure 5.23, all CB peaks are considered as β sheet.⁴⁴

The TEM image of HM IBs purified by PBS and wash buffer supported β sheet structure. The width of each fibril is ~ 5 nm within the typical range of amyloid. However, the amyloid couldn't be strongly evidenced by Congo red staining experiment. But the small orange spot supported there is a possibility that amyloid exists. Combining all the evidence including the existence of cross peaks matching α -helix and β sheet structure, and the fibril observed in TEM, it is safe to conclude that the overall conformation of HM IBs is mixture of major α -helix and minor β sheet. Overall, the data analysis of the multidimensional spectra evidenced that the major conformation of HM as insoluble fractions is α -helix, consistent with the native structure of gp41. The minor β sheet portion is a unique feature of HM IBs since the purified HM in solution state is N-helices and C-helices connected by a loop. The existence of β sheet in IBs formation may be due to the perturbation of cell growth. The secondary structure characterization of beta-lactamase IBs via Raman spectroscopy suggests that the growth condition of cell to produce IBs perturbs the secondary structure of the protein relative to the native protein in solution by different extents. The higher temperature exhibited a more substantial perturbation, and the content of beta sheet became greater than low temperature growth.⁵ The higher expression temperature, more IPTG and longer expression time may cause a perturbation to cell growth resulting in the occurrence of β sheet conformation since the overexpress condition increases the protein expression rate and misfolding/formation of intermolecular β sheet would occur due to lack of time for protein folding.

5.3.3 Future work

Although we were able to assign some of the crosspeaks of amino acid type, we don't know the accurate population distribution of α -helix and β sheet since the current estimation only includes the well-resolved crosspeak with determined linewidth. Many unresolved crosspeaks whose linewidth is difficult to determine due to superposition are not included in the estimation of α helix and β sheet intensity.

We don't know the information about the secondary structure of the specific region without sequential assignment because of the lack of the chemical shift dispersion in the ^{13}C O region. The spread of ^{13}C O chemical shift is ~ 170 - 180 ppm and it is impossible to recognize a large number of ^{13}C O in a narrow chemical shift range. Site specific labeling might be an alternative method to obtain some ^{13}C O chemical shifts. For example, Met, Lys and Phe are the good candidate to be labeled because there are only a few repeating residues of each type. There is one methionine, one

phenylalanine and four lysine in HM sequence.

The signal corresponding to aromatic region are very weak in the processed spectra, which could be attributed to the attenuated signal intensity transferred from CA. The magnetization coming from CA distributes to six or more aromatic peaks, so the intensity is attenuated and more number of scans is needed.

The 1,3-¹³C-Glyc-HM displays unexpected weak cross peak intensity and limited cross peaks, possibly due to the protein were not effectively labeled. It should have a stronger peak intensity than 2-¹³C-Glyc-HM as the result of more carbons should be labeled by employing 1,3-¹³C-Glycerol as isotope source. It is worthy to make an effectively 1,3-¹³C-Glycerol labeled HM sample. As a complementary ssNMR data of 2-¹³C-Glyc-HM, assignment of NCACX spectra of 1,3-¹³C-Glyc-HM and 2-¹³C-Glyc-HM together would be an efficient and less tedious method to assign chemical shifts. The protein extracted from the cells grown in either 1,3-¹³C-glycerol or 2-¹³C-glycerol as the solely carbon supplement would follow different labelling patterns. If the CA is not labeled, then the CO signal would transfer to the nearest labeled CX. By assigning the NCACX and NCOCX spectra of the 1,3-¹³C-glycerol or 2-¹³C-glycerol labeled sample together, the residue type can be determined.^{32,45,46}

Even though many chemical shifts of the side chains have been recognized, there still exist many partial superimposed peaks that cannot be assigned. It is possible that there is a distribution of HM structure and/or the electronic environment around the ¹⁵N and ¹³C nucleus is different than one another resulting in broad linewidth of ¹⁵N and ¹³C peak. The inadequate decoupling could be another reason for the superimposed region. The development of fast MAS rate (up to 150 kHz today) at ultrahigh magnetic field might be another alternative structural determination instead of traditional ¹³C detective methods. The detection of ¹³C and ¹⁵N needs a strong ¹H dipolar decoupling to recover narrow lines and the detection sacrifices sensitivity and the long experimental times are required. Take a coupled ¹³C-¹H spin pair as an example, the external magnetic field ¹³C spin experienced is influenced by the ¹H spin. The spin-up and spin-down state of ¹H would either increase or decrease the effective local field of ¹³C spin, resulting in the changes of the resonance frequency compared with uncoupled ¹³C. ¹H-¹³C dipolar decoupling would eliminate the line broadening from fluctuating of ¹³C effective local field caused by ¹H. Figure 5.25 shows the peak of glycine powder is narrower when ¹H-¹³C decoupling is applied.⁴⁷

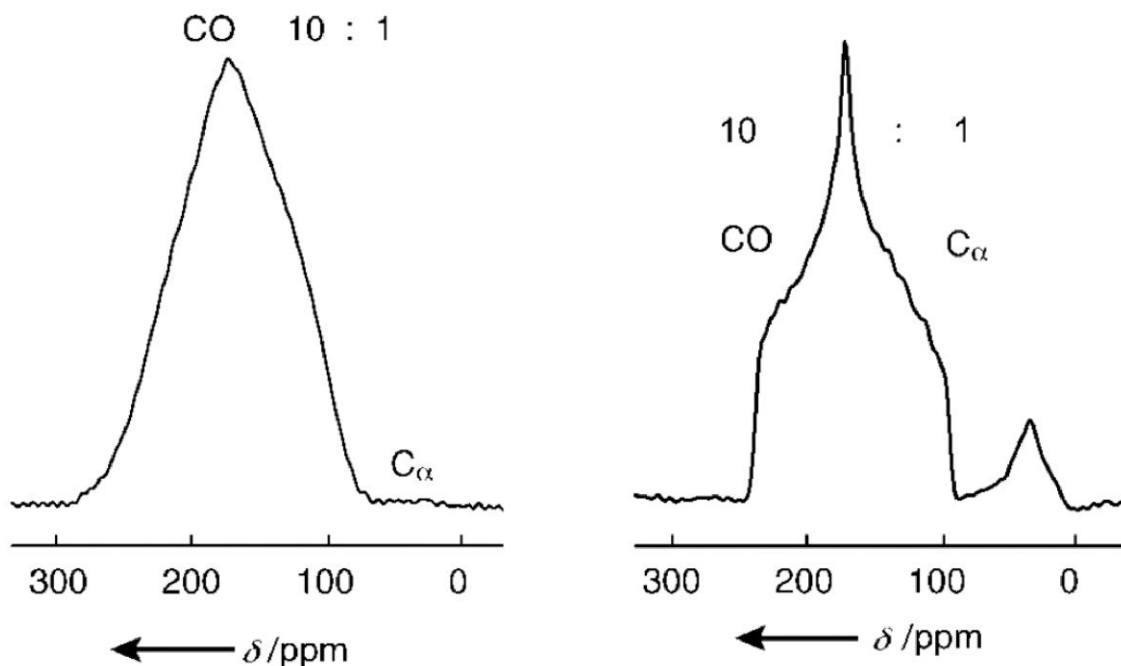


Figure 5.25. Solid-state ^{13}C spectra (125 MHz) of a $1\text{-}^{13}\text{C}$ -labeled (10%) glycine powder (90 mg) with and without ^1H - ^{13}C decoupling. Left: spectrum without decoupling; Right: spectrum with continuous wave (CW) decoupling. The broadening of the signal for C_α gives an imperceptible peak in the left spectrum.

Detection of ^1H would benefit not only from its high gyromagnetic ratio but also an enriched natural abundance. Nevertheless, protons of proteins sample always generate a network of strong dipolar coupling resulting in a severe line broadening at common MAS rates (10-20 kHz). The fast MAS rate at ultrahigh field would suppress the strong dipolar coupling network of protons so the linewidth would be narrower and peak intensity would be increased. ^1H correlation spectra (3D HN(H)-H and 4D HC(H)-(H)CH, etc) could provide more structural information.⁴⁸⁻⁵¹

REFERENCES

- (1) Bhatwa, A.; Wang, W.; Hassan, Y. I.; Abraham, N.; Li, X. Z.; Zhou, T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in *Escherichia Coli* and Strategies to Address Them for Industrial Applications. *Front Bioeng Biotechnol* 2021, 9 (February), 1–18. DOI: 10.3389/fbioe.2021.630551.
- (2) Ventura, S.; Villaverde, A. Protein Quality in Bacterial Inclusion Bodies. *Trends Biotechnol* 2006, 24 (4), 179–185. DOI: 10.1016/j.tibtech.2006.02.007.
- (3) Singh, S. M.; Panda, A. K. Solubihzation and Refolding of Bacterial Inclusion Body Proteins. *J Biosci Bioeng* 2005, 99 (4), 303–310. DOI: 10.1263/jbb.99.303.
- (4) Tsumoto, K.; Ejima, D.; Kumagai, I.; Arakawa, T. Practical Considerations in Refolding Proteins from Inclusion Bodies. *Protein Expr Purif* 2003, 28 (1), 1–8. DOI: 10.1016/S1046-5928(02)00641-1.
- (5) Przybycien, T. M.; Dunn, J. P.; Valax, P.; Georglou, G. Secondary Structure Characterization of SS-Lactamase Inclusion Bodies. *Protein Engineering, Design and Selection* 1994, 7 (1), 131–136. DOI: 10.1093/protein/7.1.131.
- (6) Curtis-Fisk, J. Structural Studies of the Influenza and HIV Viral Fusion Proteins and Bacterial Inclusion Bodies, Michigan state university, 2009.
- (7) Ramazi, S.; Zahiri, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database* 2021, 2021 (7), 1–20. DOI: 10.1093/database/baab012.
- (8) Walsh, G. Post-Translational Modifications of Protein Biopharmaceuticals. *Drug Discov Today* 2010, 15 (17–18), 773–780. DOI: 10.1016/j.drudis.2010.06.009.
- (9) Cain, J. A.; Solis, N.; Cordwell, S. J. Beyond Gene Expression: The Impact of Protein Post-Translational Modifications in Bacteria. *J Proteomics* 2014, 97, 265–286. DOI: 10.1016/j.jprot.2013.08.012.
- (10) Forrest, S.; Welch, M. Arming the Troops: Post-Translational Modification of Extracellular Bacterial Proteins. *Sci Prog* 2020, 103 (4), 1–22. DOI: 10.1177/0036850420964317.
- (11) Rinas, U.; Bailey, J. E. Protein Compositional Analysis of Inclusion Bodies Produced in Recombinant *Escherichia Coli*. *Appl Microbiol Biotechnol* 1992, 37 (5), 609–614. DOI: 10.1007/BF00240735.
- (12) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in *Escherichia Coli*: Advances and Challenges. *Front Microbiol* 2014, 5 (APR), 1–17. DOI: 10.3389/fmicb.2014.00172.
- (13) Vogel, E. P.; Weliky, D. P. Quantitation of Recombinant Protein in Whole Cells and Cell Extracts via Solid-State NMR Spectroscopy. *Biochemistry* 2013, 52 (25), 4285–4287. DOI: 10.1021/bi4007034.
- (14) Vogel, E. P.; Curtis-fisk, J.; Young, K. M.; Weliky, D. P. Erica P. Vogel, Jaime Curtis-Fisk, Kaitlin M. Young, and David P. Weliky *. *Biochemistry* 2011, No. 50, 10013–10026.

- (15) Carrió, M. M.; Corchero, J. L.; Villaverde, A. Dynamics of in Vivo Protein Aggregation: Building Inclusion Bodies in Recombinant Bacteria. *FEMS Microbiol Lett* 1998, *169* (1), 9–15. DOI: 10.1016/S0378-1097(98)00444-3.
- (16) Singhvi, P.; Saneja, A.; Srichandan, S.; Panda, A. K. Bacterial Inclusion Bodies: A Treasure Trove of Bioactive Proteins. *Trends Biotechnol* 2020, *38* (5), 474–486. DOI: 10.1016/j.tibtech.2019.12.011.
- (17) García-Fruitós, E. Inclusion Bodies: A New Concept. *Microb Cell Fact* 2010, *9*, 2010–2012. DOI: 10.1186/1475-2859-9-80.
- (18) García-Fruitós, E.; Vázquez, E.; Díez-Gil, C.; Corchero, J. L.; Seras-Franzoso, J.; Ratera, I.; Veciana, J.; Villaverde, A. Bacterial Inclusion Bodies: Making Gold from Waste. *Trends Biotechnol* 2012, *30* (2), 65–70. DOI: 10.1016/j.tibtech.2011.09.003.
- (19) Wang, L. Towards Revealing the Structure of Bacterial Inclusion Bodies. *Prion* 2009, *3* (3), 139–145. DOI: 10.4161/pri.3.3.9922.
- (20) Gil-Garcia, M.; Navarro, S.; Ventura, S. Coiled-Coil Inspired Functional Inclusion Bodies. *Microb Cell Fact* 2020, *19* (1), 1–16. DOI: 10.1186/s12934-020-01375-4.
- (21) Curtis-Fisk, J.; Spencer, R. M.; Weliky, D. P. Native Conformation at Specific Residues in Recombinant Inclusion Body Protein in Whole Cells Determined with Solid-State NMR Spectroscopy. *J Am Chem Soc* 2008, *130* (38), 12568–12569. DOI: 10.1021/ja8039426.
- (22) Julien, J. P.; Cupo, A.; Sok, D.; Stanfield, R. L.; Lyumkis, D.; Deller, M. C.; Klasse, P. J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Ward, A. B.; Wilson, I. A. Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science (1979)* 2013, *342* (6165), 1477–1483. DOI: 10.1126/science.1245625.
- (23) Tran, E. E. H.; Borgnia, M. J.; Kuybeda, O.; Schauder, D. M.; Bartesaghi, A.; Frank, G. A.; Sapiro, G.; Milne, J. L. S.; Subramaniam, S. Structural Mechanism of Trimeric HIV-1 Envelope Glycoprotein Activation. *PLoS Pathog* 2012, *8* (7), 37. DOI: 10.1371/journal.ppat.1002797.
- (24) Tan, K.; Liu, J. H.; Wang, J. H.; Shen, S.; Lu, M. Atomic Structure of a Thermostable Subdomain of HIV-1 Gp41. *Proc Natl Acad Sci U S A* 1997, *94* (23), 12303–12308. DOI: 10.1073/pnas.94.23.12303.
- (25) Banerjee, K.; Weliky, D. P. Folded Monomers and Hexamers of the Ectodomain of the HIV Gp41 Membrane Fusion Protein: Potential Roles in Fusion and Synergy between the Fusion Peptide, Hairpin, and Membrane-Proximal External Region. *Biochemistry* 2014, *53* (46), 7184–7198. DOI: 10.1021/bi501159w.
- (26) Close, W.; Neumann, M.; Schmidt, A.; Hora, M.; Annamalai, K.; Schmidt, M.; Reif, B.; Schmidt, V.; Grigorieff, N.; Fändrich, M. Physical Basis of Amyloid Fibril Polymorphism. *Nat Commun* 2018, *9* (1), 1–7. DOI: 10.1038/s41467-018-03164-5.
- (27) Howie, A. J.; Brewer, D. B.; Howell, D.; Jones, A. P. Physical Basis of Colors Seen in Congo Red-Stained Amyloid in Polarized Light. *Laboratory Investigation* 2008, *88* (3), 232–242. DOI: 10.1038/labinvest.3700714.

- (28) Howie, A. J.; Owen-Casey, M. P. Systematic Review of Accuracy of Reporting of Congo Red-Stained Amyloid in 2010–2020 Compared with Earlier. *Ann Med* 2022, *54* (1), 2511–2516. DOI: 10.1080/07853890.2022.2123558.
- (29) Hicks, A.; Escobar, C. A.; Cross, T. A.; Zhou, H. X. Fuzzy Association of an Intrinsically Disordered Protein with Acidic Membranes. *JACS Au* 2021, *1* (1), 66–78. DOI: 10.1021/jacsau.0c00039.
- (30) Hong, M. Determination of Multiple ϕ -Torsion Angles in Proteins by Selective and Extensive ^{13}C Labeling and Two-Dimensional Solid-State NMR. *Journal of Magnetic Resonance* 1999, *139* (2), 389–401. DOI: 10.1006/jmre.1999.1805.
- (31) Hong, M.; Jakes, K. Selective and Extensive ^{13}C Labeling of a Membrane Protein for Solid-State NMR Investigations. *J Biomol NMR* 1999, *14* (1), 71–74. DOI: 10.1023/A:1008334930603.
- (32) Higman, V. A.; Flinders, J.; Hiller, M.; Jehle, S.; Markovic, S.; Fiedler, S.; van Rossum, B. J.; Oschkinat, H. Assigning Large Proteins in the Solid State: A MAS NMR Resonance Assignment Strategy Using Selectively and Extensively ^{13}C -Labelled Proteins. *J Biomol NMR* 2009, *44* (4), 245–260. DOI: 10.1007/s10858-009-9338-7
- (33) Hong, M. Resonance Assignment of $^{13}\text{C}/^{15}\text{N}$ Labeled Solid Proteins by Two- and Three-Dimensional Magic-Angle-Spinning NMR. *J Biomol NMR* 1999, *15* (1), 1–14. DOI: 10.1023/A:1008334204412.
- (34) Gao, Q.; Chalmers, G. R.; Moremen, K. W.; Prestegard, J. H. NMR Assignments of Sparsely Labeled Proteins Using a Genetic Algorithm. *J Biomol NMR* 2017, *67* (4), 283–294. DOI: 10.1007/s10858-017-0101-1.
- (35) Dalgarno, D. C.; Levine, B. A.; Williams, R. J. P. Structural Information from NMR Secondary Chemical Shifts of Peptide Alpha C-H Protons in Proteins. *Biosci Rep* 1983, *3*, 443–452.
- (36) Spera, S.; Bax, A. Empirical Correlation between Protein Backbone Conformation and Ca and Cb ^{13}C Nuclear Magnetic Resonance Chemical Shifts. *J Am Chem Soc* 1991, No. 113, 5490–5492.
- (37) Caffrey, M.; Cai, M.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Gronenborn, A. M.; Clore, G. M. Determination of the Secondary Structure and Global Topology of the 44 KDa Ectodomain of Gp41 of the Simian Immunodeficiency Virus by Multidimensional Nuclear Magnetic Resonance Spectroscopy. *J Mol Biol* 1997, *271* (5), 819–826. DOI: 10.1006/jmbi.1997.1217.
- (38) Wishart, D. S. Interpreting Protein Chemical Shift Data. *Prog Nucl Magn Reson Spectrosc* 2011, *58* (1–2), 62–87. DOI: 10.1016/j.pnmrs.2010.07.004.
- (39) Cadars, S.; Lesage, A.; Emsley, L. Chemical Shift Correlations in Disordered Solids. *J Am Chem Soc* 2005, *127* (12), 4466–4476. DOI: 10.1021/ja043698f.

- (40) Su, Y.; Hong, M. Conformational Disorder of Membrane Peptides Investigated from Solid-State NMR Line Widths and Line Shapes. *Journal of Physical Chemistry B* 2011, *115* (36), 10758–10767. DOI: 10.1021/jp205002n.
- (41) Sakellariou, D.; Brown, S. P.; Lesage, A.; Hediger, S.; Bardet, M.; Meriles, C. a; Pines, A.; Emsley, L.; Lyon, D.; Chimie, S. De; Grenoble, C. E. a. High-Resolution NMR Correlation Spectra of Disordered Solids. *J Am Chem Soc* 2003, *125* (2), 4376–4380.
- (42) Melikyan, G. B.; Markosyan, R. M.; Hemmati, H.; Delmedico, M. K.; Lambert, D. M.; Cohen, F. S. Evidence That the Transition of HIV-1 Gp41 into a Six-Helix Bundle, Not the Bundle Configuration, Induces Membrane Fusion. *Journal of Cell Biology* 2000, *151* (2), 413–423. DOI: 10.1083/jcb.151.2.413.
- (43) Marti, D. N.; Bjelić, S.; Lu, M.; Bosshard, H. R.; Jelesarov, I. Fast Folding of the HIV-1 and SIV Gp41 Six-Helix Bundles. *J Mol Biol* 2004, *336* (1), 1–8. DOI: 10.1016/j.jmb.2003.11.058.
- (44) Zhang, H.; Neal, S.; Wishart, D. S. RefDB: A Database of Uniformly Referenced Protein Chemical Shifts. *J Biomol NMR* 2003, *25* (3), 173–195.
- (45) Filipp, F. V.; Sinha, N.; Jairam, L.; Bradley, J.; Opella, S. J. Labeling Strategies for ¹³C-Detected Aligned-Sample Solid-State NMR of Proteins. *Journal of Magnetic Resonance* 2009, *201* (2), 121–130. DOI: 10.1016/j.jmr.2009.08.012.
- (46) LeMaster, D. M.; Kushlan, D. M. Dynamical Mapping of E. Coli Thioredoxin via ¹³C NMR Relaxation Analysis. *J Am Chem Soc* 1996, *118* (39), 9255–9264. DOI: 10.1021/ja960877r.
- (47) David D. Lawas, Hans-Maecus L. Bitter, A. J. Solid-State NMR Spectroscopic Methods in Chemistry. *Angew Chem Int Ed Engl* 2002, *41*, 3096–3129.
- (48) Huber, M.; Böckmann, A.; Hiller, S.; Meier, B. H. 4D Solid-State NMR for Protein Structure Determination. *Physical Chemistry Chemical Physics* 2012, *14* (15), 5239–5246. DOI: 10.1039/c2cp23872a.
- (49) Le Marchand, T.; Schubeis, T.; Bonaccorsi, M.; Paluch, P.; Lalli, D.; Pell, A. J.; Andreas, L. B.; Jaudzems, K.; Stanek, J.; Pintacuda, G. ¹H-Detected Biomolecular NMR under Fast Magic-Angle Spinning. *Chem Rev* 2022, *122* (10), 9943–10018. DOI: 10.1021/acs.chemrev.1c00918.
- (50) Loquet, A.; Laage, S.; Gardiennet, C.; Elena, B.; Emsley, L.; Böckmann, A.; Lesage, A. Methyl Proton Contacts Obtained Using Heteronuclear Through-Bond Transfers in Solid-State NMR Spectroscopy. *J Am Chem Soc* 2008, *130* (32), 10625–10632. DOI: 10.1021/ja801464g.
- (51) Asami, S.; Schmieder, P.; Reif, B. High Resolution ¹H-Detected Solid-State NMR Spectroscopy of Protein Aliphatic Resonances: Access to Tertiary Structure Information. *J Am Chem Soc* 2010, *132* (43), 15133–15135. DOI: 10.1021/ja106170h.

Chapter 6 Very broad and different distributions of antiparallel β sheet registries of the wild-type and fusion-defective V2E mutant of the membrane-bound HIV fusion peptide and roles of these distributions in: (1) mutational robustness of fusion for HIV under constant immune pressure; and (2) loss of fusion and infection with V2E mutation

(This chapter is a collaborative work. All ssNMR data were obtained from REDOR experiments prepared and carried out by Dr. Scott Schmik and Dr. Li Xie.)

6.1 Introduction

Membrane-enveloped viruses are a large group that includes many families including HIV, influenza, and coronaviruses.¹⁻⁴ Cellular infection for these viruses requires fusion (joining) the viral and cellular membranes, and depending on the family, the latter is the plasma and/or an endosomal membrane. The fusion rate is typically negligible in the absence of catalyst, so each virus family has protein spikes that protrude from the viral membrane and with function that is in part fusion catalysis. There is homology in the spike sequence within a virus family but not between families. For “class I” viruses like HIV, each spike has three glycoproteins and each glycoprotein is a receptor-binding subunit and a fusion subunit.¹ For HIV, the glycoprotein 160 kD (gp160) is cleaved into the gp120 receptor-binding and gp41 fusion subunits, with ~510 and ~350 residues, respectively. Gp41 has a ~170-residue ectodomain outside the virus followed by a transmembrane domain (Tmd) and then a ~150-residue endo-domain that is located in the virus interior, Figure 6.1.⁵⁻⁸ The spike protruding from the virus has a core formed by the three gp41 ectodomains and the three gp120 subunits bound non-covalently to this core.⁹ Target T and macrophage cells are identified by gp120 binding to extracellular segments of integral plasma membrane proteins. Gp120 first binds CD4, followed by binding to either the CXCR4 or CCR5 chemokine receptor, and gp120 then moves away from the gp41 ectodomain.⁴ The gp41 residues ~25-160 then spontaneously transform to a different and thermostable trimer-of-hairpins structure, Figure 6.1(i-iv).^{7,8} Each hairpin has ~60-residue N-helix and C-helix segments separated by a loop. The N-helices from three gp41’s form an interior parallel coiled-coil and the C-helices are antiparallel and bound to the exterior grooves of the N-helix coil.

The ~23 N-terminal residues of gp41 are not part of the final hairpin and are named the “fusion peptide” (Fp).^{5, 10} The Fp sequence is fairly well-conserved among HIV isolates, with some variability.¹¹ Some engineered point mutations, e.g. V2E, result in highly-impaired gp160-mediated fusion and HIV infection.^{12,13} The Fp in the absence of the rest of gp160 binds membrane

and has been commonly proposed to bind the target membrane during fusion, Figure 6.1.^{1,14,15} Such binding could be important in overcoming activation energy barriers between different membrane structures during fusion. For example, close apposition between viral and target membranes is likely the initial step in fusion and requires ~25 kcal/mole of free energy, Figure 6.1(i).^{3,4,16} Some of this free energy could be provided through the combination of Fp in target membrane, Tmd in viral membrane, and intervening thermostable trimer-of-hairpins. There are ~10 kcal/mole barriers to form subsequent membrane structures during fusion, Figure 6.1(ii-iv), in part because these membrane transformations require large-amplitude motions of acyl chains of lipids.³ For example, formation of the “stalk” membrane intermediate, Figure 6.1(ii), requires “protrusion”, i.e. transient location of the outer leaflet lipid chains in the aqueous region.¹⁷ Previous studies have shown that large-amplitude motions and protrusion are more probable for lipids next to vs. more distant from the Fp.¹⁷⁻²⁰

In detergent-rich media, Fp is a monomer and residues 2-22 are a continuous single helix.^{21,22} In membrane without cholesterol, there are two Fp populations with distinct structures. One structure is the monomer helix observed in detergent and the other is an intermolecular β sheet oligomer with each Fp as a strand in the sheet.²³⁻²⁵ There is a positive correlation between the mole fraction cholesterol in the membrane and the β sheet oligomer population, with >90% sheet population when the cholesterol fraction is ~0.3, which is typical for the plasma membrane of host cells of HIV.²⁵⁻²⁸ The β sheet structure has been the focus of some previous investigations. One question was whether the strands form predominant parallel, antiparallel, or a mixture of parallel and antiparallel sheets. Infrared spectra of membrane-bound Fp with >8 sequential backbone ¹³CO labels were interpreted to support in-register parallel sheets while rotational-echo double-resonance (REDOR) NMR spectra of samples with equimolar mixtures of Fp's having either three ¹³CO or three ¹⁵N sequential labels were interpreted to support a mixture of parallel and antiparallel registries.^{29,30} These interpretations were confounded by ambiguity due to the extensive labelings. For example, the infrared data were spectral shifts due to dipole-dipole couplings and the interpretation that supported parallel sheet only considered couplings between adjacent Fp molecules but did not consider the effect of larger couplings within a molecule which would support antiparallel sheet. A much more definitive result was from a different REDOR NMR study with samples with mixtures of Fp's having one ¹³CO or two sequential ¹⁵N labels, and ¹³CO-Fp:¹⁵N-Fp = 1:2.³¹ Both the data and analysis supported predominant antiparallel sheets with an

upper limit of ~10% on the parallel population. NMR spectra of a large membrane-bound gp41 construct with Fp and hairpin regions supported Fp with predominant sheet rather than helix structure.³² Other NMR data evidenced predominant antiparallel vs. parallel Fp sheet.³³ NMR data also support a relatively small number of molecules in the sheet, perhaps ~10.^{15,34}

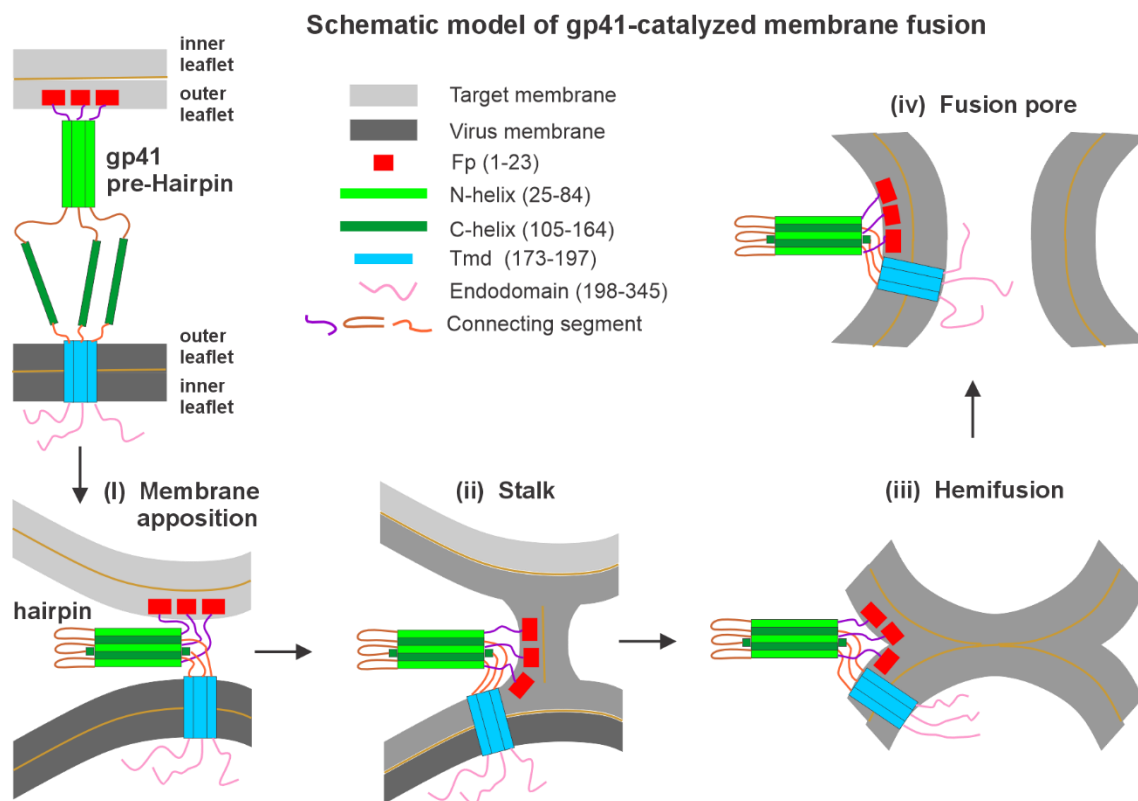


Figure 6.1. Schematic model for changes in gp41 and membrane structures during fusion. The approximate residue numbers are shown for the fusion peptide, N-helix, C-helix, transmembrane, and endo- domains. After binding of the gp120 subunit to the CD4 and chemokine receptors, gp120 moves away from gp41 and gp41 adopts the extended pre-hairpin structure. The Fp binds the target membrane followed by change to the final hairpin structure which promotes (i) initial close apposition of the two membranes. The membranes transform to the (ii) stalk in which the outer- but not inner- leaflets of the two membranes are contiguous, followed by topological change to (iii) hemifusion in which the Inner leaflets are joined, and then (iv) membrane pore with consequent single membrane that encloses the virus and cell. In the absence of gp41, calculations support ~25 kcal/mole barrier for initial apposition and ~10 kcal/mole barriers to form stalk, hemifusion, and pore. The endodomain forms a well-defined structure that is not displayed in this figure.

The present study addresses the distribution of registries (residue alignments) of adjacent antiparallel Fp molecules in the membrane-bound β sheets. A convenient index for a registry is t , the number of hydrogen-bonded Fp residues in the neighboring strands, starting from the N-termini. Figure 6.2a shows a schematic of the $t=16$ registry for a Fp with a ^{13}C O label at L12 and a ^{15}N

label at G5 (relevant for the present study). There have been limited experimental data from previous studies of the registry distribution. There was one analysis of REDOR NMR data of samples with A14 ^{13}CO + G3 ^{15}N Fp or with A14 ^{13}CO + I4 ^{15}N Fp and a separate analysis of REDOR NMR data from a sample with L12 ^{13}CO Fp and G5+A6 ^{15}N Fp in 1:2 molar ratio.^{31,35} The REDOR signals probe ^{13}CO and ^{15}N nuclei for which the ^{13}CO - ^{15}N distance, r , satisfies $r < 7$ Å. This condition could be satisfied for nuclei in neighboring molecules that have specific antiparallel registries. For these previous samples, the largest contributions to the REDOR signals would be from the $t=16$ and $t=17$ registries because these registries have the smallest r 's between the ^{13}CO and ^{15}N nuclei on adjacent molecules, e.g. $t=16$ aligns the A14 and G3 residues in neighboring molecules, Figure 6.2a. The analyses of the samples described above gave a best-fit sum of populations, $f(16) + f(17)$, in the 0.5-0.6 range, where $f(t)$ is the fractional population of Fp's with registry t . The present study provides a complete and much more accurate determination of all $f(t)$'s for the $t=8-24$ range, based on global analysis of REDOR data from 17 differently-labeled samples.

A second important contribution of the present study is a new experimentally-based model that explains how Fp structure contributes to fusion. This development is based on acquisition of REDOR data for membrane samples with Fp with wild-type (WT) vs. the fusion-defective V2E sequences, and then data fitting to determine the $f(t)$'s for WT vs. V2E β sheets. Earlier studies with cells expressing gp160 and cells expressing CD4 and a chemokine receptor showed complete loss of cell-cell fusion with V2E vs. WT gp160.^{12, 13} Similarly, HIV with V2E spikes is not infectious. Samples for the present study contain Fp without the rest of gp160 but the relevance of our study for understanding WT vs. V2E gp160 is supported by earlier observation of $\sim 10\times$ more extensive fusion between vesicles after exogenous addition of WT vs. V2E Fp.³⁶

Another notable earlier result is V2E-dominant reduction of fusion and infection, as was respectively observed with cells and viruses with spikes containing mixed trimers of WT and V2E gp160. For example, relative to spikes with only WT gp160, spikes with WT:V2E = 10:1 exhibited more than two-fold loss in function.¹³ The loss-in-function vs. WT:V2E data have been analyzed to estimate the number of spikes required for efficient fusion and infection. The estimates from different analyses vary between 1 and 19 spikes, with the large range due to the different assumptions of the analysis models.³⁷ The larger number estimates in this range are comparable to the numbers of spikes (14 ± 7) observed in single virions.³⁸ The requirement of multiple spikes for

fusion is supported by clustering of gp160 spikes in microscopy images of virions bound to cells.³⁹ To our knowledge, dominance has not been observed for gp160 with other Fp mutations. Mixtures of WT and V2E Fp's do not clearly exhibit V2E-dominant loss of vesicle fusion.⁴⁰ However, V2E-dominant loss was observed for vesicle fusion induced by a large Fp+hairpin construct that included most of the gp41 ectodomain and adopted trimer-of-hairpins structure.⁴¹ In addition, relative to WT Fp+hairpin, the V2E mutant exhibited ~15% loss in helicity. There were quantitatively-similar dependences of V2E-dominant losses vs. WT:V2E ratio for gp160-induced cell-cell fusion, HIV infection, Fp+hairpin-induced vesicle fusion, and Fp+hairpin helicity. As shown in Figure 6.1(i), one mechanistic hypothesis is that when the target and viral membranes respectively have bound WT Fp and Tmd, the thermostable hairpin enables the initial and required fusion step of close membrane apposition. The loss in helicity for V2E Fp+hairpin could be due to a shorter hairpin with functional consequence of larger distance between apposed membranes and higher energy barrier for fusion. The V2E mutation is separated by ~20-residues from the start of the hairpin and it isn't clear why the hairpin would be shorter with V2E. Addressing this question requires more detailed knowledge of the structure of membrane-bound Fp, and motivates the quantitative determinations of the distributions of registries for WT and V2E Fp's.

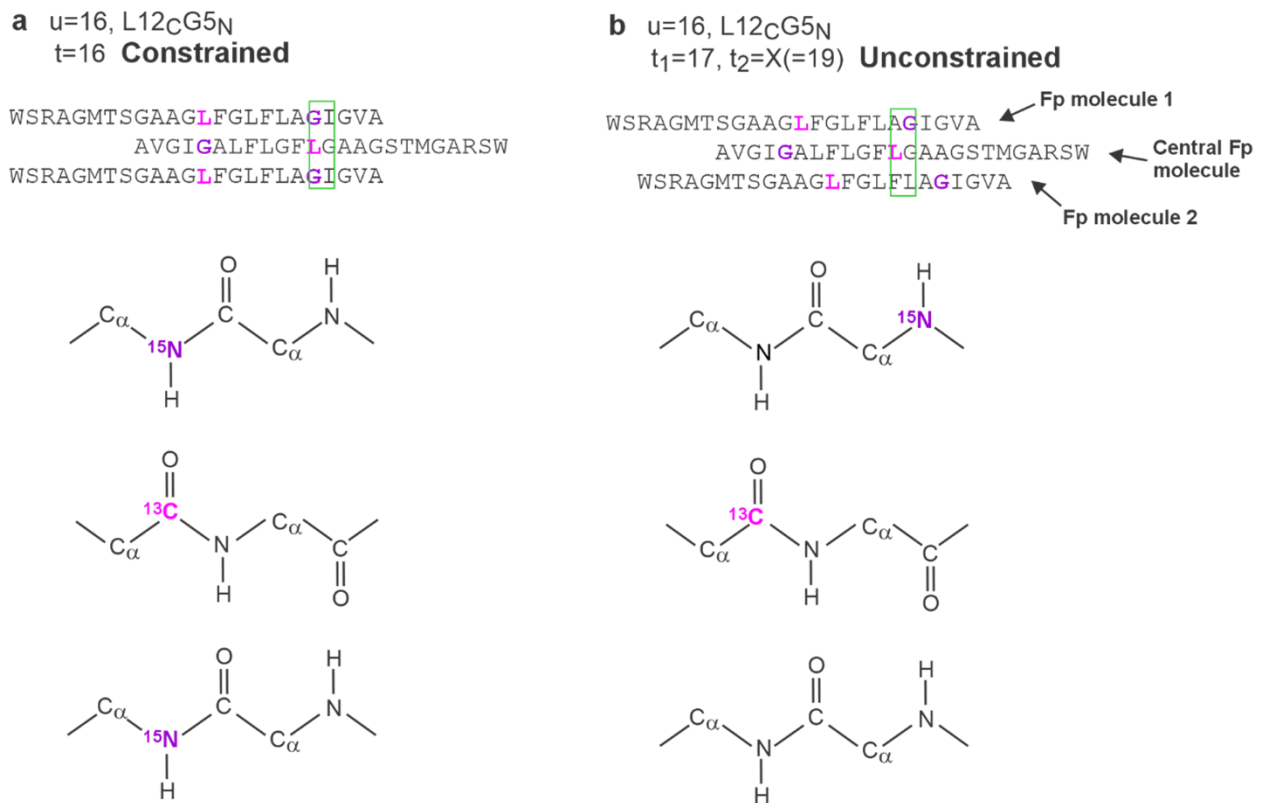


Figure 6.2. Schematic representations of antiparallel β sheet registries and $^{13}CO/^{15}N$ spin systems of the WT Fp. The Fp labeling is $u=16$ ($L12C^{13}G5N$), the $L12/^{13}CO$ are magenta, and the $G5/^{15}N$ are purple. Panel a shows three adjacent Fp's in a β sheet in the constrained model with $t=16$ for all registries in the sheet. The spin geometry for the region in the green rectangle is displayed below. Panel b shows three adjacent Fp's in the unconstrained model with $t_1=17$ and $t_2=X(=19)$. The spin geometry for the region in the green rectangle is displayed below. The t_1 in panel b refers to the registry adopted between the central Fp molecule and Fp molecule 1. The t_2 refers to the registry adopted between the central Fp molecule and Fp molecule 2. The central Fp ^{13}CO group is hydrogen bonded to a backbone HN of Fp molecule 1. There are likely more than three molecules in a Fp β sheet.

Another more general contribution of the present study is quantitative determination of a broad distribution of molecular structures, which is typically challenging to do by experiment.⁴² NMR approaches to quantitation of structural populations have typically relied on chemical shift resolution among signals from different structures so that the integrated signal intensity is proportional to the relative population of the structure.^{22, 23} Another NMR approach is measurement of non-radiative relaxation rates and then analysis using a model in which there is one structure with high population and a second structure with low population.⁴³ This approach was recently applied to describe the minor population of lipid chains that protrude towards the

headgroup region.²⁰ Analyses were done using the chain ^2H transverse relaxation rate, R_2 , and the increase in chain ^{13}C R_2 , Γ_2 , in the presence vs. absence of paramagnetic Mn^{2+} . The substantially larger $^2\text{H}R_2$'s and $^{13}\text{C}\Gamma_2$'s in the presence vs. absence of viral fusion peptides supported increased chain protrusion for lipids next to vs. further from fusion peptides, and the $^{13}\text{C}\Gamma_2$ analysis evidenced ~10% vs. ~1% protruded populations in membrane with vs. without fusion peptides. Structural populations can often be obtained computationally from molecular dynamics simulations although it isn't typically known whether the simulations reflect thermodynamic equilibrium.⁴⁴

For the Fp samples of the present study, the more common NMR approaches to determine the broad distribution of registry populations aren't applicable because the NMR signals of the different registries aren't resolved and the NMR relaxation rates aren't straightforwardly analyzable to determine populations. These circumstances motivated the approach of the present study in which registry populations were determined by global analysis of REDOR NMR data from a large group of differently-labeled Fp samples, with data of each sample dependent on only a few specific registries.

6.2 Experimental

6.2.1 Sample preparation

WT Fp peptides had sequence AVGIGALFLGFLGAAGSTMGARSWKKKKKKA, with underlining for the 23 N-terminal residues of HIV gp41, HXB2 laboratory strain, followed by a non-native W which served as a A280 chromophore, and a polylysine tag which resulted in Fp monomers in aqueous solution prior to membrane binding.^{14,45} The peptides were synthesized manually by solid-phase peptide synthesis using 9-Fluorenylmethoxycarbonyl (Fmoc) chemistry, followed by cleavage from the resin with a trifluoroacetic acid solution, and then purification with reverse-phase high-performance liquid chromatography with a semi-preparative C4 column. The synthesis and purification followed published methods, with typical final Fp purity >95% as assessed by MALDI-TOF mass spectrometry.²⁵ Each Fp had one residue with a backbone ^{13}C label and a different residue with a backbone ^{15}N label. Labeled amino acids were purchased from Cambridge Isotopes with Fmoc-protection done by published methods.⁴⁶ Each labeled Fp was indexed by an integer u . When $u=t$, where t is the registry length, the ^{13}C -labeled residue on one strand is aligned with the ^{15}N -labeled residue on the adjacent antiparallel strand. The integer value of u is the ^{13}C residue number plus ^{15}N residue number minus 1, e.g. the Fp with L12C_G5_N

labeling has $u=16$, Figure 6.2a. Both WT and V2E Fp's were produced with 17 different labelings that correspond to all values of u in the 8-24 range. WT Fp with $u = 28$ was also produced.

The lipid composition of the samples was 1,2-di-*O*-tetradecyl-*sn*-glycero-3-phosphocholine (DTPC), 1,2-di-*O*-tetradecyl-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (DTPG), and cholesterol in an 8:2:5 molar ratio. This composition has correspondence with that of plasma membranes of host cells of HIV which have a significant fraction of phosphatidylcholine.²⁸ In addition, the mole fraction DTPG is similar to the fraction negatively-charged lipid of host cells and there are also similar cholesterol fractions in host cell membranes.²⁸ Ether- rather than the ester-linked phospholipids were used so that there wouldn't be lipid natural abundance (na) ¹³CO NMR signal. This study relies on analysis of the labeled (lb) Fp ¹³CO NMR signals and this analysis would likely be less accurate if there were lipid contributions to the ¹³CO signals. Earlier studies showed that the lipid composition of the samples adopted membrane bilayer phase and the bound Fp had β sheet structure.^{35, 47, 48}

The DTPC, DTPG, and cholesterol, ~ 32 , 8, and 20 μ mole, were dissolved in chloroform followed by chloroform removal with nitrogen gas and vacuum. The solid was suspended in 2 mL of 5 mM HEPES buffer (pH 7.0) with 0.01% NaN₃ preservative and large unilamellar vesicles were formed with 10 freeze-thaw cycles followed by extrusion through 100 nm diameter pores of a polycarbonate filter. A solution containing ~ 5 mg Fp in ~ 30 mL HEPES buffer was added dropwise into the vesicle solution followed by gentle stirring overnight. Earlier analytical ultracentrifugation data showed that Fp is a monomer in this buffer.⁴⁵ Vesicles with bound Fp were pelleted by centrifugation at $\sim 150000g$ for 4 h and the bound Fp:total lipid mole ratio is estimated to be $\sim 1:60$, based on an earlier study, with unbound Fp in the supernatant.¹⁵ The pellet was lyophilized and transferred to a NMR magic angle spinning rotor with 4 mm outer diameter, followed by sample hydration with ~ 20 μ L water.

6.2.2 REDOR NMR spectroscopy

Spectra were acquired with a 9.4 T spectrometer with Varian Infinity Plus console. The samples were maintained at ~ -30 °C by cooling with nitrogen gas at -50 °C. This cooling helped to maintain sample stability and hydration, and reduced molecular motion so that internuclear dipolar couplings were close to the rigid values, with consequent larger signals from ¹H \rightarrow ¹³C cross-polarization (CP) and more accurate analysis of the ¹³CO-¹⁵N REDOR data.⁴⁹ Cooling did not modify Fp structure, as evidenced by earlier spectra showing very similar ¹³C shifts of samples

near ambient and cooled temperatures, with typical difference ≤ 0.5 ppm.^{15, 34} The rotor with sample was in a probe tuned to ^1H , ^{13}C , and ^{15}N frequencies. The ^{13}C transmitter was typically at 153 ppm with ^{13}C shift referencing done using the adamantane $^{-13}\text{CH}_2$ signal at 40.5 ppm. The REDOR experiment was done with 10 kHz magic angle spinning frequency, 2 s recycle delay between scans, and temporal sequence: (1) 50 kHz ^1H $\pi/2$ pulse; (2) 2.2 ms $^1\text{H} \rightarrow ^{13}\text{C}$ CP with a 60 kHz ^1H field and 63-68 kHz ramped ^{13}C field; (3) time period τ_k which alternated between S_0 reference scans with refocusing 54 kHz ^{13}C π pulses at the end of each rotor cycle except the last cycle, and S_1 scans with ^{13}C - ^{15}N dipolar recoupling because of ^{13}C π and 45 kHz ^{15}N π pulses at the end and in the middle of each rotor cycle, respectively; and (4) ^{13}C detection.^{24, 50} The rf pulses were set using a lyophilized helical peptide containing a single labeled ^{13}CO - ^{15}N spin pair with $r = 4.1$ Å.²⁴ There was XY-8 phase cycling for the ^{13}C and ^{15}N π pulses, and 80 kHz ^1H TPPM decoupling during periods 3 and 4.⁵¹ The phase cycle of S_0 and S_1 acquisitions was: ^1H $\pi/2$, 0, 180, 0, 180; ^1H CP, 90, 90, 90, 90; ^{13}C CP, 270, 270, 180, 180; final ^{13}C π , 270, 270, 180, 180; receiver, 180, 0, 90, 270. Typically $\sim 20,000$ S_0 or S_1 acquisitions were summed for each $\tau_k = 2.2, 8.2, 16.2, 24.2, 32.2, 40.2,$ and 48.2 ms with $k = 1, 2, 3, 4, 5, 6,$ and 7 , respectively.

S_0 and S_1 data were separately processed with 200 Hz Gaussian line broadening, Fourier transformation, and baseline correction followed by integration about the ^{13}CO peak with 3 ppm window that was the same for all spectra of a single sample and resulted in $S_0(u, \tau_k)$ and $S_1(u, \tau_k)$, with $u=8-24, 28$ and $k=1-7$. The uncertainty, $\sigma(u, \tau_k)$, was the root mean squared deviation (RMSD) of 24 different 3 ppm integrations of noise regions of both the S_0 and S_1 spectra. For either WT or V2E samples, the experimental $[S_1/S_0](u, \tau_k)$ ratios are the basis for determination of populations, $f(t)$'s, using χ^2 fitting that includes the $\sigma_{S_1/S_0}(u, \tau_k)$ calculated with error propagation. The data are typically graphically presented as a dephasing buildup, i.e. $\Delta S/S_0 = 1 - S_1/S_0$.

6.2.3 f(t) fitting

The total $S_0^{\text{tot}}(\tau_k)$ and total $S_1^{\text{tot}}(\tau_k)$ signals are each a sum of labeled (lb) and natural abundance (na) ^{13}CO signals. The $S_0^{\text{lb}}(\tau_k)$ is assigned to be 1.0 so that $S_0^{\text{na}}(\tau_k) = 0.33$, based on the 30 other backbone carbonyl nuclei. Both the $S_1^{\text{lb}}(u, \tau_k)$, which is u -dependent, and $S_1^{\text{na}}(\tau_k)$ signals are sums of contributions from different ^{13}CO populations that experience different ^{13}CO - ^{15}N dipolar dephasings. The $S_1^{\text{lb}}(u, \tau_k)$ signal includes an attenuated $S_1^{\text{lb,na}}(\tau_k)$ contribution from the lb ^{13}CO nuclei that experience dephasing from nearby na ^{15}N nuclei. The other $S_1^{\text{lb}}(u, \tau_k)$ signal is from lb

^{13}CO nuclei not near ^{15}N nuclei and is divided into two categories that are registry-dependent: (1) $S_1^{\text{lb,lb}}(u, \tau_k)$ is the signal from lb ^{13}CO nuclei in registries with t values close to u , and is attenuated because of dipolar couplings to the nearby lb ^{15}N nuclei; and (2) $S_1^{\text{lb,X}}(u, \tau_k)$ is the S_1 signal of the lb ^{13}CO in all other “X” registries and is not attenuated so that $S_1^{\text{lb,X}}(u, \tau_k) = S_0^{\text{lb,X}}(u, \tau_k)$. The $S_1^{\text{na}}(\tau_k)$ signal of the na ^{13}CO nuclei is similarly separated and includes one attenuated contribution that is denoted $S_1^{\text{na,lb}}(\tau_k)$ and is from the na ^{13}CO nuclei near lb ^{15}N nuclei. The S_1 signal of the other na ^{13}CO nuclei is not attenuated so that $S_1(\tau_k) = S_0(\tau_k)$. The lb $^{13}\text{CO}/\text{na } ^{15}\text{N}$ and na $^{13}\text{CO}/\text{lb } ^{15}\text{N}$ populations are calculated using the ^{13}C and ^{15}N na probabilities of 0.011 and 0.0037, respectively. These probabilities are small and only isolated spin pairs are considered. Selection of a specific pair in the β sheet as lb $^{13}\text{CO}/\text{na } ^{15}\text{N}$ or na $^{13}\text{CO}/\text{lb } ^{15}\text{N}$ is based on the magnitude of the $^{13}\text{CO}-^{15}\text{N}$ dipolar coupling, d , with $d(\text{Hz}) = 3080/[r(\text{\AA})]^3$, where r is the $^{13}\text{CO}-^{15}\text{N}$ internuclear distance. Among all samples, the smallest dephasing is for the WT $u=28$ (F8_CA21_N) sample, and this dephasing is used to validate a model of dephasing only due to lb $^{13}\text{CO}/\text{na } ^{15}\text{N}$ and na $^{13}\text{CO}/\text{lb } ^{15}\text{N}$ spin pairs and not lb $^{13}\text{CO}/\text{lb } ^{15}\text{N}$ pairs. For this model, the largest dephasing contributions are from the 4 pairs with $r < 5 \text{ \AA}$ in a model β sheet. Three of the pairs are intra-strand with $r = 1.3, 2.4, \text{ and } 4.6 \text{ \AA}$ and one pair is inter-strand with $r = 4.1 \text{ \AA}$. For example, the labeled F8 ^{13}CO has natural abundance intra-strand ^{15}N at L9, F8, and G10, respectively, and natural abundance inter-strand $^{15}\text{NH}\cdots\text{O}^{13}\text{C}$ (F8) hydrogen bond. The ratio $\gamma = S_1/S_0$ is calculated for each pair, with a pair indexed by $m = 1, 2, 3, \text{ or } 4$. More specifically, the $\gamma^{\text{lb,na}}(d_m, \tau_k)$ is calculated using an expression with n^{th} -order Bessel functions J_n of the first kind and the dimensionless parameter $\lambda_{m,k} = d_m \times \tau_k$.⁵²

$$\gamma^{\text{lb,na}}(d_m, \tau_k) = [J_0(\sqrt{2}\lambda_{m,k})]^2 - \left\{ 2 \times \sum_{n=1}^5 \frac{[J_n(\sqrt{2}\lambda_{m,k})]^2}{16n^2 - 1} \right\} \quad (6.1)$$

The calculated $S_1^{\text{lb,na}}$ and $S_1^{\text{na,lb}}$ signals:

$$S_1^{\text{lb,na}}(\tau_k) = 0.0037 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m, \tau_k) \quad (6.2)$$

$$S_1^{\text{na,lb}}(\tau_k) = 0.011 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m, \tau_k) \quad (6.3)$$

For any sample, the non-dephased S_1^{na} signal is $0.33 - (4 \times 0.011) = 0.286$. For a sample for which

$S_0^{lb,lb} = 0$, i.e. no populated registries with t close to u , the non-dephased $S_1^{lb,X} = S_0^{lb,X} = 1 - (4 \times 0.0037) = 0.9852$ so that:

$$\begin{aligned} S_1^{tot}(\tau_k) &= 1.2712 + S_1^{lb,na}(\tau_k) + S_1^{na,lb}(\tau_k) \\ &= 1.2712 + \left[0.0147 \times \sum_{m=1}^4 \gamma^{lb,na}(d_m, \tau_k) \right] \end{aligned} \quad (6.4)$$

This expression is used in the dephasing expression:

$$\frac{\Delta S}{S_0}(\tau_k) = \frac{[1.33 - S_1^{tot}(\tau_k)]}{1.33} \quad (6.5)$$

and is compared with the experimental $\left(\frac{\Delta S}{S_0}\right)^{exp}$ ($u=28, \tau_k$).

For the other samples, the $S_1^{na}(\tau_k)$ is the sum of the dephased signal from the 4 na ^{13}CO nuclei sites close to a lb ^{15}N nucleus and the remaining unde phased signal = 0.286 from the other na ^{13}CO nuclei. The combined $S_1^{na} + S_1^{lb,na}$:

$$S_1^{na}(\tau_k) + S_1^{lb,na}(\tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^4 \gamma^{lb,na}(d_m, \tau_k) \right] \quad (6.6)$$

The population of lb ^{13}CO that have not been dephased by na ^{15}N is $1 - [4 \times 0.0037] = 0.9852$. Calculation of the S_1^{lb} signal is done with two different models, referred to as constrained and unconstrained. In the constrained model, the sample is considered to contain separate β sheets with a single registry in each sheet. The determination of the $f(t)$ populations is done by fitting the $S_1^{lb,lb} + S_1^{lb,X}$ signal contributions from the $u=8-24$ samples. The $S_1^{lb,lb}$ contributions are from lb ^{13}CO nuclei in sheets with the $t = u, u+1, u-1, u+2$, or $u-2$ registries, i.e. registries with substantial lb ^{13}CO -lb ^{15}N dipolar coupling. The $\gamma^{lb,lb}_{t=u}(\tau_k)$, $\gamma^{lb,lb}_{t=u\pm 1}(\tau_k)$, and $\gamma^{lb,lb}_{t=u\pm 2}(\tau_k)$ values were calculated using the SIMPSON simulation program and the relevant geometry with one ^{13}CO and two ^{15}N spins.⁵³ As one example, Figure 6.2a displays schematic representations of the constrained $t=16$ registry for the $u=16$ sample, and the geometry of the three spins. In general, the spin geometries were based on atomic coordinates of the crystal structure of β barrel outer membrane protein G (OMPG, PDB file 2IWW). Simulation inputs were determined using these coordinates and the SIMMOL program and included the dipolar couplings and the Euler angles for each coupling vector and for the principal axis system of the ^{13}CO chemical shift anisotropy (CSA), as described in an earlier study.^{31, 54} The ^{13}CO CSA principal values were 247, 176, and 99 ppm. Each $\gamma(\tau_k)$ was an average from ~ 10 SIMPSON simulations that were each based on coordinates of different atoms

in OMPG. Neither ^1H 's nor relaxation were considered in the simulations. Table D2 presents the $\gamma^{\text{lb,lb}}_{t=u}(\tau_k)$, $\gamma^{\text{lb,lb}}_{t=u\pm 1}(\tau_k)$, and $\gamma^{\text{lb,lb}}_{t=u\pm 2}(\tau_k)$ values determined from these simulations. The approximations $\gamma^{\text{lb,lb}}_{t=u+1}(\tau_k) = \gamma^{\text{lb,lb}}_{t=u-1}(\tau_k)$ and $\gamma^{\text{lb,lb}}_{t=u+2}(\tau_k) = \gamma^{\text{lb,lb}}_{t=u-2}(\tau_k)$ are based on the differences in γ values between the two spin geometries being smaller than the differences due to variations in β sheetstructure.

The lb ^{15}N nuclei in the other (X) registries are considered too distant to dephase the lb ^{13}CO nuclei so that $S_1^{\text{lb,X}}(u,\tau_k) = S_0^{\text{lb,X}}(u,\tau_k) = f_X(u)$ and:

$$f_X(u) = 1 - \sum_{t=u-2}^{t=u+2} f(t) \quad (6.7)$$

$$S_1^{\text{lb,X}}(u,\tau_k) + S_1^{\text{lb,lb}}(u,\tau_k) = 0.9852 \times \left\{ f_X(u) + \sum_{t=u-2}^{t=u+2} [f(t) \times \gamma_t^{\text{lb,lb}}(\tau_k)] \right\} \quad (6.8)$$

$$S_1^{\text{tot}}(u,\tau_k) = S_1^{\text{na}}(\tau_k) + S_1^{\text{lb,na}}(\tau_k) + S_1^{\text{lb,lb}}(u,\tau_k) + S_1^{\text{lb,X}}(u,\tau_k) \quad (6.9)$$

The $\frac{S_1^{\text{tot}}(u,\tau_k)}{1.33} = \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u,\tau_k)$ so that:

$$1.33 \times \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u,\tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m,\tau_k) \right] + (0.9852) \times \left\{ 1 + \sum_{t=u-2}^{t=u+2} [f(t) \times \{\gamma_t^{\text{lb,lb}}(\tau_k) - 1\}] \right\} \quad (6.10)$$

Algebra is used to place the $f(t)$ terms on the left-side and the other terms on the right-side:

$$1 - \sum_{t=u-2}^{t=u+2} [f(t) \times \{1 - \gamma_t^{\text{lb,lb}}(\tau_k)\}] = \frac{\left\{ 1.33 \times \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u,\tau_k) - 0.286 - [0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m,\tau_k)] \right\}}{0.9852} \quad (6.11)$$

The $f(t)$, $t=8-24$, are determined by χ^2 fitting using Python code and the $\left(\frac{S_1}{S_0}\right)^{\text{exp}}(u,\tau_k)$ data with $\sigma_{\frac{S_1}{S_0}}^{\text{exp}}(u,\tau_k)$ uncertainties. The $f(6)$, $f(7)$, $f(25)$, and $f(26)$ were set to 0 in the fittings. Somewhat smaller χ^2 were sometimes obtained when the SIMPSON-calculated $\gamma_t^{\text{lb,lb}}(\tau_k)$ were all multiplied by a scaling parameter, b , with $0.95 < b < 1$.

$$1 - \sum_{t=u-2}^{t=u+2} [f(t) \times \{1 - b \times \gamma_t^{\text{lb,lb}}(\tau_k)\}] = \frac{\left\{ 1.33 \times \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u,\tau_k) - 0.286 - [0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m,\tau_k)] \right\}}{0.9852} \quad (6.12)$$

Better fitting with smaller $\gamma_t^{lb,lb}(\tau_k)$ could reflect inclusion of contributions from couplings to more distant lb ^{15}N nuclei.

The second approach is unconstrained fitting for which there can be different registries for the two Fp molecules, denoted 1 and 2, that are hydrogen bonded to the central Fp. The respective registries are denoted t_1 and t_2 , and when $t_1 = u$, the central Fp ^{13}CO is hydrogen-bonded to the H^{15}N of molecule 1. In Figure 6.62a, the schematic β sheet has $t_1 = t_2 = u = 16$. Relative to the constrained model, there are a larger number of distinct lb $^{13}\text{CO}/\text{lb}^{15}\text{N}$ spin geometries in the unconstrained model and the unconstrained analysis is done based on $\gamma_{t_1,t_2}(\tau_k) < 1$ only when $t_1 = u, u+1, \text{ or } u-1$ and/or $t_2 = u, u+1, \text{ or } u-1$. Figure 6.62b displays schematic representations of three adjacent Fp molecules in an unconstrained sheet with $u=16, t_1=17, \text{ and } t_2=X(=19)$. The SIMPSON-calculated $\gamma_{t_1,t_2}(\tau_k)$ are presented in Table D1. Equations 6.1 - 6.6 are valid for the unconstrained analysis and Equations 6.7 and 6.8 become:

$$f_X(u) = 1 - \sum_{t=u-1}^{t=u+1} f(t) \quad (6.13)$$

$$S_1^{lb,X}(u, \tau_k) + S_1^{lb,lb}(u, \tau_k) = 0.9852 \times \left\{ \sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}(\tau_k) \right] \right\} \quad (6.14)$$

And:

$$S_1^{\text{tot}}(u, \tau_k) = S_1^{\text{na}}(\tau_k) + S_1^{\text{lb,na}}(\tau_k) + S_1^{\text{lb,lb}}(u, \tau_k) + S_1^{\text{lb,X}}(u, \tau_k) \quad (6.15)$$

The $S_1^{\text{na}}(\tau_k) + S_1^{\text{lb,na}}(\tau_k)$ are described by Equation 6.6 and $\frac{S_1^{\text{tot}}(u, \tau_k)}{1.33} = \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u, \tau_k)$ so that:

$$1.33 \times \left(\frac{S_1}{S_0}\right)^{\text{exp}}(u, \tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m, \tau_k) \right] + (0.9852) \times \left\{ \sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}(\tau_k) \right] \right\} \quad (6.16)$$

Algebra is used to place the $f(t)$ terms on the left-side and the other terms on the right-side:

$$\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}(\tau_k) \right] = \frac{\left\{ 1.33 \times \left(\frac{S_1}{S_0} \right)^{\text{exp}}(u, \tau_k) - 0.286 - \left[0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m, \tau_k) \right] \right\}}{0.9852} \quad (6.17)$$

The $f(t)$, $t=8-24$, are determined by fitting to the $\left(\frac{S_1}{S_0} \right)^{\text{exp}}(u, \tau_k)$ data, with $f(7) = f(25) = 0$. Similar to constrained fitting, unconstrained fitting was also done with $\gamma_{t_1,t_2}(\tau_k)$ multiplied by a scaling factor, $b_{t_1=u, u \pm 1, t_2=u, u \pm 1} = 0.98$, $b_{t_1=u, u \pm 1, t_2=X} = b_{t_1=X, t_2=u, u \pm 1} = 0.99$, and $b_{t_1=X, t_2=X} = 1$. For conciseness, this scaling is often referred to as “ $b=0.98$ ”.

$$\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times b_{t_1,t_2} \times \gamma_{t_1,t_2}(\tau_k) \right] = \frac{\left\{ 1.33 \times \left(\frac{S_1}{S_0} \right)^{\text{exp}}(u, \tau_k) - 0.286 - \left[0.0147 \times \sum_{m=1}^4 \gamma^{\text{lb,na}}(d_m, \tau_k) \right] \right\}}{0.9852} \quad (6.18)$$

6.3 Results

6.3.1 Spectra and lineshape fitting

Figure 6.3 displays plots of the ^{13}CO regions of the REDOR NMR S_0 and S_1 spectra at $\tau_k = 40.2$ ms for the $u=8-24$ samples, WT and V2E, and $u=28$, WT. Figure 6.4 displays expanded views of the S_0 and $\Delta S = S_0 - S_1$ spectra for the $u=17$ and $u=20$ samples. The S_0 spectra include both lb and na ^{13}CO contributions whereas the ΔS spectra are predominantly the lb ^{13}CO signals. Both S_0 and ΔS lineshapes are well-fitted by a single Gaussian function with example fittings shown in Figure 6.4. Table 6.1 lists the fitted peak chemical shifts, δ_{peak} 's, and full-width at half-maximum linewidths, δ_{FWHM} 's, with labeled ^{13}CO sites of A6, L7, F8, L9, and L12. For both WT and V2E Fp's, the δ_{peak} values correlate with β sheet structure, and the typical δ_{FWHM} is between 3 and 4 ppm.⁵⁵ For a particular sample, the δ_{peak,S_0} and $\delta_{\text{peak},\Delta S}$ typically agree within 0.3 ppm. The δ_{FWHM,S_0} is usually larger than $\delta_{\text{FWHM},\Delta S}$, typically by 0.2-0.5 ppm, which likely reflects the presence vs. absence of na contributions in S_0 vs. ΔS spectra. For either WT or V2E, the F8, L9, and L12 sites are ^{13}CO labeled in multiple samples, Table 6.1, and the spectrum for a particular site is expected to be similar among samples. This expectation is supported by RMSD's that are typically <0.3 ppm for average values of δ_{peak,S_0} , $\delta_{\text{peak},\Delta S}$, δ_{FWHM,S_0} , and $\delta_{\text{FWHM},\Delta S}$, Table 6.1. This spectral similarity

evidences the reproducibility of the sample preparation and NMR methods. For a specific ^{13}CO site and u , there are also similar δ_{peak} values for WT and V2E samples.

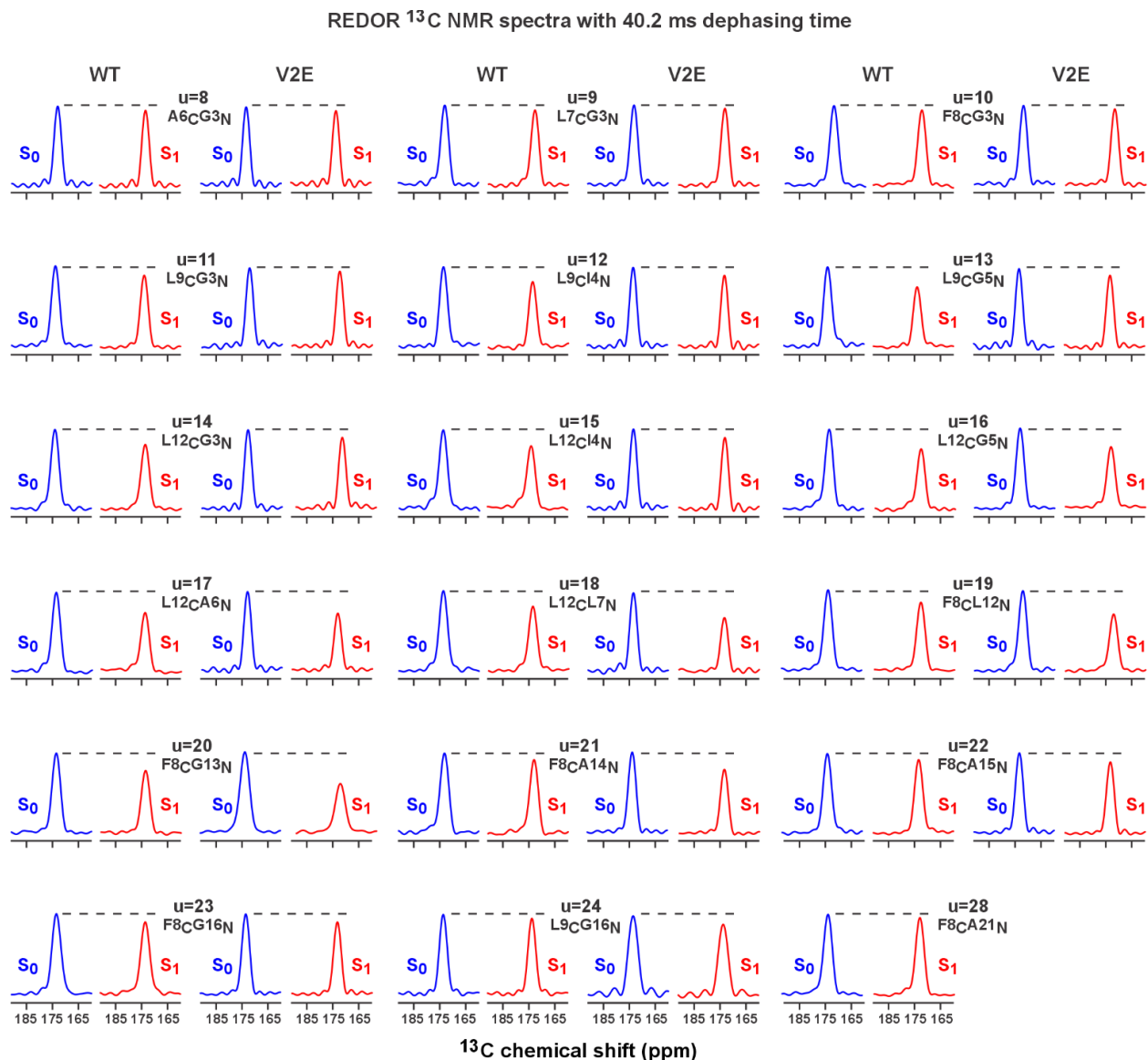


Figure 6.3. Plots of the ^{13}CO region of REDOR NMR S_0 (blue) and S_1 (red) spectra with $\tau_k = 40.2$ ms. The samples are membrane-bound Fp with $u=8-24$ labeling, WT or V2E, and $u=28$, WT. Each column of spectra is for either WT or V2E samples. The ^{13}CO and ^{15}N labelings are shown for each Fp, e.g. L12CG5N for $u=16$. The peak intensities are the same for all the S_0 spectra and are marked with dashed lines. Spectra were processed with 100 Hz Gaussian line broadening and 5th order polynomial baseline correction.

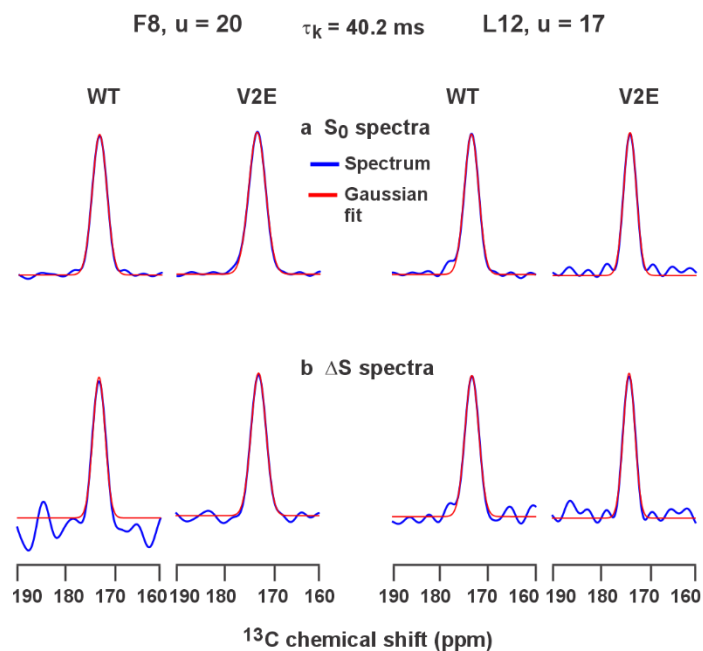


Figure 6.4. Plots of a S_0 and b ΔS ^{13}C REDOR NMR spectra for 40.2 dephasing time, blue traces, and Gaussian fits, red traces. Spectra are displayed for (left) F8, u=20 samples, and (right) L12, u=17 samples. Spectra were processed with 100 Hz Gaussian line broadening and 5th order polynomial baseline correction. The S_0 spectra have contributions from both labeled and natural abundance ^{13}C signals whereas the ΔS spectra are predominantly the labeled ^{13}C signals. Table 6.1 presents the peak shifts and linewidths for all samples, u=8-24, derived from the Gaussian fittings.

Table 6.1. ^{13}C O peak chemical shifts and linewidths determined from REDOR spectra with $\tau_k = 40.2 \text{ ms}^a$.

Residue	u	WT				V2E			
		S_0		ΔS		S_0		ΔS	
		δ_{peak} ppm	δ_{FWHM}	δ_{peak} ppm	δ_{FWHM}	δ_{peak} ppm	δ_{FWHM}	δ_{peak} ppm	δ_{FWHM}
A6	8	173.87	2.95			173.94	2.81		
L7	9	173.59	3.34			173.78	3.21		
	10	172.85	3.44			172.75	3.18		
	19	172.58	3.63	172.35	3.57	173.00	3.55	172.66	3.39
F8	20	173.07	3.58	173.13	3.12	173.12	4.15	172.79	3.59
	21	172.78	3.85			173.05	3.24	173.40	3.14
	22	173.17	3.63			173.02	3.27		
	23	173.14	3.88			172.94	3.21		
Average(RMSD)		172.93(23)	3.67(17)	172.74(55)	3.35(32)	172.98(13)	3.43(38)	172.95(40)	3.37(23)
	11	173.50	3.38			173.22	2.96		
L9	12	173.39	3.29	173.62	3.12	173.74	2.87		
	13	173.63	3.38	173.79	3.02	173.88	2.87		
	24	173.33	3.26			173.10	2.86		
Average(RMSD)		173.46(13)	3.33(6)	173.71(12)	3.07(7)	173.49(38)	2.89(5)		
	14	173.49	3.51	173.24	3.08	173.69	2.98		
	15	173.54	3.66	173.77	3.29	174.06	3.59		
L12	16	173.75	3.53	173.61	3.45	173.92	3.02	174.04	3.16
	17	173.32	3.54	173.33	3.38	173.83	2.98	173.96	2.86
	18	173.31	3.70			173.95	3.04	173.93	2.94
Average(RMSD)		173.48(18)	3.59(9)	173.49(25)	3.30(16)	173.89(14)	3.12(27)	173.98(6)	2.99(16)

^a ^{13}C O peak shifts and full-width at half-maximum linewidths from fitting REDOR S_0 and ΔS spectra with $\tau_k = 40.2 \text{ ms}$. The spectra were processed with 100 Hz Gaussian line broadening and 5th order baseline correction and fitted to a Gaussian line profile, see Figure 6.4 for examples. The ΔS spectra were only fitted when there was reasonable signal-to-noise. Averages for a particular ^{13}C O with RMSD's in parentheses are also listed using the convention that the RMSD corresponds to the right-most digits in the average, e.g. 173.48(18) means 173.48 ± 0.18 .

6.3.2 Unconstrained and constrained fittings

Table 6.2 numerically presents the experimental $\Delta S/S_0$ data. Figure D1 and Table D1 display the experimental $\Delta S/S_0$ of two WT $u=20$ samples that were separately-prepared. The agreement within error supports reproducibility of the $\Delta S/S_0$ values for two similarly-prepared samples. There is similar agreement between the $\Delta S/S_0$ values for replicate $u=13, 16,$ and 17 samples, Table D1. Figure 6.5 displays plots of experimental $\Delta S/S_0$ vs. τ_k for WT and V2E samples, $u=8-24$ and $\tau_k=2.2-40.2$ ms, as well as the best-fit $\Delta S/S_0$ from unconstrained fittings, Figure 6.2b, using Equation 6.18 and “ $b=0.98$ ”, i.e. $b_{t1=u, u\pm 1, t2=u, u\pm 1} = 0.98$, $b_{t1=u, u\pm 1, t2=X} = b_{t1=X, t2=u, u\pm 1} = 0.99$, and $b_{t1=X, t2=X} = 1$. Figure 6.6 and Table 6.3 present the best-fit $f(t)^{WT}$ and $f(t)^{V2E}$ from these fittings, and Tables D3 (WT) and D4 (V2E) numerically present the experimental and calculated $\Delta S/S_0$. The quality of the unconstrained fitting model is evidenced by the best-fit χ^2 of 107 for WT and 145 for V2E, which are close to the number of data, 102. Figure 6.5 also displays the experimental $\Delta S/S_0$ for the WT $u=28$ sample and the calculated $\Delta S/S_0$ for dipolar dephasing from na $^{13}\text{CO}/\text{lb}$ ^{15}N and lb $^{13}\text{CO}/\text{na}$ ^{15}N spin pairs with $r < 5 \text{ \AA}$, Equations 6.1-6.5. The quantitative agreement within error between the experimental and calculated $\Delta S/S_0$ supports the na dephasing model.

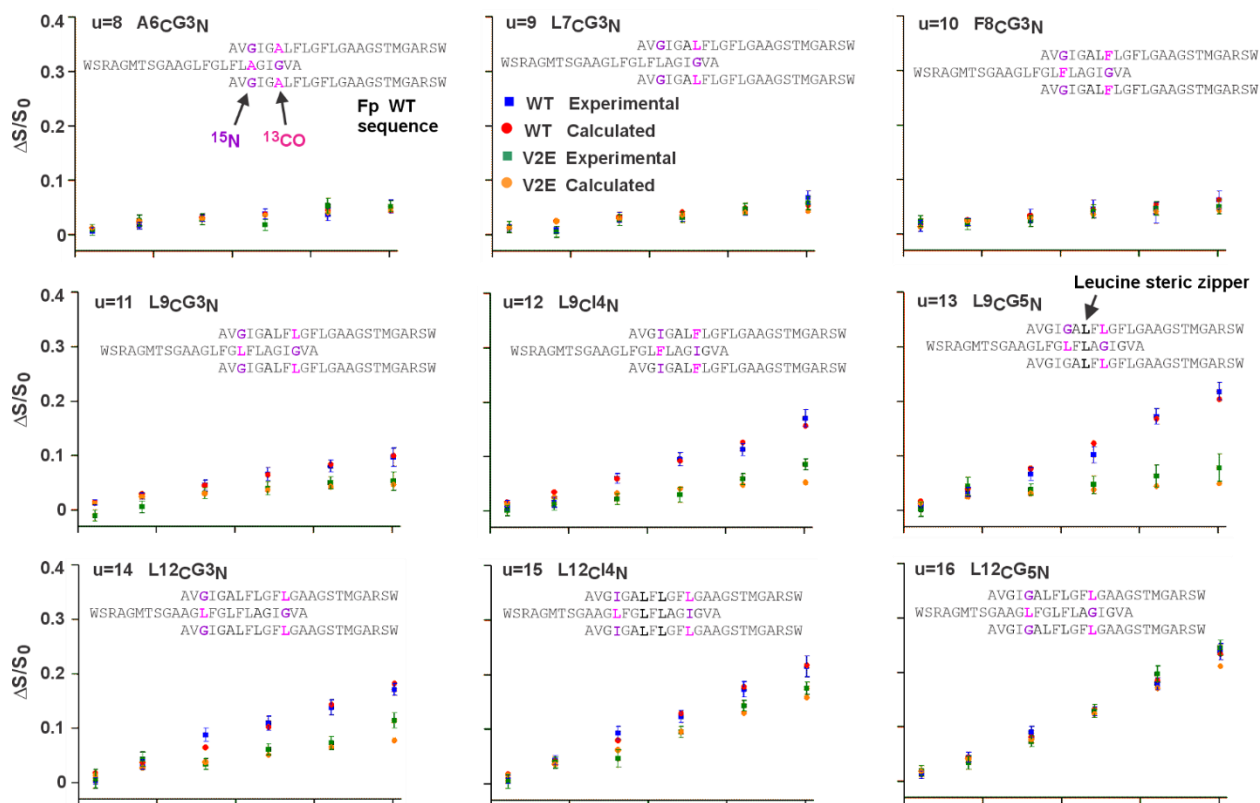


Figure 6.5. Plots of ^{13}CO - ^{15}N REDOR NMR $\Delta S/S_0$ vs. dephasing time (τ_k) for the u=8-24, 28 samples with membrane-bound Fp, WT and V2E, and $\tau_k=2.2$ -40.2 ms ($k=1$ -6). Both experimental and calculated $\Delta S/S_0$ are displayed. For u=8-24, the calculated $\Delta S/S_0$ are based on unconstrained fitting of the u=8-24 data with $b=0.98$, Equation 6.18. For u=28, natural abundance dephasing is dominant and the calculated $\Delta S/S_0$ are based on Equations 6.1-6.5. The numerical values of the experimental and calculated $\Delta S/S_0$ are presented in Tables D3, D4. Each Fp has one residue with a backbone ^{13}CO label and a different residue with a backbone ^{15}N label. Each antiparallel registry is indexed by t , the number of aligned residues of the adjacent strands, starting from the N-termini. The sample index u is the value of t that aligns the labeled ^{13}CO and ^{15}N residues in a constrained β sheet, as displayed schematically in each panel. The ^{13}CO - and ^{15}N -labeled residues in the WT sequence are bolded in magenta and purple, respectively. The leucines that are aligned in the β sheet are also bolded. The membrane-bound β sheets likely have more than three Fp molecules.

Figure 6.5 (cont'd)

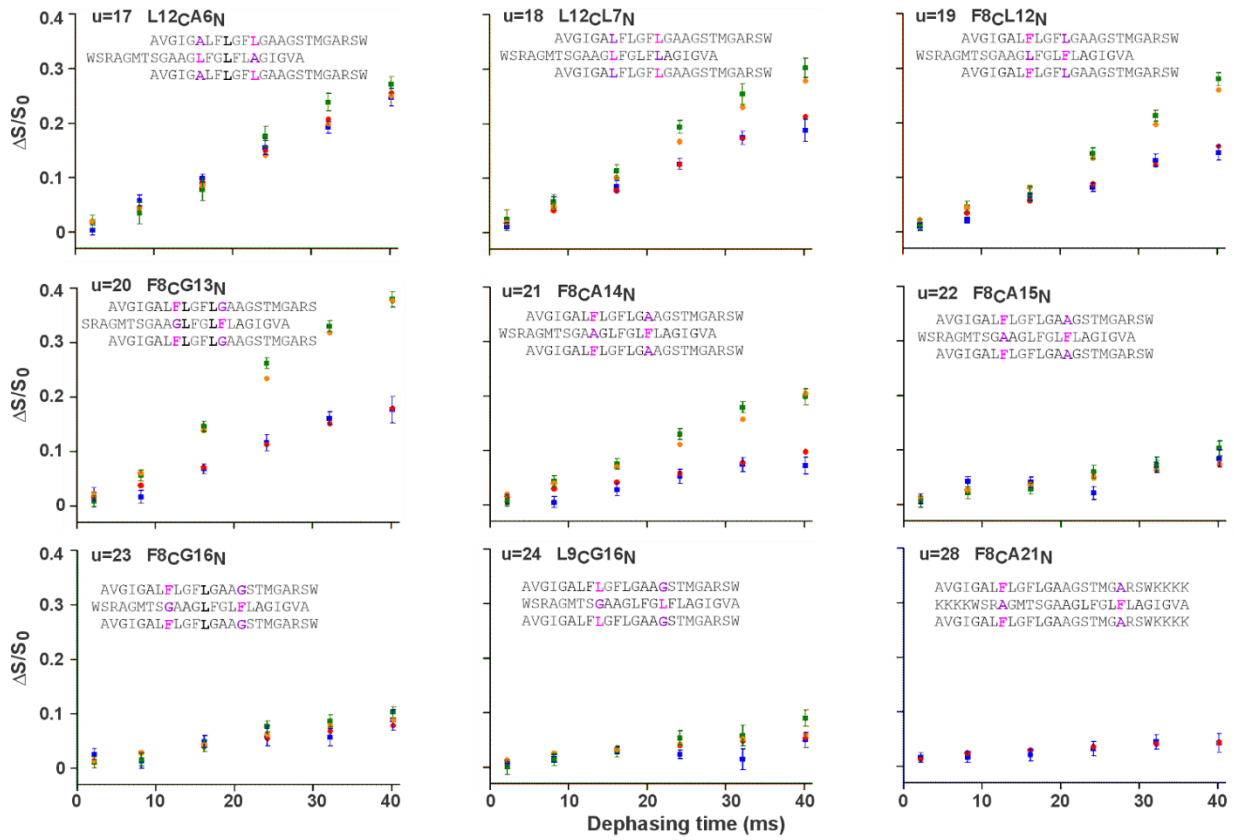


Table 6.2. Experimental REDOR $\Delta S/S_0$ ^a.

u	WT							V2E						
	Dephasing time (ms)							Dephasing time (ms)						
	2.2	8.2	16.2	24.2	32.2	40.2	48.2	2.2	8.2	16.2	24.2	32.2	40.2	48.2
8	0.006(7)	0.017(6)	0.030(6)	0.038(9)	0.037(11)	0.052(12)	0.057(16)	0.009(10)	0.026(10)	0.028(10)	0.018(10)	0.053(14)	0.052(10)	0.055(12)
9	0.012(6)	0.009(6)	0.032(9)	0.033(9)	0.047(11)	0.068(12)	0.063(18)	0.014(10)	0.005(10)	0.027(10)	0.032(10)	0.047(10)	0.058(12)	0.062(17)
10	0.015(10)	0.022(7)	0.033(12)	0.046(16)	0.039(19)	0.062(17)	0.111(21)	0.024(10)	0.018(10)	0.024(10)	0.041(12)	0.047(12)	0.050(13)	0.068(16)
11	0.014(5)	0.026(6)	0.046(9)	0.066(12)	0.081(11)	0.097(17)	0.151(22)	-0.010(10)	0.006(11)	0.031(10)	0.040(12)	0.050(11)	0.053(17)	0.062(13)
12	0.011(9)	0.016(9)	0.060(9)	0.095(12)	0.113(11)	0.170(16)	0.215(23)	0.001(10)	0.012(10)	0.022(10)	0.030(14)	0.059(10)	0.086(10)	0.085(16)
13	0.010(5)	0.034(8)	0.067(12)	0.102(15)	0.172(14)	0.218(16)	0.256(25)	0.002(13)	0.044(17)	0.038(10)	0.047(16)	0.063(20)	0.078(25)	0.129(25)
14	0.003(6)	0.033(9)	0.088(12)	0.109(13)	0.138(14)	0.171(11)	0.235(21)	0.007(17)	0.043(13)	0.034(10)	0.061(10)	0.073(12)	0.114(14)	0.154(12)
15	0.008(8)	0.043(8)	0.093(12)	0.123(11)	0.173(14)	0.215(19)	0.244(19)	0.004(13)	0.039(10)	0.046(16)	0.095(10)	0.143(10)	0.175(11)	0.195(10)
16	0.012(7)	0.044(9)	0.090(10)	0.128(8)	0.179(11)	0.238(15)	0.253(15)	0.019(10)	0.034(12)	0.073(10)	0.129(12)	0.197(15)	0.245(14)	0.287(23)
17	0.004(9)	0.058(10)	0.099(7)	0.155(13)	0.192(11)	0.247(16)	0.275(21)	0.019(12)	0.035(20)	0.078(20)	0.175(19)	0.238(16)	0.271(14)	0.310(20)
18	0.011(7)	0.055(11)	0.085(11)	0.126(10)	0.174(12)	0.188(20)	0.201(21)	0.025(17)	0.057(13)	0.113(12)	0.194(12)	0.254(19)	0.302(18)	0.303(22)
19	0.01(7)	0.022(5)	0.064(6)	0.082(8)	0.131(11)	0.145(13)	0.157(13)	0.013(11)	0.046(10)	0.069(14)	0.144(10)	0.213(10)	0.280(12)	0.346(12)
20	0.022(12)	0.017(12)	0.068(9)	0.116(15)	0.161(12)	0.177(24)	0.175(15)	0.009(10)	0.056(10)	0.146(10)	0.262(10)	0.330(10)	0.379(14)	0.398(17)
21	0.010(9)	0.005(10)	0.028(11)	0.052(13)	0.074(13)	0.072(16)	0.112(16)	0.007(10)	0.043(10)	0.075(10)	0.130(10)	0.179(10)	0.198(15)	0.257(11)
22	0.011(9)	0.042(9)	0.041(9)	0.021(12)	0.070(11)	0.084(16)	0.096(13)	0.005(10)	0.022(11)	0.029(10)	0.060(12)	0.075(12)	0.103(14)	0.155(15)
23	0.026(10)	0.014(14)	0.049(11)	0.059(18)	0.057(16)	0.089(19)	0.113(19)	0.011(10)	0.016(11)	0.041(10)	0.077(10)	0.087(11)	0.103(10)	0.113(11)
24	0.006(5)	0.016(7)	0.031(6)	0.024(8)	0.015(19)	0.050(14)	0.046(14)	0.001(13)	0.014(11)	0.029(10)	0.054(13)	0.058(19)	0.090(15)	0.101(18)
28	0.016(9)	0.017(10)	0.021(11)	0.032(13)	0.045(13)	0.043(17)	0.044(21)							

^a The uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$, e.g. 0.015(10) means 0.015 ± 0.010 .

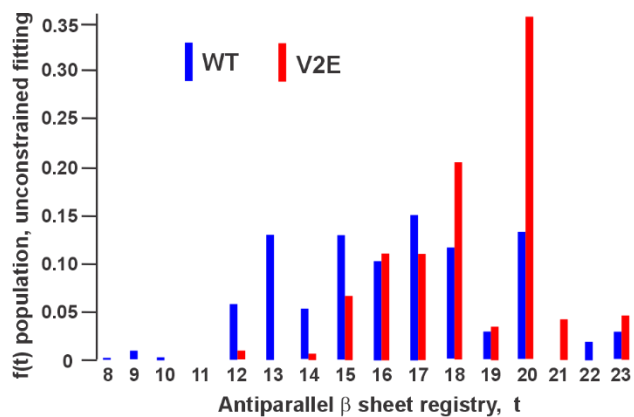


Figure 6.6. Plot of WT (blue bar) and V2E (red bar) fractional populations, $f(t)$, vs. antiparallel β sheet registry, t , where t is the number of aligned residues in adjacent strands, starting at the N-termini, Figures. 6.2, 6.5. The numerical values of $f(t)^{\text{WT}}$ and $f(t)^{\text{V2E}}$ are presented in Table 6.3 and were determined using the 102 $\Delta S/S_0$ data from the $u=8-24$ samples with $\tau_k = 2.2-40.2$ ms and the unconstrained model with $b=0.98$, Equation 6.18.

Table 6.3. The $f(t)$ for fittings with $b = 0.98$ and 2.2-40.2 ms data^a.

	WT; Unconstrained	WT; Constrained	V2E; Unconstrained	V2E; Constrained
t	$\chi^2=107; \langle t \rangle=16.1843$	$\chi^2=131; \langle t \rangle=16.1734$	$\chi^2=145; \langle t \rangle=18.4585$	$\chi^2=231; \langle t \rangle=18.4957$
8	0.0015	0	0	0
9	0.0090	0.0027	0	0
10	0.0032	0	0	0
11	0.0355	0.0247	0	0
12	0.0579	0.0672	0.0092	0
13	0.1306	0.1384	0	0
14	0.0524	0.0545	0.0065	0
15	0.1297	0.1266	0.0769	0.0806
16	0.1035	0.1069	0.1113	0.1114
17	0.1514	0.1678	0.1106	0.1254
18	0.1159	0.1206	0.2054	0.1993
19	0.0290	0.0138	0.0351	0.0058
20	0.1325	0.1490	0.3564	0.4275
21	0	0	0.0425	0.0137
22	0.0193	0.0033	0	0
23	0.0285	0.0244	0.0460	0.0364
24	0	0	0	0

^a Unconstrained and constrained fittings were done using $u = 8-24$, $k = 1-6$, $\tau_k = 2.2-40.2$ ms data. For unconstrained fittings, $b_{t1=u, u\pm 1, t2=u, u\pm 1} = 0.98$, $b_{t1=u, u\pm 1, t2=X} = b_{t1=X, t2=u, u\pm 1} = 0.99$, and $b_{t1=X, t2=X} = 1$.

Table 6.3 also presents the best-fit $f(t)^{\text{WT}}$ and $f(t)^{\text{V2E}}$ from constrained fitting, Figure 6.2a, using Equation 6.12 with $b=0.98$. The best-fit χ^2 are 131 for WT and 231 for V2E and the calculated $\Delta S/S_0$ are presented in Tables D3, D4. For either WT or V2E, the unconstrained and constrained fittings both show similar trends for $f(t)$ vs. t and the numerical $f(t)$ typically agree within 0.02 between the two fitting models. Table D5 presents the $f(t)$ and χ^2 values from different unconstrained and constrained fittings of WT and V2E data, including fittings based on either $\tau_k=2.2-40.2$ or $2.2-48.2$ ms data and with different b values for γ , Equations 6.12 and 6.18. The χ^2 values are typically smaller for unconstrained vs. constrained fittings. Comparison of $f(t)$ values among the different unconstrained fittings or among the different constrained fittings show very similar values for fittings without vs. with $\tau_k=48.2$ ms data and for different b values in the 0.95-1.0 range. For a specific $f(t)$, there is typically agreement within 0.01 among fittings. The average value of t , denoted $\langle t \rangle$, is highly-conserved. For all WT fittings, the $\langle t \rangle^{\text{WT}} = 16.132 \pm 0.048$ and for all V2E fittings, $\langle t \rangle^{\text{V2E}} = 18.475 \pm 0.028$.

6.3.3 Free energy contributions to $f(t)$

For either WT or V2E, the substantial REDOR $\Delta S/S_0$ data for many differently-labeled samples are the basis for the broad distributions of populated registries and for the population weighting towards longer registries for V2E vs. WT, Figures 6.5, 6.6 and Tables 6.3, 6.4. The reproducibility of the sample preparation and REDOR NMR approaches is supported by typical agreement within uncertainties between $\Delta S/S_0$ values from replicate samples, Figure D1 and Table D1. The $f(t)^{\text{WT}}$ and $f(t)^{\text{V2E}}$ distributions from unconstrained fittings with $b=0.98$ were quantitatively-analyzed with thermodynamic models:

$$f(t)^{\text{WT}} = C^{\text{WT}} \times \exp \left\{ \frac{- \left[(G_{\beta}^{\text{WT}} \times t) + (G_{\text{zip}}^{\text{WT}} \times L(t)) + (G_{\text{sc}}^{\text{WT}}(t) \times g^{\text{WT}}) \right]}{RT} \right\} \quad (6.19)$$

$$f(t)^{\text{V2E}} = C^{\text{V2E}} \times \exp \left\{ \frac{- \left[(G_{\beta}^{\text{V2E}} \times t) + (G_{\text{zip}}^{\text{V2E}} \times L(t)) \right]}{RT} \right\} \quad (6.20)$$

Table 6.4. The $f(t)$ for unconstrained and constrained fittings with different b values, and based on $k=1-6$, $\tau_k=2.2-40.2$ ms, or $k=1-7$, $\tau_k=2.2-48.2$ ms data from $u=8-24$ samples. For all WT fittings, the average value of $\langle t \rangle$ and RMSD is $\langle \langle t \rangle^{WT} \rangle = 16.132 \pm 0.048$. For all V2E fittings, $\langle \langle t \rangle^{V2E} \rangle = 18.475 \pm 0.028$.

	WT	WT	WT	WT	WT	WT	WT	V2E	V2E	V2E	V2E	V2E	V2E	V2E	V2E	
	Uncons	Uncons	Uncons	Uncons	Cons.	Cons.	Cons.	Uncons	Uncons	Uncons	Uncons	Cons.	Cons.	Cons.	Cons.	
	$b=0.98$	$b=0.98$	$b=1$	$b=1$	$b=0.98$	$b=1$	$b=1$	$b=0.98$	$b=0.98$	$b=1$	$b=1$	$b=0.98$	$b=1$	$b=1$	$b=0.9641$	$b=0.9554$
	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-6$	$k=1-7$	$k=1-6$	$k=1-7$
	$\chi^2=107$	$\chi^2=130$	$\chi^2=97$	$\chi^2=123$	$\chi^2=131$	$\chi^2=117$	$\chi^2=163$	$\chi^2=145$	$\chi^2=221$	$\chi^2=168$	$\chi^2=250$	$\chi^2=231$	$\chi^2=277$	$\chi^2=422$	$\chi^2=220$	$\chi^2=333$
	$\langle t \rangle=16.18$	$\langle t \rangle=16.08$	$\langle t \rangle=16.19$	$\langle t \rangle=16.09$	$\langle t \rangle=16.17$	$\langle t \rangle=16.11$	$\langle t \rangle=16.09$	$\langle t \rangle=18.46$	$\langle t \rangle=18.42$	$\langle t \rangle=18.49$	$\langle t \rangle=18.45$	$\langle t \rangle=18.50$	$\langle t \rangle=18.49$	$\langle t \rangle=18.49$	$\langle t \rangle=18.50$	$\langle t \rangle=18.49$
t																
8	0.0015	0.0029	0	0.0009	0	0	0	0	0	0	0	0	0	0	0	0
9	0.0090	0.0066	0.0090	0.0073	0.0027	0	0	0	0	0	0	0	0	0	0	0
10	0.0032	0.0130	0	0.0100	0	0	0	0	0	0	0	0	0	0	0	0
11	0.0355	0.0379	0.0346	0.0375	0.0247	0.0230	0.0293	0	0	0	0	0	0	0	0	0
12	0.0579	0.0608	0.0598	0.0626	0.0672	0.0681	0.0688	0.0092	0.0095	0.0067	0.0073	0	0	0	0	0
13	0.1306	0.1335	0.1294	0.1321	0.1384	0.1369	0.1400	0	0	0	0	0	0	0	0	0
14	0.0524	0.0543	0.0555	0.0575	0.0545	0.0577	0.0586	0.0065	0.0242	0.0038	0.0223	0	0	0	0	0.0017
15	0.1297	0.1285	0.1285	0.1268	0.1266	0.1301	0.1282	0.0769	0.0743	0.0776	0.0756	0.0806	0.0762	0.0950	0.0843	0.1013
16	0.1035	0.0990	0.1073	0.1025	0.1069	0.1113	0.1095	0.1113	0.1098	0.1118	0.1096	0.1114	0.1199	0.1140	0.1047	0.0976
17	0.1514	0.1509	0.1536	0.1524	0.1678	0.1720	0.1712	0.1106	0.1061	0.1104	0.1063	0.1254	0.1083	0.0972	0.1389	0.1322
18	0.1159	0.1071	0.1152	0.1060	0.1206	0.1182	0.1064	0.2054	0.1903	0.2014	0.1861	0.1993	0.2108	0.2020	0.1895	0.1781
19	0.0290	0.0260	0.0342	0.0302	0.0138	0.0280	0.0255	0.0351	0.0444	0.0434	0.0523	0.0058	0.0104	0.0123	0.0020	0.0034
20	0.1325	0.1240	0.1302	0.1218	0.1490	0.1384	0.1341	0.3564	0.3437	0.3549	0.3409	0.4275	0.4263	0.4225	0.4291	0.4286
21	0	0	0	0	0	0	0	0.0425	0.0413	0.0463	0.0465	0.0137	0.0183	0.0203	0.0091	0.00076
22	0.0193	0.0245	0.0184	0.0242	0.0033	0.0070	0.0152	0	0.0107	0	0.0081	0	0	0	0	0
23	0.0285	0.0309	0.0244	0.0282	0.0244	0.0094	0.0131	0.0460	0.0455	0.0436	0.0450	0.0364	0.0298	0.0367	0.0424	0.0495
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The G_{β} is the free energy-per-residue of β sheetformation. The G_{Leu} is the free energy when leucines are aligned in adjacent strands in a β sheet, with $L(t)=1$ when at least one residue position is aligned, and $L(t)=0$ in the absence of such alignment. These leucines are bolded in the schematic registries of Figure 6.5. The $G(t)_{\text{sc}}^{\text{WT}}$ is the sum of free energies of membrane insertion of sidechains for residues V2 to t-1, with sidechain energy relative to Ala, and g^{WT} is a scaling factor that is <1 and may help to account for the positive free energy of membrane insertion of the Fp backbone.⁵⁶ There are earlier studies that support membrane insertion of WT Fp starting near V2.¹⁵

²⁶ The $f(t)^{\text{WT}}$, $t=11-20$, were fitted with Equation 6.19 and encompass $\sim 95\%$ of the total WT Fp population. The $L(t)=1$ for $t=13, 15, 17, 18$, and 20 , and $L(t)=0$ for $t=11, 12, 14, 16$, and 19 . Fitting was done using $RT = 0.6$ kcal/mole and variation of the parameters C^{WT} , G_{β}^{WT} , $G_{\text{Leu}}^{\text{WT}}$, and g^{WT} . Figure 6.7a displays bar plot comparison between $f(t)^{\text{WT}}$ and values from Equation 6.19 best-fitting, along with a bar plot of the three contributions to $G(t)$. Table D6 presents the numerical values. The fitting $R^2 = 0.88$, the typical magnitude of a residual is ~ 0.01 , $G_{\beta}^{\text{WT}} = -0.113 \pm 0.038$ kcal/mole, $G_{\text{Leu}}^{\text{WT}} = -0.350 \pm 0.079$ kcal/mole, and $g^{\text{WT}} = 0.129 \pm 0.040$. The negative-signed $G_{\text{Leu}}^{\text{WT}}$ contribution is illustrated by larger $f(t)^{\text{WT}}$ for $t=15$ vs. 14 or 16 . These three registries all have the same value of $g^{\text{WT}} \times G_{\text{sc}}(t)^{\text{WT}}$ but differ in the presence ($t=15$) vs. absence ($t=14, 16$) of aligned leucines, Figure 6.5. The extension of the β sheet for registry t over the region from A1 to t is supported by significant intensity assigned to β sheet chemical shifts in ^{13}C NMR spectra of membrane-bound Fp, with ^{13}C -labeled sites between A1 and A21.³⁵ Models different than Equation 6.19 resulted in poorer fitting, i.e. smaller R^2 . Examples of these alternate models were: (1) not including G_{β}^{WT} and/or not including $G_{\text{Leu}}^{\text{WT}}$ contributions to free energy; (2) setting $L(t)$ as the number of leucine and/or phenylalanine residue positions that are aligned in the registry; (3) not including the g^{WT} parameter; and (4) calculating $G_{\text{sc}}^{\text{WT}}(t)$ as the sum starting at a specific residue between G3 and L7 and ending at the corresponding residue between $t-2$ and $t-6$.

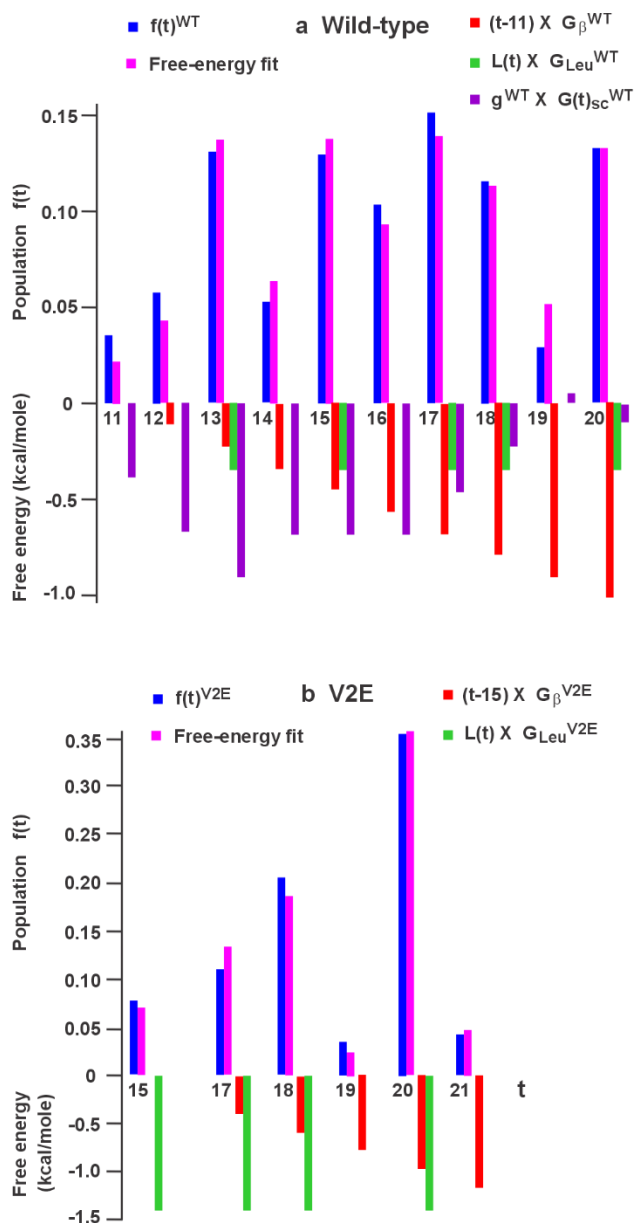


Figure 6.7. Plots of $f(t)$ based on REDOR data, blue bars, and the values calculated from fitting to a sum of free energy contributions, magenta bars, using Equation 6.19 for panel a, WT and Equation 6.20 for panel b, V2E. The $f(t)$ values were determined by unconstrained fittings with $b=0.98$, Figure 6.6 and Table 6.3. The free energy contributions from β sheet length, leucine alignment, and membrane insertion are also displayed as red, green, and purple bars. The fittings were done using $t=11-20$ for WT and $t=15, 17-21$ for V2E. These ranges include $\sim 95\%$ and $\sim 85\%$ of the total registry populations, respectively. The $f(t)$ values and free energies are numerically presented in Tables D6 (WT) and D7 (V2E).

The $f(t)^{V2E}$, $t=15-21$, were also fitted and encompass ~95% of the total V2E Fp population. Fittings were first done using Equation 6.19, with $G_{sc}(t)^{V2E}$ calculated as the sum of sidechain free energies. Separate fittings were done for insertion of the I4→t-3, A6→t-5, and L7→t-6 regions, and in all cases, the best-fit $g^{V2E} \approx 0$. This result correlates with the shallower membrane insertion for V2E vs. WT Fp that has been observed in previous studies.^{15,26} The $f(t)^{V2E}$ fitting was then done using Equation 6.20, i.e. without a contribution to free energy from membrane insertion, and $R^2 = 0.78$, $G_{\beta}^{WT} = -0.184 \pm 0.056$ kcal/mole, and $G_{Leu}^{WT} = -1.21 \pm 0.44$ kcal/mole. The typical residual had magnitude ≈ 0.01 except for $t=16$ which fitted poorly with residual ≈ 0.1 . When $f(16)^{V2E}$ wasn't included, $R^2 = 0.98$, $G_{\beta}^{V2E} = -0.195 \pm 0.021$ kcal/mole, and $G_{Leu}^{V2E} = -1.40 \pm 0.22$ kcal/mole, Figure 6.7b and Table D7.

6.4 Discussion

6.4.1 Broad registry distributions for WT and fusion-defective V2E Fp

This study elucidates the quantitative populations of strand registries of the intermolecular antiparallel β sheet of the membrane-bound Fp, the N-terminal domain of HIV glycoprotein 41 kD. The Fp plays a critical role in gp41-catalyzed fusion between the HIV and target cell membranes which is an early and required step in infection. Fp significance is evidenced by point mutations like V2E which don't affect gp160 spike density but result in complete loss of fusion and infection.^{12,13,41} Most fusion mechanisms propose that the Fp binds the target membrane early in the fusion process. Initial apposition of target and HIV membranes has a 25 kcal/mole barrier in uncatalyzed fusion, and this barrier may be reduced by target membrane with bound Fp in conjunction with thermostable final hairpin structure of gp41 and HIV membrane with bound Tmd, Figure 6.1(i).³ The acyl chains of lipid molecules next to the Fp are more disordered than chains of more distant lipids and this disordering likely reduces the energy barriers between different membrane intermediates during fusion.¹⁹

Prior to the present study, earlier work had shown that the Fp adopts intermolecular antiparallel β sheet structure when bound to membrane with mole fraction cholesterol ≈ 0.3 which is similar to the fraction in membranes of host cells of HIV.^{28,35} Analyses of REDOR NMR spectra of two samples with differently-labeled Fp's were consistent with a fraction of the Fp's adopting the $t=16$ and $t=17$ registries and with another significant population of Fp's adopting other registries that could not be determined from the data.³¹ The present study provides REDOR NMR data from 35 samples with differently-labeled Fp's and the subsequent analyses result in the full antiparallel

registry distributions for both WT and V2E Fp, Tables 6.3, D5. There are many populated registries, with ~95% of the total population in $t=11-20$ for WT and $15-21$ for V2E. Comparison of $\Delta S/S_0$ plots for WT vs. V2E shows that for smaller vs. larger u , the $\Delta S/S_0$ are typically greater for WT vs. V2E, with crossover at $u=16$, Figure 6.5. These differences correlate with the registry distribution weighted towards smaller vs. larger t for WT vs. V2E, Figure 6.6.

6.4.2 NMR linewidths support broad registry distributions

For either S_0 or ΔS spectra, the peak ^{13}CO shifts are typical for β sheet rather than other secondary structures.⁵⁵ Each spectral line profile is typically well-fitted to a single Gaussian function with 3-4 ppm linewidth, Figure 6.4 and Table 6.1. This profile may be due to the superposition of unresolved signals with different peak shifts from the individual β sheet registries within the distribution. The labeled ^{13}CO nucleus would have a different surrounding chemical environment in each registry. The Gaussian function is often associated with a population distribution and the Fp spectral linewidths are much broader than the ~1 ppm linewidth typical for a single-site backbone ^{13}C signal in a membrane-bound protein with a unique structure.^{23, 57} The multiple-registry explanation for the broad Fp linewidths is also supported by $\delta_{\text{FWHM,WT}} > \delta_{\text{FWHM,V2E}}$ for most u , with typical difference ≈ 0.5 ppm, Table 6.1. The greater linewidths for WT vs. V2E correlate with the larger number of populated registries for WT vs. V2E. For example, the unconstrained $b=0.98$ fittings result in 12 values of t with $f(t)^{\text{WT}} > 0.01$ vs. only 8 values of t with $f(t)^{\text{V2E}} > 0.01$, Figure 6.6 and Table 6.3.

6.4.3 Registry distributions are similar with unconstrained and constrained models and for Fp with and without C-terminal hairpin

Figure 6.6 and Table 6.3 display $f(t)$ populations based on unconstrained fittings of data with τ_k between 2.2 and 40.2 ms. The fitting quality is evidenced by $\chi^2 = 107$ for WT and 145 for V2E which are close to 102, the number of fitted data. The fitting model is based on rigorous spin physics and there is quantitative agreement between the model-calculated $\Delta S/S_0$ in the absence of nearby lb $^{13}\text{CO}/\text{lb } ^{15}\text{N}$ spin pairs and the experimental $\Delta S/S_0$ in the $u=28$ sample, Figure 6.5 and Table D2.^{31, 50, 53, 54} For $u=8-24$, the largest deviations between experimental and fitted $\Delta S/S_0$ are typically for $\tau_k = 40.2$ ms. Relative to shorter τ_k , the larger deviations for $\tau_k = 40.2$ ms may be due to the greater effect of ^{13}CO couplings to more distant ^{15}N that aren't included in the $\gamma_{t1,t2}(\tau_k)$ calculations, Figure 6.2 and Table D2. This reasoning is supported by: (1) typically smaller χ^2 for

fittings done with SIMPSON-derived $\gamma(\tau_k)$ multiplied by a factor $b=0.98$, with largest effect on the $\tau_k = 40.2$ ms fitting residuals; and (2) when the $\tau_k = 48.2$ ms data are included in the fitting, the increase in χ^2 is typically greater than 17, the increase in number of data, Table D5. For a specific t , the typical variation in $f(t)$ is only ~ 0.01 among these different unconstrained fittings, even though the χ^2 values can vary by >100 .

The unconstrained fitting model is based on absence of correlation between adjacent registries, Figure 6.2b and Equation 6.18. The REDOR data were also fitted by the constrained model in which all Fp's within one β sheet have a single registry, Figure 6.2a and Equation 6.12. The $f(t)$ vs. t trends are similar for unconstrained and constrained fittings, as is reasonable because the same experimental data are fitted, and the $f(t)$ for a specific t typically differ by <0.02 , Tables 6.3, D5. The $\gamma(\tau_k)$ are a little different for the unconstrained vs. constrained models, Table D2. The average value of t is very similar across all unconstrained and constrained fittings, with $\langle t \rangle^{\text{WT}} = 16.132 \pm 0.048$ and $\langle t \rangle^{\text{V2E}} = 18.475 \pm 0.028$.

Earlier NMR studies of membrane-bound WT Fp+hairpin protein evidenced Fp's with antiparallel β sheet structure and multiple populated registries, Figure 6.8a.^{32, 33} This result supported interleaved strands from two hairpin trimers. The parallel coiled-coil alignment of the three N-helices in each trimer could favor constrained Fp registries within the β sheet. The populations of a few specific registries in Fp+hairpin were probed using NMR detection of proximity between single ^{13}CO labels in adjacent strands, in particular by measurement of ^{13}CO signal dephasing due to dipolar coupling between nearby ^{13}C nuclei. When residue v is ^{13}CO -labeled, the couplings are largest for registries $t = 2v-1$ and $t = 2v$. After accounting for n_a contributions, the long-time dephasings for $v = 4, 7, 8, 11, \text{ and } 12$ were $\sim 0.04, 0.3, 0.3, 0.06, \text{ and } 0.02$, which agree semi-quantitatively with the $f(2v-1)^{\text{WT}} + f(2v)^{\text{WT}}$ sums of the present study of $\sim 0.001, 0.18, 0.23, 0.02, \text{ and } 0.03$ for unconstrained fitting and $\sim 0, 0.19, 0.23, 0.003, \text{ and } 0.02$ for constrained fitting, Table 6.3. This agreement supports the $f(t)^{\text{WT}}$ distribution of HIV gp41 in its final hairpin state to be similar to the $f(t)^{\text{WT}}$ distribution of the present study.

6.4.4 Broad registry distribution may be advantageous for chronic infection by HIV

To our knowledge, the broad registry distributions of Fp are largely unprecedented in peptides and proteins. The unconstrained $f(t)^{\text{WT}}$, $t=11-20$, were fitted with Equation 6.19, see Figure 6.7a and Table D6. The total free energies, $G(t)^{\text{WT}}$, were sums of contributions: (1) $t \times G_{\beta}^{\text{WT}}$, β sheet length;

(2) $L(t) \times G_{\text{Leu}}^{\text{WT}}$, with $L(t) = 1$ or 0 for presence vs. absence of aligned leucines in adjacent strands, Figure 6.5; and (3) $g^{\text{WT}} \times G(t)_{\text{sc}}^{\text{WT}}$, sidechain membrane insertion. The typically negative values of all three contributions may be from the hydrophobic effect, specifically release of water solvating the Fp because there is: (1) β sheet hydrogen bonding; (2) packing of leucine sidechains; and (3) Fp solvation by lipid acyl chains. As t increases from $t = 13$, there is generally a tradeoff between the $t \times G_{\beta}$ and $g^{\text{WT}} \times G_{\text{sc}}(t)^{\text{WT}}$ contributions which are typically becoming more negative and positive, respectively, Figure 6.7a and Table D6. Relative to $t = 11, 12, 14, 16,$ and 19 , the larger $f(t)^{\text{WT}}$ values for $t = 13, 15, 17, 18,$ and 20 correlate with the free energy contribution from aligned leucines, $G_{\text{Leu}}^{\text{WT}} = -0.35$ kcal/mole. The populated β sheet registries will likely bind the target membrane, which in conjunction with thermostable hairpin will reduce the barrier to initial membrane apposition, Figure 6.8a. The bound Fp will also reduce barriers between later membrane intermediates, in part by inducing larger amplitudes of acyl chain motions with resulting increased rates of formation of new bilayers during fusion. This is evidenced by earlier NMR and X-ray studies showing when Fp is bound, lipids still adopt bilayer phase but have greater chain mobility, particularly for lipids next to Fp.^{19, 48, 58}

The Fp β sheet is likely inserted in a single leaflet rather than traversing the bilayer, Figure 6.8, as is reasonable for interleaved Fp's from different hairpin trimers and also consistent with earlier observation that multiple residues within the G5-L12 region contact the lipid chain termini.¹⁵ Sheets with more negative $G_{\text{sc}}^{\text{WT}}$ are likely more deeply-inserted and will induce larger perturbations of neighboring lipids. For the unconstrained model, the $G_{\text{sc}}^{\text{WT}}$ might be an average over the different registries, and for the constrained model, each sheet has a single specific t and therefore $G_{\text{sc}}(t)^{\text{WT}}$ and membrane insertion depth. This registry-dependent depth hypothesis is supported by an earlier NMR study that probed proximity between specific ^{13}C O nuclei in the Fp and specific ^2H nuclei in the lipid acyl chains. The NMR data were only reasonably understood with two Fp β sheet populations, one inserted close to the bilayer center and the other with shallower insertion.⁵⁹ The magnitude of lipid perturbation will also correlate positively with $N_{\text{lipid,nb}}(t)$, the number of sheet-neighboring lipids, with $N_{\text{lipid,nb}}(t) \propto \text{sheet area} \propto t$. The dependence of WT fusion activity on t is considered using $t = 12-20$ which includes >90% of the total registry population, Table 6.3. There may be similar fusion activities for most of these registries, based on $t=12-16$ having more negative $g^{\text{WT}} \times G_{\text{sc}}^{\text{WT}}(t)$, range between -0.91 and -0.67 kcal/mole, and smaller $N_{\text{lipid,nb}}(t)$, whereas $t=17-20$ have less negative $g^{\text{WT}} \times G_{\text{sc}}^{\text{WT}}(t)$, range between -0.46 and

0.01 kcal/mole, but larger $N_{lipid,nb}(t)$, Figure 6.7a and Table D6. Similar fusion activities among most populated registries would also confer fusion activity to unconstrained sheets with mixed registries.

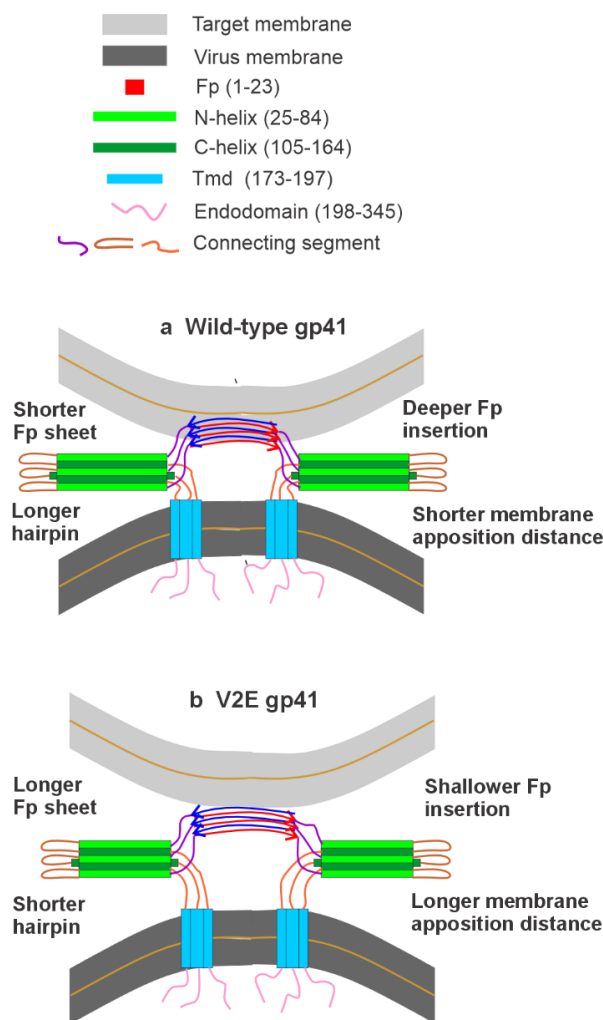


Figure 6.8. Structural models for (a) WT and (b) V2E gp41. The structural differences are proposed to result in longer membrane apposition distance and slower fusion rate for V2E vs. WT. The longer vs. shorter Fp sheets are observed in the present study, Figure 6.6, and are correlated with shorter vs. longer hairpins observed in an earlier study. There are ~12 fewer helical hairpin residues in V2E vs. WT. The longer V2E Fp sheets are proposed to result in unfolding of ~6 N-helix residues (those closest to Fp). This leads to unfolding of ~6 C-helix residues (those closest to the Tmd) and therefore longer distance between the apposed membranes. One consequence of the longer distance is that stalk formation will require larger-amplitude lipid chain motions, and happen at a slower rate for V2E vs. WT, Figure 6.1. Relative to WT, there is also shallower membrane location of V2E Fp sheets which will reduce the probability of lipid chain protrusion into the aqueous region. The endodomain forms a well-defined structure that is not displayed in this figure.

HIV is a chronic infection that relies on constant mutation to escape neutralization by the immune system, with Fp being one of the neutralization epitopes.⁶⁰ Relative to a narrow registry distribution, the broad distribution for the Fp and hypothesized similar fusion activities for most registries may permit escape mutants that moderately change the $f(t)$ distribution while remaining fusion-competent. This evolutionary advantage hypothesis is supported by comparison between the non-homologous fusion peptides of HIV gp41 and influenza virus Ha2. Influenza is also membrane-enveloped, and fusion is likely catalyzed by a mechanism similar to Figure 6.1, with Ha2 replacing gp41.¹ However, influenza is an acute rather than chronic infection and doesn't experience long-time immune pressure within a single person. The long- vs. short-time immune pressures of HIV vs. influenza may be manifested in the contrasts between: (1) HIV fusion peptide which has sequence variety among patient isolate strains and an intermolecular β sheet structure with broad registry distribution; vs. (2) influenza fusion peptide which exhibits very high sequence conservation among viral isolates and adopts two very similar and monomeric helical hairpin structures.^{11, 61}

6.4.5 Longer V2E registries with shallower membrane insertion can explain V2E-dominant loss of fusion, infection, and hairpin helicity

Comparison of the $\Delta S/S_0$ data among WT samples shows largest values for $u=17$ whereas V2E samples show the largest values for $u=20$, Figure 6.5. This correlates with the registry distribution weighted towards larger t for V2E, Figure 6.6 and Table 6.3. The $f(t)^{V2E}$ distribution, $t=15, 17-21$ was fitted with a free energy function that doesn't include the sidechain membrane insertion contribution, $g \times G_{sc}(t)$, that was needed to fit the $f(t)^{WT}$ distribution, Figure 6.7b and Table D7. This result correlates with earlier observation of deeper membrane insertion for WT vs. V2E Fp.²⁶ There would therefore be greater lipid chain displacement for WT vs. V2E which correlates with $\sim 10\times$ greater vesicle fusion induced by WT vs. V2E Fp.³⁶ The ratios of best-fit $G_{Leu}^{V2E}/G_{Leu}^{WT} \approx 4$ and $G_{\beta}^{V2E}/G_{\beta}^{WT} \approx 1.7$ might reflect larger hydrophobic effect in the higher water content environment of shallower V2E vs. deeper WT Fp, both for leucines and for other hydrophobic sidechains that become more tightly-packed in β sheets vs. monomeric peptides. When $f(16)^{V2E}$ (≈ 0.11) is included in the fitting, there is poor agreement with the fitted value of ~ 0.01 . This disagreement might be due to a free energy contribution from membrane insertion specific to $t = 16$, which is the longest registry that doesn't include polar residues, other than E2, Figure 6.5. The $t = 17$ registry is used for comparison, with $f(17)^{V2E} = 0.11$ and the Equation 6.20 calculated

$G(17)^{V2E} = -4.7$ kcal/mole. The $G(16)^{V2E}$ could be similar to the calculated $G(17)^{V2E}$ using a sum of $16 \times G_{\beta}^{V2E}$, -3.1 kcal/mole, and a contribution proportional to $G_{sc}(16)^{V2E} = -6.3$ kcal/mole, which is calculated for insertion going from I4, G5, or A6 to t-3, t-4, or t-5, respectively. There isn't a substantial contribution to $G(23)^{V2E}$ from E2-R22 salt bridges, based on $f(23)^{V2E} < 0.05$ which is only ~ 0.015 larger than $f(23)^{WT}$, Figure 6.6 and Table 6.3.

Understanding the Fp role in fusion has typically been based on: (1) point mutations which impair HIV fusion and infection; (2) differences in structure and motion of lipid molecules in membranes with vs. without Fp; and (3) computational studies. V2E has been the most well-studied mutation because V2E results in complete loss of HIV gp160-mediated fusion and infection and because V2E is dominant in mixed WT/V2E gp160 trimers.^{12,13} A recent study showed V2E-dominant losses of helicity and vesicle fusion for mixed WT/V2E Fp+hairpin trimers, and also that these losses were quantitatively-similar to the losses in fusion and infection with mixed gp160 trimers.⁴¹ The mole fraction V2E-dependences of losses in fusion, infection, and helicity of gp160 and Fp+hairpin were globally-fitted and supported a requirement of cooperativity between at least 6 WT molecules for efficient fusion and infection, i.e. 2 WT trimers, Figure 6.8.⁴¹ One conundrum raised in this earlier study is the mechanism by which V2E, which is ~ 20 residues N-terminal of the hairpin, changes the hyperthermostable and autonomously-folding structure of the hairpin trimer.^{8, 62, 63} This question is addressed by the finding in the present study that relative to WT, the V2E registry distribution is weighted to larger t. Several unstructured residues are likely required between the C-terminus of the Fp β sheet and the hairpin N-helix, so the longer V2E β sheets likely result in unfolding of the N-helix region closest to the Fp, Figure 6.8b. The C-helices pack in the exterior grooves of the N-helix trimeric bundle, so the loss of N-helix residues likely results in unfolding of the C-helix region closest to the Tmd, Figure 6.8b. There are ~ 12 fewer helical residues in the hairpin for V2E vs. WT which would correspond to ~ 6 fewer N-helix and ~ 6 fewer C-helix residues.⁴¹ The N- and C-helices have heptad repeat sequences, so these losses correspond to about one repeat in both the N- and C-helices.⁶³ Relative to WT, the unfolding of N- and C-helix segments will result in a longer distance between the apposed membranes, Figure 6.8b. Stalk formation, Figure 6.1(ii), will therefore require larger-amplitude lipid chain protrusion into the aqueous phase and will happen at a slower rate. The protrusion probability will also be smaller because of shallower membrane location of V2E Fp sheets. V2E-dominant phenotypes in mixed WT/V2E trimers are understood by trimeric bundle formation by the N-helices and the bundle N-

terminus being determined by the longest registry in the Fp β sheet. There are consequently V2E-dominant shorter lengths of the C-helices and two hairpins, and longer membrane apposition distance, Figure 6.8b. This hypothesis is supported by V2E-dominant loss of hairpin helicity and vesicle fusion induced by Fp+hairpin trimers whereas V2E isn't dominant for vesicle fusion induced by Fp-only.^{40, 41}

6.4.6 Quantitative determination of broad structural distributions using REDOR NMR of multiple differently-labeled samples

To our knowledge, this study is one of only a few reports of experimentally-based determination of the populations of >10 different structures of a molecule in a sample. Structural populations can sometimes be obtained in computational simulations, although it can be difficult to ascertain whether the populations represent a thermodynamic equilibrium distribution, and there is dependence of the energies that underlie the distribution on the force field parameters of the simulation.^{44, 64} The present study highlights an underutilized strength of solid-state NMR to experimentally determine broad population distributions, particularly with a pulse sequence like REDOR whose data are less-sensitive to: (1) small variations in instrument parameters during acquisition over several days; and (2) similar small variations in parameters among acquisitions for different samples, Figure D1 and Table D1.⁵⁰ Determination of the full registry distribution required REDOR NMR data for eighteen differently labeled samples, both because the distribution is broad and because the ¹³C NMR signals from different registries aren't resolved.³⁵ Earlier REDOR NMR studies provided distributions that were incomplete because only a few samples were used.^{29,31,35} There was also uncertainty when there were multiple labels in individual samples.^{29,31} The >10 structural populations determined in this study is larger than the few structures typically distinguished with other experimental approaches such as crystal diffraction or cryo-EM.⁶⁵ There are typically only a few structures in earlier NMR studies based on spectrally-resolved signals among the structures.^{22,57,66}

The REDOR approach with multiple differently-labeled samples may be applicable to determining structural distributions for other important systems such as intrinsically-disordered proteins or amyloid formation.⁶⁷ The extent to which broad (but defined) structural distributions are important in biological processes is an open question. For processes like membrane fusion which are molecular movement rather than chemical reactions, the present study shows that a broad structural distribution can confer catalytic function that is mutationally-robust. This may be evolutionarily

advantageous for a pathogen like HIV which requires constant mutation to escape neutralization by the immune system.

6.5 Conclusions

This study describes the quantitative determinations of the populations of registries of the intermolecular antiparallel β sheets of the membrane-bound HIV gp41 Fp, both WT and fusion-defective V2E mutant. The highest energy barrier in fusion is likely initial close apposition of membranes. This barrier can be partially compensated by a combination of N-terminal Fp in target cell membrane, C-terminal Tmd in HIV membrane, and intervening thermostable hairpin, Figure 6.1. The Fp sheets also induce larger-amplitude motions of chains of neighboring lipids, with consequential increase in the rates of transformation between different membrane structures during fusion. For both WT and V2E, the data of the present study are the time-dependent ^{13}C - ^{15}N REDOR NMR dephasings of 17 differently-labeled Fp's, Figure 6.5. Each registry is denoted by t , the number of aligned residues in adjacent Fp molecules, and the REDOR data for each labeled sample depend on only a few values of t . For both WT and V2E, the REDOR data from all 17 samples were globally-analyzed to determine the populations, $f(t)$'s, using either a constrained model in which all the molecules within a single sheet have a single registry (value of t), and an unconstrained model without this restriction. There isn't spectral resolution of ^{13}C signals from different registries so the approach of the present study differs from the more typical NMR approach of determining populations from the intensities of resolved signals. For either fitting model, the derived $f(t)^{\text{WT}}$ and $f(t)^{\text{V2E}}$ distributions have a large number of populated registries, i.e. there are typically ~ 10 values of t for which $f(t) > 0.01$, Figure 6.6 and Table 6.3. There are quite different distributions for WT vs. V2E, with weighting to shorter vs. longer registries. For a specific t , the $f(t)$ populations typically agree within 0.01 among different constrained fittings, with similar agreement among different unconstrained fittings. For a specific t , the $f(t)$ agree typically within 0.02 between the constrained and unconstrained models. The average value of t is highly-conserved for all unconstrained and constrained fittings, with $\langle t \rangle^{\text{WT}} = 16.132 \pm 0.048$ and $\langle t \rangle^{\text{V2E}} = 18.475 \pm 0.028$.

The $f(t)^{\text{WT}}$ are well-fitted using a sum of free energy contributions that depend on β sheet length, presence vs. absence of aligned leucines, and sidechain membrane insertion, Figure 6.7. There are likely similar lipid perturbations and fusion activities of most populated registries of WT sheets. This is based on the tradeoff for shorter vs. longer registries of smaller vs. larger numbers of sheet-

neighboring lipids and larger vs. smaller free energies of membrane insertion, where the latter likely correlate with deeper vs. shallower location of the sheet in the membrane. HIV is a chronic infection that relies on mutation to escape neutralization by the immune system. It is likely that many Fp mutations will result in moderate changes in the registry distribution. However, fusion competence will typically be retained because most registries are fusion-active. The broad registry distribution of FP may therefore be more mutationally-robust than a single or narrow distribution of Fp structures that would be more easily disrupted by continual mutation. The $f(t)^{V2E}$ are well-fitted using a sum of free energy contributions for β sheet length and aligned leucines but not sidechain membrane insertion, which correlates with V2E insertion that is shallower than WT and results in less perturbation of neighboring lipids by V2E Fp. The longer V2E sheets are also correlated with shorter C-terminal hairpins with consequential larger distance between initially-apposed membranes, Figure 6.8. This distance change is expected to significantly slow the rate of stalk formation, where the stalk is a membrane intermediate for which the outer- but not inner-leaflets of the fusing bilayers are contiguous. Longer apposition distance means that stalk formation will require larger-amplitude motions of the lipid chains in the initially-apposed HIV and target cell membranes. In mixtures of WT and V2E gp41, V2E exhibits dominance. The present study supports V2E dominance in Fp sheet length with consequent dominance in hairpin shortening and reductions in fusion and infection. More generally, the present study provides a rare and informative demonstration of the power of solid-state NMR to determine a broad distribution of molecular structures in the absence of spectral resolution between the structures.

REFERENCES

- (1) White, J. M.; Delos, S. E.; Brecher, M.; Schornberg, K. Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* 2008, 43 (3), 189-219.
- (2) Harrison, S. C. Viral membrane fusion. *Virology* 2015, 479, 498-507. DOI: 10.1016/j.virol.2015.03.043. Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research* 2020, 178, 104792. DOI: 10.1016/j.antiviral.2020.104792.
- (3) Boonstra, S.; Blijleven, J. S.; Roos, W. H.; Onck, P. R.; van der Giessen, E.; van Oijen, A. M. Hemagglutinin-mediated membrane fusion: A biophysical perspective. *Ann. Revs. Biophys.* 2018, 47, 153-173. DOI: 10.1146/annurev-biophys-070317-033018.
- (4) Blumenthal, R.; Durell, S.; Viard, M. HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* 2012, 287 (49), 40841-40849.
- (5) Gallaher, W. R. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* 1987, 50 (3), 327-328.
- (6) Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997, 89 (2), 263-273. Kwon, B.; Lee, M.; Waring, A. J.; Hong, M. Oligomeric structure and three-dimensional fold of the HIV gp41 membrane-proximal external region and transmembrane domain in phospholipid bilayers. *J. Am. Chem. Soc.* 2018, 140 (26), 8246-8259, Article. DOI: 10.1021/jacs.8b04010. Piai, A.; Fu, Q. S.; Sharp, A. K.; Bigli, B.; Brown, A. M.; Chou, J. J. NMR model of the entire membrane-interacting region of the HIV-1 fusion protein and its perturbation of membrane morphology. *J. Am. Chem. Soc.* 2021, 143 (17), 6609-6615, Article. DOI: 10.1021/jacs.1c01762.
- (7) Caffrey, M.; Cai, M.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Covell, D. G.; Gronenborn, A. M.; Clore, G. M. Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* 1998, 17 (16), 4572-4584. Yang, Z. N.; Mueser, T. C.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Hyde, C. C. The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution. *J. Struct. Biol.* 1999, 126 (2), 131-144.
- (8) Buzon, V.; Natrajan, G.; Schibli, D.; Campelo, F.; Kozlov, M. M.; Weissenhorn, W. Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* 2010, 6 (5), e1000880.
- (9) Pancera, M.; Zhou, T. Q.; Druz, A.; Georgiev, I. S.; Soto, C.; Gorman, J.; Huang, J. H.; Acharya, P.; Chuang, G. Y.; Ofek, G.; et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 2014, 514 (7523), 455-461. DOI: 10.1038/nature13808. Ward, A. B.; Wilson, I. A. The HIV-1 envelope glycoprotein structure: nailing down a moving target. *Immunol. Revs.* 2017, 275 (1), 21-32. DOI: 10.1111/imr.12507.
- (10) Durell, S. R.; Martin, I.; Ruyschaert, J. M.; Shai, Y.; Blumenthal, R. What studies of fusion peptides tell us about viral envelope glycoprotein-mediated membrane fusion. *Mol.*

- Membr. Biol. 1997, 14 (3), 97-112.
- (11) Ochsenbauer, C.; Edmonds, T. G.; Ding, H. T.; Keele, B. F.; Decker, J.; Salazar, M. G.; Salazar-Gonzalez, J. F.; Shattock, R.; Haynes, B. F.; Shaw, G. M.; et al. Generation of Transmitted/Founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages. *J. Virol.* 2012, 86 (5), 2715-2728. DOI: 10.1128/jvi.06157-11. Parrish, N. F.; Gao, F.; Li, H.; Giorgi, E. E.; Barbian, H. J.; Parrish, E. H.; Zajic, L.; Iyer, S. S.; Decker, J. M.; Kumar, A.; et al. Phenotypic properties of transmitted founder HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110 (17), 6626-6633, Article. DOI: 10.1073/pnas.1304288110.
 - (12) Freed, E. O.; Myers, D. J.; Risser, R. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. *Proc. Natl. Acad. Sci. U.S.A.* 1990, 87 (12), 4650-4654.
 - (13) Freed, E. O.; Delwart, E. L.; Buchschacher, G. L., Jr.; Panganiban, A. T. A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U.S.A.* 1992, 89 (1), 70-74.
 - (14) Yang, J.; Prorok, M.; Castellino, F. J.; Weliky, D. P. Oligomeric b-structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers. *Biophys. J.* 2004, 87 (3), 1951-1963.
 - (15) Jia, L. H.; Liang, S.; Sackett, K.; Xie, L.; Ghosh, U.; Weliky, D. P. REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. *J. Magn. Reson.* 2015, 253, 154-165. DOI: 10.1016/j.jmr.2014.12.020.
 - (16) Chernomordik, L. V.; Kozlov, M. M. Mechanics of membrane fusion. *Nat. Struct. Mol. Biol.* 2008, 15 (7), 675-683.
 - (17) Larsson, P.; Kasson, P. M. Lipid tail protrusion in simulations predicts fusogenic activity of influenza fusion peptide mutants and conformational models. *PLoS Comp. Biol.* 2013, 9 (3), e1002950. DOI: 10.1371/journal.pcbi.1002950.
 - (18) Legare, S.; Lagguee, P. The influenza fusion peptide promotes lipid polar head intrusion through hydrogen bonding with phosphates and N-terminal membrane insertion depth. *Proteins-Struc. Func. Bioinform.* 2014, 82 (9), 2118-2127. DOI: 10.1002/prot.24568. Victor, B. L.; Lousa, D.; Antunes, J. M.; Soares, C. M. Self-assembly molecular dynamics simulations shed light into the interaction of the influenza fusion peptide with a membrane bilayer. *J. Chem. Inform. Model.* 2015, 55 (4), 795-805. DOI: 10.1021/ci500756v. Michalski, M.; Setny, P. Membrane-bound configuration and lipid perturbing effects of Hemagglutinin subunit 2 N-terminus investigated by computer simulations. *Front. Mol. Biosci.* 2022, 9, 826366, Article. DOI: 10.3389/fmolb.2022.826366.
 - (19) Ghosh, U.; Weliky, D. P. 2H nuclear magnetic resonance spectroscopy supports larger amplitude fast motion and interference with lipid chain ordering for membrane that contains beta sheet human immunodeficiency virus gp41 fusion peptide or helical hairpin influenza virus hemagglutinin fusion peptide at fusogenic pH. *Biochim. Biophys. Acta* 2020, 1862 (10), 183404. DOI: 10.1016/j.bbamem.2020.183404. Ghosh, U.; Weliky, D. P. Rapid 2H NMR transverse relaxation of perdeuterated lipid acyl chains of membrane with

- bound viral fusion peptide supports large-amplitude motions of these chains that can catalyze membrane fusion. *Biochemistry* 2021, 60 (35), 2637-2651. DOI: 10.1021/acs.biochem.1c00316.
- (20) Zhang, Y. J.; Ghosh, U.; Xie, L.; Holmes, D.; Severin, K. G.; Weliky, D. P. Lipid acyl chain protrusion induced by the influenza virus hemagglutinin fusion peptide detected by NMR paramagnetic relaxation enhancement. *Biophys. Chem.* 2023, 299, 107028, Article. DOI: 10.1016/j.bpc.2023.107028.
- (21) Jaroniec, C. P.; Kaufman, J. D.; Stahl, S. J.; Viard, M.; Blumenthal, R.; Wingfield, P. T.; Bax, A. Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain. *Biochemistry* 2005, 44 (49), 16167-16180. Li, Y. L.; Tamm, L. K. Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers. *Biophys. J.* 2007, 93 (3), 876-885.
- (22) Gabrys, C. M.; Weliky, D. P. Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles. *Biochim. Biophys. Acta* 2007, 1768 (12), 3225-3234.
- (23) Wasniewski, C. M.; Parkanzky, P. D.; Bodner, M. L.; Weliky, D. P. Solid-state nuclear magnetic resonance studies of HIV and influenza fusion peptide orientations in membrane bilayers using stacked glass plate samples. *Chem. Phys. Lipids* 2004, 132 (1), 89-100.
- (24) Zheng, Z.; Yang, R.; Bodner, M. L.; Weliky, D. P. Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry* 2006, 45, 12960-12975.
- (25) Qiang, W.; Weliky, D. P. HIV fusion peptide and its cross-linked oligomers: efficient syntheses, significance of the trimer in fusion activity, correlation of β strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry* 2009, 48, 289-301.
- (26) Qiang, W.; Sun, Y.; Weliky, D. P. A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106 (36), 15314-15319.
- (27) Lai, A. L.; Moorthy, A. E.; Li, Y. L.; Tamm, L. K. Fusion activity of HIV gp41 fusion domain is related to its secondary structure and depth of membrane insertion in a cholesterol-dependent fashion. *J. Mol. Biol.* 2012, 418 (1-2), 3-15.
- (28) Lorizate, M.; Sachsenheimer, T.; Glass, B.; Habermann, A.; Gerl, M. J.; Krausslich, H. G.; Brugger, B. Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines. *Cell. Microbiol.* 2013, 15 (2), 292-304. DOI: 10.1111/cmi.12101.
- (29) Yang, J.; Weliky, D. P. Solid state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry* 2003, 42, 11879-11890.
- (30) Sackett, K.; Shai, Y. The HIV fusion peptide adopts intermolecular parallel β -sheet

- structure in membranes when stabilized by the adjacent N-terminal heptad repeat: A ¹³C FTIR study. *J. Mol. Biol.* 2005, 350 (4), 790-805.
- (31) Schmick, S. D.; Weliky, D. P. Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide. *Biochemistry* 2010, 49 (50), 10623-10635.
- (32) Ratnayake, P. U.; Sackett, K.; Nethercott, M. J.; Weliky, D. P. pH-dependent vesicle fusion induced by the ectodomain of the human immunodeficiency virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation. *Biochim. Biophys. Acta* 2015, 1848 (1), 289-298. DOI: 10.1016/j.bbamem.2014.07.022.
- (33) Sackett, K.; Nethercott, M. J.; Zheng, Z. X.; Weliky, D. P. Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J. Mol. Biol.* 2014, 426 (5), 1077-1094.
- (34) Bodner, M. L.; Gabrys, C. M.; Parkanzky, P. D.; Yang, J.; Duskin, C. A.; Weliky, D. P. Temperature dependence and resonance assignment of ¹³C NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn. Reson. Chem.* 2004, 42 (2), 187-194.
- (35) Qiang, W.; Bodner, M. L.; Weliky, D. P. Solid-state NMR spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *J. Am. Chem. Soc.* 2008, 130 (16), 5459-5471.
- (36) Pereira, F. B.; Goni, F. M.; Nieva, J. L. Liposome destabilization induced by the HIV-1 fusion peptide effect of a single amino acid substitution. *FEBS Lett.* 1995, 362 (2), 243-246. Gabrys, C. M.; Qiang, W.; Sun, Y.; Xie, L.; Schmick, S. D.; Weliky, D. P. Solid-state nuclear magnetic resonance measurements of HIV fusion peptide ¹³CO to lipid ³¹P proximities support similar partially inserted membrane locations of the a Helical and b sheet peptide structures. *J. Phys. Chem. A* 2013, 117 (39), 9848-9859. DOI: 10.1021/jp312845w.
- (37) Brandenburg, O. F.; Magnus, C.; Regoes, R. R.; Trkola, A. The HIV-1 entry process: A stoichiometric view. *Trends Microbiol.* 2015, 23 (12), 763-774. DOI: 10.1016/j.tim.2015.09.003.
- (38) Zhu, P.; Liu, J.; Bess, J.; Chertova, E.; Lifson, J. D.; Grise, H.; Ofek, G. A.; Taylor, K. A.; Roux, K. H. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 2006, 441 (7095), 847-852.
- (39) Sougrat, R.; Bartesaghi, A.; Lifson, J. D.; Bennett, A. E.; Bess, J. W.; Zabransky, D. J.; Subramaniam, S. Electron tomography of the contact between T cells and SIV/HIV-1: Implications for viral entry. *PLoS Pathogens* 2007, 3 (5), e63.
- (40) Kliger, Y.; Aharoni, A.; Rapaport, D.; Jones, P.; Blumenthal, R.; Shai, Y. Fusion peptides

- derived from the HIV type 1 glycoprotein 41 associate within phospholipid membranes and inhibit cell-cell fusion. Structure-function study. *J. Biol. Chem.* 1997, 272 (21), 13496-13505.
- (41) Rokonujjaman, M.; Sahyouni, A.; Wolfe, R.; Jia, L. H.; Ghosh, U.; Weliky, D. P. A large HIV gp41 construct with trimer-of-hairpins structure exhibits V2E mutation-dominant attenuation of vesicle fusion and helicity very similar to V2E attenuation of HIV fusion and infection and supports: (1) hairpin stabilization of membrane apposition with larger distance for V2E; and (2) V2E dominance by an antiparallel beta sheet with interleaved fusion peptide strands from two gp41 trimers. *Biophys. Chem.* 2023, 293, 106933. DOI: 10.1016/j.bpc.2022.106933.
- (42) Yang, H.; Staveness, D.; Ryckbosch, S. M.; Axtman, A. D.; Loy, B. A.; Barnes, A. B.; Pande, V. S.; Schaefer, J.; Wender, P. A.; Cegelski, L. REDOR NMR reveals multiple conformers for a Protein Kinase C ligand in a membrane environment. *ACS Cent. Sci.* 2018, 4 (1), 89-96, Article. DOI: 10.1021/acscentsci.7b00475.
- (43) Clore, G. M.; Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 2009, 109 (9), 4108-4139, Review. DOI: 10.1021/cr900033p. Lorieau, J. L.; Louis, J. M.; Schwieters, C. D.; Bax, A. pH-triggered, activated-state conformations of the influenza hemagglutinin fusion peptide revealed by NMR. *Proc. Natl. Acad. Sci. U.S.A.* 2012, 109, 19994-19999.
- (44) Panahi, A.; Feig, M. Conformational sampling of Influenza fusion peptide in membrane bilayers as a function of termini and protonation states. *J. Phys. Chem. B* 2010, 114 (3), 1407-1416.
- (45) Yang, R.; Prorok, M.; Castellino, F. J.; Weliky, D. P. A trimeric HIV-1 fusion peptide construct which does not self-associate in aqueous solution and which has 15-fold higher membrane fusion rate. *J. Am. Chem. Soc.* 2004, 126 (45), 14722-14723.
- (46) Lapatsanis, L.; Miliadis, G.; Froussios, K.; Kolovos, M. Synthesis of N-2,2,2-(trichloroethoxycarbonyl)-L-amino acids and N-(9-fluorenylmethoxycarbonyl)-L-amino acids involving succinimidoxyl anion as a leaving group in amino-acid protection. *Synthesis* 1983, 8, 671-673.
- (47) Yang, J.; Parkanzky, P. D.; Khunte, B. A.; Canlas, C. G.; Yang, R.; Gabrys, C. M.; Weliky, D. P. Solid state NMR measurements of conformation and conformational distributions in the membrane-bound HIV-1 fusion peptide. *J. Mol. Graph. Model.* 2001, 19 (1), 129-135. Yang, J.; Gabrys, C. M.; Weliky, D. P. Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide. *Biochemistry* 2001, 40 (27), 8126-8137.
- (48) Gabrys, C. M.; Yang, R.; Wasniewski, C. M.; Yang, J.; Canlas, C. G.; Qiang, W.; Sun, Y.; Weliky, D. P. Nuclear magnetic resonance evidence for retention of a lamellar membrane phase with curvature in the presence of large quantities of the HIV fusion peptide. *Biochim. Biophys. Acta* 2010, 1798 (2), 194-201.
- (49) Kashefi, M.; Malik, N.; Struppe, J. O.; Thompson, L. K. Carbon-nitrogen REDOR to

- identify ms-timescale mobility in proteins. *J. Magn. Reson.* 2019, 305, 5-15, Article. DOI: 10.1016/j.jmr.2019.05.008.
- (50) Gullion, T. Introduction to rotational-echo, double-resonance NMR. *Concepts Magn. Reson.* 1998, 10 (5), 277-289.
- (51) Gullion, T.; Baker, D. B.; Conradi, M. S. New, compensated Carr-Purcell sequences. *J. Magn. Reson.* 1990, 89 (3), 479-484. Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. Heteronuclear decoupling in rotating solids. *J. Chem. Phys.* 1995, 103 (16), 6951-6958.
- (52) Mueller, K. T. Analytical solutions for the time evolution of dipolar-dephasing NMR signals. *J. Magn. Reson. Ser. A* 1995, 113 (1), 81-93.
- (53) Bak, M.; Rasmussen, J. T.; Nielsen, N. C. SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* 2000, 147 (2), 296-330.
- (54) Bak, M.; Schultz, R.; Vosegaard, T.; Nielsen, N. C. Specification and visualization of anisotropic interaction tensors in polypeptides and numerical simulations in biological solid-state NMR. *J. Magn. Reson.* 2002, 154 (1), 28-45.
- (55) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99 (26), 16742-16747. Zhang, H. Y.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* 2003, 25 (3), 173-195.
- (56) Moon, C. P.; Fleming, K. G. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* 2011, 108 (25), 10174-10177. DOI: 10.1073/pnas.1103979108.
- (57) Bhate, M. P.; Wylie, B. J.; Tian, L.; McDermott, A. E. Conformational dynamics in the selectivity filter of KcsA in response to potassium ion concentration. *J. Mol. Biol.* 2010, 401 (2), 155-166, Article. DOI: 10.1016/j.jmb.2010.06.031.
- (58) Tristram-Nagle, S.; Chan, R.; Kooijman, E.; Uppamoochikkal, P.; Qiang, W.; Weliky, D. P.; Nagle, J. F. HIV fusion peptide penetrates, disorders, and softens T-cell membrane mimics. *J. Mol. Biol.* 2010, 402 (1), 139-153.
- (59) Xie, L.; Jia, L. H.; Liang, S.; Weliky, D. P. Multiple locations of peptides in the hydrocarbon core of gel-phase membranes revealed by peptide ¹³C to lipid ²H rotational-echo double-resonance solid-state nuclear magnetic resonance. *Biochemistry* 2015, 54 (3), 677-684. DOI: 10.1021/bi501211x.
- (60) Kong, R.; Xu, K.; Zhou, T. Q.; Acharya, P.; Lemmin, T.; Liu, K.; Ozorowski, G.; Soto, C.; Taft, J. D.; Bailer, R. T.; et al. Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science* 2016, 352 (6287), 828-833. DOI: 10.1126/science.aae0474. Xu, K.; Acharya, P.; Kong, R.; Cheng, C.; Chuang, G.-Y.; Liu, K.; Louder, M. K.; O'Dell, S.; Rawi, R.; Sastry, M.; et al. Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. *Nat. Medicine* 2018, 24 (6), 857-867.

- DOI: 10.1038/s41591-018-0042-6. Yuan, M.; Cottrell, C. A.; Ozorowski, G.; van Gils, M. J.; Kumar, S.; Wu, N. C.; Sarkar, A.; Torres, J. L.; de Val, N.; Copps, J.; et al. Conformational plasticity in the HIV-1 fusion peptide facilitates recognition by broadly neutralizing antibodies. *Cell Host & Microbe* 2019, 25 (6), 873-883, Article. DOI: 10.1016/j.chom.2019.04.011.
- (61) Nobusawa, E.; Aoyama, T.; Kato, H.; Suzuki, Y.; Tateno, Y.; Nakajima, K. Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virology* 1991, 182 (2), 475-485. DOI: 10.1016/0042-6822(91)90588-3. Lorieau, J. L.; Louis, J. M.; Bax, A. The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107 (25), 11341-11346. Ghosh, U.; Xie, L.; Jia, L. H.; Liang, S.; Weliky, D. P. Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis. *J. Am. Chem. Soc.* 2015, 137 (24), 7548-7551. DOI: 10.1021/jacs.5b04578.
- (62) Sackett, K.; Nethercott, M. J.; Epand, R. F.; Epand, R. M.; Kindra, D. R.; Shai, Y.; Weliky, D. P. Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. *J. Mol. Biol.* 2010, 397 (1), 301-315.
- (63) Liang, S.; Ratnayake, P. U.; Keinath, C.; Jia, L.; Wolfe, R.; Ranaweera, A.; Weliky, D. P. Efficient fusion at neutral pH by Human Immunodeficiency Virus gp41 trimers containing the fusion peptide and transmembrane domains. *Biochemistry* 2018, 57 (7), 1219-1235. DOI: 10.1021/acs.biochem.7b00753.
- (64) Worch, R.; Dudek, A.; Borkowska, P.; Setny, P. Transient excursions to membrane core as determinants of influenza virus fusion peptide activity. *Int. J. Mol. Sci.* 2021, 22 (10), 5301, Article. DOI: 10.3390/ijms22105301.
- (65) Cai, Y. F.; Zhang, J.; Xiao, T. S.; Peng, H. Q.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 2020, 369 (6511), 1586-1592. DOI: 10.1126/science.abd4251.
- (66) Cady, S. D.; Mishanina, T. V.; Hong, M. Structure of amantadine-bound M2 transmembrane peptide of Influenza A in lipid bilayers from magic-angle-spinning solid-state NMR: The role of Ser31 in amantadine binding. *J. Mol. Biol.* 2009, 385 (4), 1127-1141, Article. DOI: 10.1016/j.jmb.2008.11.022. Qiang, W.; Yau, W. M.; Lu, J. X.; Collinge, J.; Tycko, R. Structural variation in amyloid- β fibrils from Alzheimer's disease clinical subtypes. *Nature* 2017, 541 (7636), 217-221. DOI: 10.1038/nature20814.
- (67) Qiang, W.; Yau, W. M.; Tycko, R. Structural evolution of Iowa mutant β -amyloid fibrils from polymorphic to homogeneous states under repeated seeded growth. *J. Am. Chem. Soc.* 2011, 133 (11), 4018-4029. DOI: 10.1021/ja109679q. Jeon, J.; Yau, W. M.; Tycko, R. Early events in amyloid- β self-assembly probed by time-resolved solid state NMR and light scattering. *Nat. Comm.* 2023, 14 (1), 2964. DOI: 10.1038/s41467-023-38494-6.

Chapter 7 Summary and future work

My research in the past six years was mainly focus on the function of fusion peptide of HA of influenza virus in facilitating viral and target membranes merging together, secondary structure of HM, and REDOR data analysis to determine the distribution of registry of β sheet oligomer when binding to lipid bilayers. ssNMR was the major characterization for all the three projects described in the thesis.

Fusion peptide of influenza virus has several functions which are critical to promote fusion. It can reduce the curvature energy, reduce interstitial void energy, and dehydrate membranes. Other than those functions, another major role of fusion peptide is to induce the lipid protrusion as the prerequisite of intermediate stalk formation. Lipid protrusion is the movement of lipid acyl chains out of the hydrophobic interior of the membrane into the aqueous phase. Such protrusion could aid joining the leaflets of the viral and target cell membranes, which is critical for fusion between influenza virus and host cells. Increased lipid protrusion in membrane with IFP has been detected in computational molecular dynamics but there aren't yet direct experimental data that support protrusion.¹ We investigated lipid protrusion in membrane with IFP using paramagnetic relaxation enhancement (PRE) of ^{13}C nuclear magnetic resonance (NMR) spectra. We found a greater enhancement of the ^{13}C NMR transverse relaxation rate (R_2) of lipid acyl chain sites with Mn^{2+} at the presence of IFP. The enhancement is inversely proportional to the separation between ^{13}C and location of Mn^{2+} and greater enhancement implies increased population of protruded chain caused by IFP. Therefore, we successfully provided directly experimental evidence of the increased probability of lipid protrusion with the presence of IFP.

We observed there was an obvious greater enhancement at 2,2' and 3,3' carbons which are close to lipid headgroups and the location of paramagnetic source. To avoid peptide aggregation of IFP, the optimal peptide concentration in lipid membrane system is ~ 4 molar % of the lipids. Consideration the low concentration of IFP, the lipid protrusion induced by the presence of IFP is a minor effect. We found that for 5 % $[\text{Mn}^{2+}]$ samples, all carbons exhibited similar extent of relaxation enhancement with and without the presence of IFP. We believe the reason is that $[\text{Mn}^{2+}]$ is too concentrated that the enhancement caused by Mn^{2+} covered minor effect from the protruded chain because of the IFP. The 0.75 % and 1 % $[\text{Mn}^{2+}]$ increased the relaxation rate up to three folds and it should not mask the influence from the existence of IFP. PRE measurement with those two $[\text{Mn}^{2+}]$ would be further support the major function of IFP is to increase the probability of

lipid protrusion.

Other than that, this project is a following-up of Dr. Shuang Liang's work as a PhD candidate. She probed the PRE enhancement of the dipalmitoylphosphatidylcholine DPPC-D8 and DPPC-D10, which locate at the middle and the tail of the acyl chain respectively. DPPC-D10 is the methylene and methyl groups at the end of the acyl chain substituted with deuterons. The signals of deuterons in ssNMR T_2 experiment overlapped together so it was difficult to determine the relaxation rate by integrated intensity of the peak. To solve the problem, we switched the sample system to DPPC-D6 in which only protons of methyl groups are replaced with deuterons. However, the experiment results were not reproducible because of the failure in controlling rehydration level.

It is worth to evaluate the effect of IFP on protons since the protons are more sensitive to the influence of IFPs. Dipolar coupling of ^1H to electron spins is 4 times larger than ^{13}C to electrons so the enhancement Γ_2 of ^1H would be 16 times greater than that of ^{13}C so the protons are more sensitive to the relaxation rate change caused by paramagnetic source. Proton detection might provide a more detailed picture about the movement of acyl chain with respect to the presence of IFPs.

Chapter 5 described NMR assignment and structural probing of a small protein HM in bacterial inclusion bodies. Recombinant protein production in bacterial hosts nearly always results in deposition of a large Rp fraction in intracellular solid aggregates that are termed inclusion bodies. The IBs fraction is often discarded because solubilization and subsequent refolding is difficult. There is little information about the structure(s) of any Rp in IBs and such information may be useful for developing better solubilization and refolding. The assignment of the spectral crosspeaks based on amino acid type supported that there exist major α helical and minor β sheet structure in HM IBs. The TEM images of HM supported the existence of β sheet structure as well. The assignment is amino acid type assignment which only reflects the type of the amino acid. To better reveal the structure of HM IBs, sequential assignment is necessary. To achieve that, high quality spectra (a well-resolved spectrum with decent peak intensity) of NCOCX, CONCACX, etc are needed. We noticed the fair amount of spectral degeneracy in the collected spectra, indicating it is not possible to do sequential assignment with all the experiments mentioned earlier. Perhaps site specific labelling would aid to solve this problem. For example, Lysine is usually hard to be determined in HM because: (1) the determination heavily relies on the signal of sidechains, especially $\text{C}\gamma$ and $\text{C}\delta$, requiring a strong polarization transfer and long mixing time; (2) HM has a

fair amount of Glu and Gln whose $C\alpha$ and $C\beta$ signal are superimposed with that of Lysine which hinders the unambiguous assignment. If HM could be labeled with Lysine, it would be easier to find out Lysine chemical shift in HM. Moreover, label Ala and Arg simultaneously will be helpful to sequential assignment as the result of there exist AR and RA connectivity in HM. At the slice of arginine ^{15}N chemical shift, there should be CO and CX crosspeak corresponding to alanine in NCOCX spectrum. Similarly, at the ^{13}C O plane of alanine, there should be CA and CX crosspeaks of arginine transferred from ^{15}N of arginine at CONCACX spectrum. In general, a different labeling strategy is a good choice to decipher the structure of HM based on ssNMR characterization.

Detection of ^1H might be an alternative to HM structure determination. ^1H -detection correlation spectra have several advantages: higher sensitivity compared with ^{13}C -detection method because of greater gyromagnetic ratio of protons; less sample amount is needed as ^1H -detection multidimensional spectra are always recorded in a rotor with a very small diameter spun at a relative high speed. ^1H -detection of proteins in ssNMR was prevented from the strong homogenous line broadening arising from proton homonuclear dipolar couplings and its high natural abundance. With the development of NMR hardware to generate strong magnetic field and the ultrafast spinning speed, the detection of ^1H signal becomes practical since the ultrafast spinning speed can effectively reduce the dipolar coupling leading to a narrow linewidth and strong magnetic field improve the sensitivity and resolution.¹ Even with high spinning speed, it is still challenging to probe fully protonated samples due to unfavorable ^1H chemical shift dispersion. The ^1H linewidths under fast MAS at ultrahigh magnetic field are narrow but ^1H chemical shift of proteins are distributed in 10 ppm. Without a good ^1H chemical shift dispersion, it is not possible to do site-specific assignment for most of the protons in the protein. An alternative method to reduce the proton linewidth is to start with perdeuterated proteins and subsequent back-substitute of deuterons with protons. Typically, the perdeuterated sample would be mixed with a buffer containing H_2O and D_2O , only the exchangeable protons in the backbone and side chains can be detected in the subsequent ^1H -detection ssNMR experiment.^{2,3} The shortcoming about ^1H -detection of perdeuterated proteins is the folding of the proteins would affect H-D exchange. Even though the inclusion bodies were reported to have a porous structure, it is likely that part of the folded region would be inaccessible by water.⁴ If a perdeuterated proteins is heavily folded and the region deeply buried in the folded proteins would not be accessible to H_2O , that region will not be protonated,

and no information will be collected in ^1H -detection measurements. On the other hand, lack of protonation of particular amino acid types (like hydrophobic amino acids) could be evidence for folding.

What is really interesting in HM project is the ssNMR cross peaks related to β sheet conformation. TEM images also support the existence of filament. It might be useful to acquire a Cryo-EM image of the HM Ibs to help us get more information about HM structure in inclusion bodies.

Chapter 6 discusses the quantitative determination of the registry distribution of antiparallel β sheet wild-type HFP and its fusion-impaired V513E mutant. Like influenza, HIV infection requires fusion between viral and host cell membranes. There is a ~25-residue fusion peptide (HFP) N-terminal region of the HIV gp41 subunit protein that plays a critical role in fusion and whose sequence is very different from IFP. Qualitative analysis of the NMR data shows there is a distribution of registries (alignments) of hydrogen-bonded residues in neighboring strands. This distribution is significantly different for HFP with the V513E point mutation, and fusion is highly-impaired for gp160 with this mutation. The free-energy fit result supports V513E-HFP mutant lies on the surface of the membrane while WT-HFP inserts into the membranes. V513E-HFP tends to form longer registries than WT-HFP does. Without membrane insertion, it is not likely that V513E-HFP mutant could disrupt the membrane structure to form a stalk intermediate assisting fusion. The longer registries would lead to the C-helix at the hairpin with a shorter length and then create a larger separation between viral and target membranes. Those could be the two major reasons that V513E-HFP is non-fusogenic.

REFERENCES

- (1) Le Marchand, T.; Schubeis, T.; Bonaccorsi, M.; Paluch, P.; Lalli, D.; Pell, A. J.; Andreas, L. B.; Jaudzems, K.; Stanek, J.; Pintacuda, G. ^1H -Detected Biomolecular NMR under Fast Magic-Angle Spinning. *Chem Rev* 2022, 122 (10), 9943–10018. DOI: 10.1021/acs.chemrev.1c00918.
- (2) Fricke, P.; Chevelkov, V.; Zinke, M.; Giller, K.; Becker, S.; Lange, A. Backbone Assignment of Perdeuterated Proteins by Solid-State NMR Using Proton Detection and Ultrafast Magic-Angle Spinning. *Nature Protocols* 2017, 12 (4), 764–782. DOI: 10.1038/nprot.2016.190.
- (3) Chevelkov, V.; Rehbein, K.; Diehl, A.; Reif, B. Ultrahigh Resolution in Proton Solid-State NMR Spectroscopy at High Levels of Deuteration. *Angewandte Chemie - International Edition* 2006, 45 (23), 3878–3881. DOI: 10.1002/anie.200600328.
- (4) Singh, S. M.; Panda, A. K. Solubilization and Refolding of Bacterial Inclusion Body Proteins. *Journal of Bioscience and Bioengineering* 2005, 99 (4), 303–310. DOI: 10.1263/jbb.99.303.

APPENDIX A NMR FILE LOCATION

Chapter 3:

Data were saved at `sftp://nmrsu@ssnmr1.cem.msu.edu`

Figure 3.2: `/opt/nmrdata/weliky/Bruker3.2mmHCN/KBr_adam/12`

Figure 3.5: `/opt/nmrdata/weliky/Bruker3.2mmHCN/KBr_adam/13`

Figure 3.6: `/opt/nmrdata/weliky/Bruker3.2mmHCN/13CO_Ala/1`

Figure 3.8: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/2`

Figure 3.9: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/3`

Figure 3.11: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/8`

Figure 3.13: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/4`

Figure 3.14: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/4`

Figure 3.15: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/7`

Figure 3.17: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/14`

`/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/16`

Figure 3.18: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/11`

Figure 3.19: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/14`

Figure 3.20: `/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/12`

`/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/10`

`/opt/nmrdata/weliky/Bruker3.2mmHCN/MLF_5C_8kHz/20`

Chapter 4:

Data were saved at `sftp://nmrsu@ssnmr1.cem.msu.edu`

Figure 4.3:

(a). `/opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/4`

(b). `/opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/46`

(c). `/opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/62`

(d). `/opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/66`

Table 4.1:

0% Mn²⁺: `/opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/23`

0.5% Mn²⁺: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/47
0.75% Mn²⁺: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/50
1% Mn²⁺: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/55
1.25% Mn²⁺: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/59

Figure 4.5:

(a). /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/23
(b). /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/47
(c). /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/70
(d). /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/69

Table B1:

PC_PG_Rep.1: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/23
PC_PG_Rep.2: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/14
PC_PG_0.5%Mn²⁺_Rep.1: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/47
PC_PG_0.5%Mn²⁺_Rep.2: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/47
PC_PG_Fp_Rep.1: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/70
PC_PG_Fp_Rep.2: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/63
PC_PG_Fp_0.5%Mn²⁺_Rep.1: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/69
PC_PG_Fp_0.5%Mn²⁺_Rep.2: /opt/nmrdata/Weliky/Bruker3.2mmHCN/Yijin_PC_PG/67

Chapter 5:

Data were saved at sftp://nmr@nmr800b8.cem.msu.edu unless noted

U-HM:

NCACX: /opt/nmrdata/user/data/zhan1128/nmr/HM/62
NCACX (25%NUS): /opt/nmrdata/user/data/zhan1128/nmr/HM/67
NCOCX: /opt/nmrdata/user/data/zhan1128/nmr/HM/58
CONCACX: /opt/nmrdata/user/data/zhan1128/nmr/HM/210

Leu-Rev-HM:

DARR-30ms: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/5
DARR-100ms: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/18

DARR-500ms: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/6

DARR-100ms (13.5 kHz): /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/24

NCACX: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/201

NCOCX: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/301

1-3-¹³C-Glycerol-HM:

DARR-30ms (NUS50%): /opt/nmrdata/user/data/zhan1128/nmr/1_3_13C_Glec_HM/2

DARR-100ms (NUS50%): /opt/nmrdata/user/data/zhan1128/nmr/1_3_13C_Glec_HM/3

NCACX (NUS25%): /opt/nmrdata/user/data/zhan1128/nmr/1_3_13C_Glec_HM/401

2-¹³C-Glycerol-HM:

DARR-30ms: /opt/nmrdata/user/data/zhan1128/nmr/2_13C_Glec_HM/4

DARR-100ms: /opt/nmrdata/user/data/zhan1128/nmr/2_13C_Glec_HM/5

NCACX (NUS25%): /opt/nmrdata/user/data/zhan1128/nmr/2_13C_Glec_HM/25

Figure 5.17:

U-HM: /opt/nmrdata/user/data/zhan1128/nmr/HM/25

Leu-Rev-HM: /opt/nmrdata/user/data/zhan1128/nmr/Leu_Rev_HM/1

1,3-¹³C-Glyc-HM: /opt/nmrdata/user/data/zhan1128/nmr/1_3_13C_Glec_HM/1

2-¹³C-Glyc-HM: /opt/nmrdata/user/data/zhan1128/nmr/2_13C_Glec_HM/1

Figure 5.16:

Data were saved at sftp://nmrsu@ssnmr1.cem.msu.edu

CP: /opt/nmrdata/weliky/Bruker3.2mmHCN/HM/50

INEPT: /opt/nmrdata/weliky/Bruker3.2mmHCN/HM/51

APPENDIX B SUPPORTING INFORMATION FOR CHAPTER 4

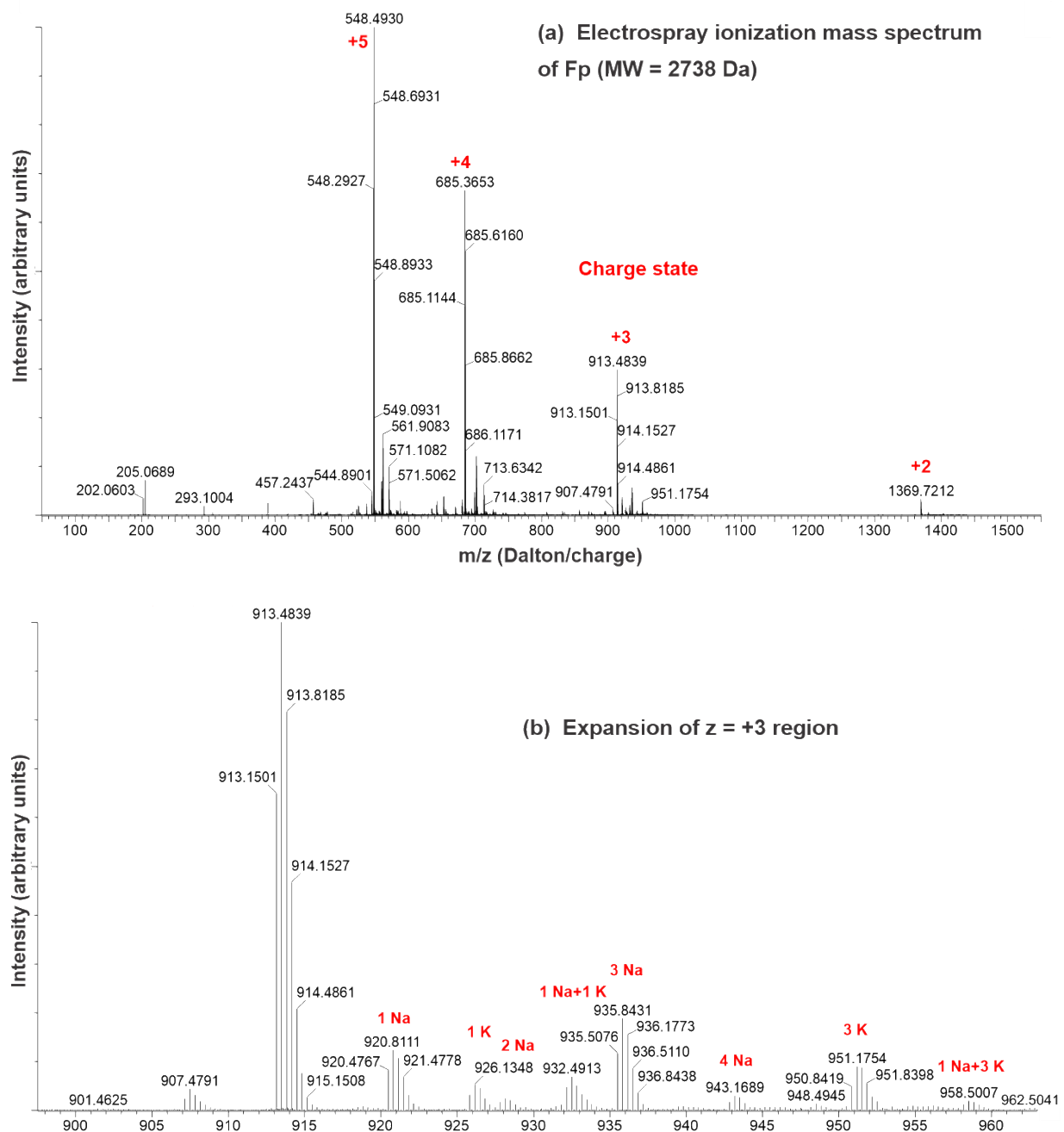


Figure B1. (a) Electrospray ionization mass spectrum of Fp and (b) expansion of $z=+3$ region. The calculated Fp mass is 2738 Da and peak regions in **a** are assigned to charge (z) states. Panel **b** shows clusters of isotopomer peaks with assignments of higher mass clusters to adducts with Na^+ and K^+ ions replacing H^+ .

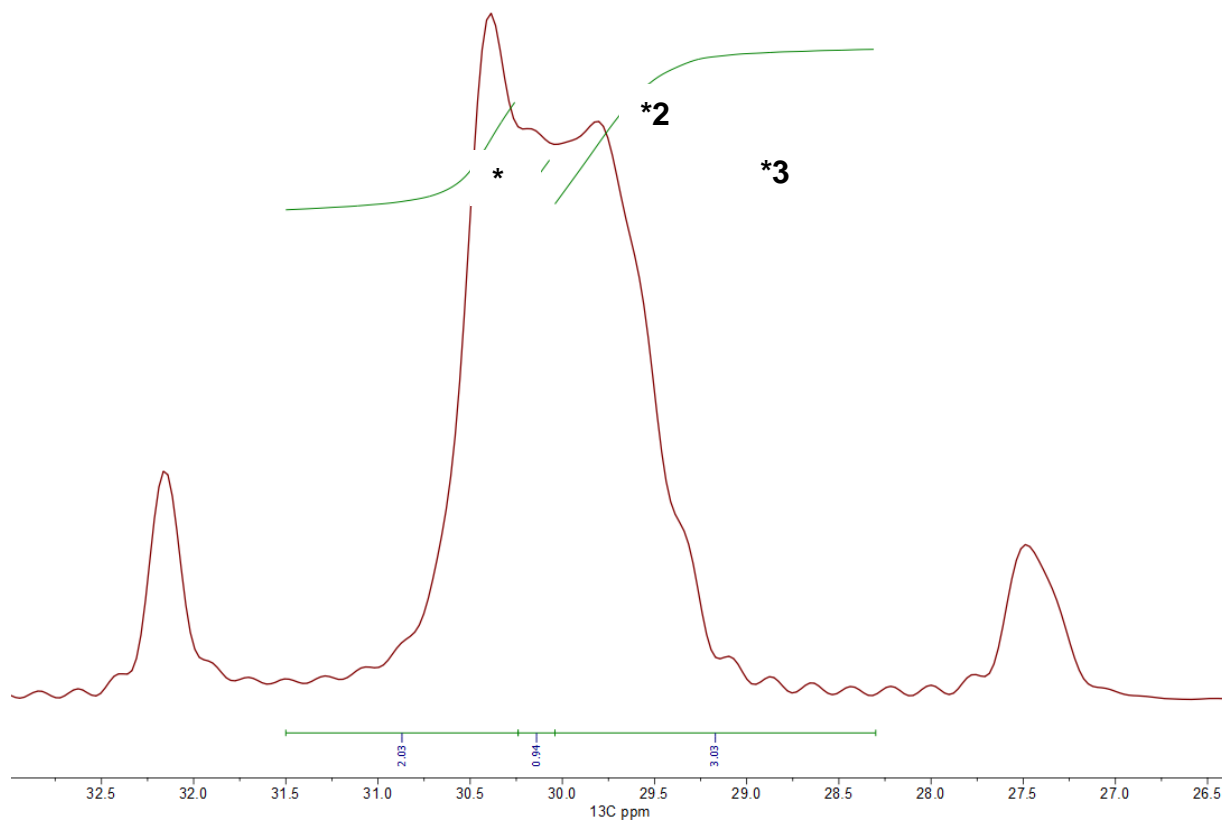


Figure B2. *1, *2, and *3 integration ranges.

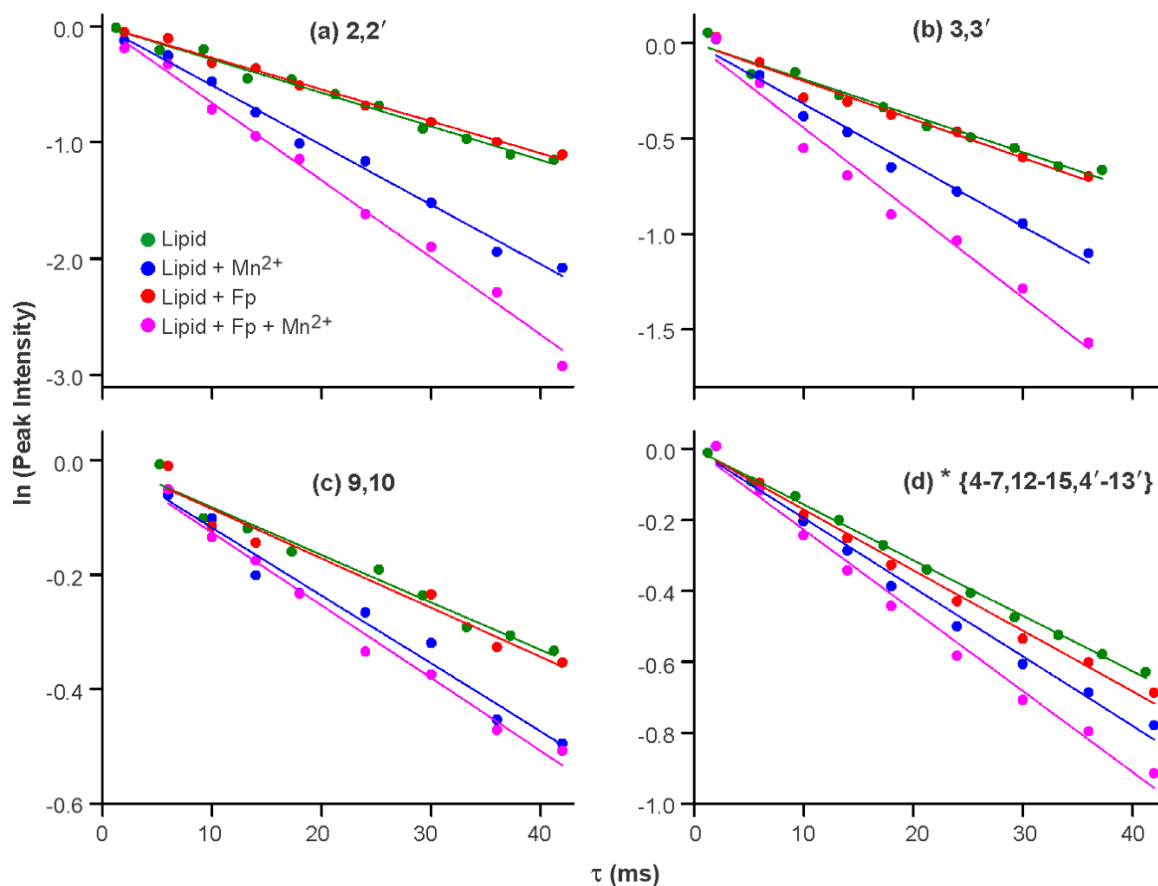


Figure B3. Plots of $\ln(\text{Integrated peak intensity})$ vs. dephasing time (τ) and linear fitting for the Lipid, Lipid + Mn^{2+} , Lipid + Fp, and Lipid + Fp + Mn^{2+} samples. Data and fittings are displayed for the (a) 2,2'; (b) 3,3'; (c) 9,10; and (d) * peaks. The * peak is a superposition of signals from the 4-7, 12-15, and 4'-13' sites. The plots of Integrated peak intensity vs. τ and exponential decay fitting are presented in Figure 4.5 in chapter 4.

Table B1. Site-specific ^{13}C R_2 's of acyl chains of POPC:POPG (4:1) membrane and Mn^{2+} dependence (uncertainties in parentheses)^a.

^{13}C	% Mn^{2+}				
	0	0.5	0.75	1.0	1.25
2,2'	28.9(1.0)	50.4(1.8)	79.8(3.3)	91.9(3.6)	97.4(5.9)
3,3'	20.0(1.2)	35.6(2.4)	56.6(3.8)	63.0(4.1)	68.0(4.7)
8,11	9.2(0.7)	12.5(1.0)	18.2(1.1)	18.9(1.3)	21.5(1.5)
9,10	9.4(0.9)	12.5(0.8)	19.6(1.8)	21.8(2.1)	21.5(1.3)
16,14'	15.3(0.7)	15.1(0.6)	18.0(1.1)	21.8(1.3)	22.1(1.1)
17,15'	6.8(1.4)	6.6(0.9)	9.9(1.3)	10.9(1.9)	10.5(1.6)
* [4-7, 12-15, 4'-13']	15.8(0.3)	20.6(0.9)	26.1(1.0)	28.5(1.4)	29.3(1.3)
*1 [6'-9']	23.7(0.2)	30.5(0.4)	34.4(0.4)	36.2(1.1)	36.1(0.7)
*2 [7,10',11']	15.1(0.2)	16.4(0.8)	23.3(1.0)	25.2(1.4)	25.8(1.3)
*3 [4-6,12-15, 4',5',12',13']	11.0(0.3)	15.4(1.0)	20.4(1.4)	22.6(1.6)	24.4(1.7)

^a Each ^{13}C transverse relaxation rate (R_2) was determined from best-fitting the integrated NMR peak intensity S vs. delay time τ using $S = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. The fitting uncertainty of R_2 is given in parentheses. The * peak is the superposition of the 4-7, 12-15, and 4'-13' signals. Typical ppm integration ranges for peaks are: 2,2', 33.00-37.00; 3,3', 24.00-26.30; 8,11, 26.50-28.20; 9,10, 128.00-131.00; 16,14', 31.50-33.00; 17,15', 21.50-23.60; *, 28.30-31.50; *1, 30.24-31.50; *2, 30.04-30.24; *3, 28.30-30.04 (Figure B2). The ^{13}C sites that make the largest contributions to the *1, *2, and *3 integration ranges are listed between the brackets. The % $\text{Mn}^{2+} = (\text{mole bound Mn}^{2+})/(\text{mole lipid}) \times 100$.

Table B2. Site-specific ^{13}C Γ_2 's of acyl chains of POPC:POPG (4:1) and Mn^{2+} dependence (uncertainties in parentheses)^a.

^{13}C	% Mn^{2+}			
	0.5	0.75	1.0	1.25
2,2'	21.5(2.1)	51.0(3.5)	63.0(3.8)	68.6(6.0)
3,3'	15.6(2.7)	36.5(4.0)	42.9(4.3)	47.9(5.0)
8,11	3.2(1.2)	8.9(1.3)	9.7(1.4)	12.3(1.6)
9,10	3.1(1.2)	10.2(2.0)	12.4(2.3)	12.1(1.5)
16,14'	-0.2(0.9)	2.7(1.3)	6.5(1.5)	6.8(1.3)
17,15'	-0.2(1.6)	3.1(1.9)	4.1(2.4)	3.7(2.1)
* [4-7, 12-15, 4'-13']	4.7(0.9)	10.3(1.1)	12.7(1.5)	13.5(1.4)
*1 [6'-9']	6.8(0.5)	10.6(0.5)	12.5(1.1)	12.4(0.7)
*2 [7,10',11']	1.3(0.9)	8.2(1.0)	10.1(1.5)	10.7(1.4)
*3 [4-6,12-15, 4',5',12',13']	4.4(1.1)	9.3(1.4)	11.5(1.6)	13.4(1.8)

^aThe Γ_2 values are the differences between the best-fit R_2 values of samples with vs. without Mn^{2+} (Table B1) and the fitting uncertainty of Γ_2 is given in parentheses. The % $\text{Mn}^{2+} = (\text{mole bound } \text{Mn}^{2+})/(\text{mole lipid}) \times 100$.

Table B3. Site-specific ^{13}C transverse relaxation rates of acyl chains of POPC:POPG (4:1) membrane and Mn^{2+} and Fp dependences, with fitted values for replicate datasets (uncertainties in parentheses)^a.

^{13}C	R_2 (s^{-1})							
	w/o Fp, w/o Mn^{2+}		w/o Fp, 0.5% Mn^{2+}		3% Fp, w/o Mn^{2+}		3% Fp, 0.5% Mn^{2+}	
	Rep. 1	Rep. 2	Rep. 1	Rep. 2	Rep. 1	Rep. 2	Rep. 1	Rep. 2
2,2'	28.8(1.4)	30.9(1.2)	51.5(2.1)	50.4(1.8)	28.1(1.2)	31.2(1.2)	63.5(3.2)	64.0(6.0)
3,3'	20.0(1.5)	15.8(1.4)	35.2(2.4)	35.6(2.4)	21.3(2.0)	20.1(1.6)	51.1(4.0)	53.8(2.2)
8,11	9.1(0.7)	8.0(0.9)	12.6(0.9)	12.5(0.8)	13.6(1.6)	11.8(1.5)	19.8(1.7)	22.8(1.0)
9,10	8.4(0.6)	9.0(1.0)	11.9(0.8)	12.5(1.0)	8.7(1.0)	13.9(1.2)	13.0(1.2)	18.5(1.9)
16,14'	15.5(0.7)	13.5(0.8)	15.0(0.6)	15.1(0.6)	13.8(0.8)	16.7(0.8)	15.4(0.7)	18.9(1.0)
17,15'	5.9(0.7)	5.8(0.9)	7.3(0.4)	6.6(0.9)	8.6(0.8)	8.6(1.1)	9.7(1.0)	11.8(1.3)
* [4-7, 12-15, 4'-13']	15.8(0.3)	15.1(0.7)	20.5(0.9)	20.6(0.9)	17.8(0.7)	19.0(0.8)	24.1(1.1)	27.9(0.4)
*1 [6'-9']	24.8(0.2)	24.4(0.3)	27.2(1.5)	30.9(0.4)	25.4(0.5)	30.2(0.3)	30.3(0.3)	34.7(1.6)
*2 [7,10',11']	15.1(0.2)	13.6(0.7)	16.4(0.9)	16.4(0.8)	16.0(0.7)	17.3(0.6)	23.2(1.0)	26.7(0.4)
*3 [4-6,12-15,4',5',12', 13']	11.7(0.3)	10.9(0.9)	18.8(0.9)	15.4(1.0)	15.8(1.2)	14.4(1.0)	23.1(1.8)	24.3(0.8)

^a Each ^{13}C transverse relaxation rate (R_2) was determined from best-fitting the integrated NMR peak intensity S vs. delay time τ using $S(\tau) = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. The fitting uncertainty of R_2 is given in parentheses. The R_2 's are given for fitting of replicate datasets, Rep. 1 and Rep. 2, with Rep. 1 values reported in Table 4.2 in the main manuscript. Typical ppm integration ranges for peaks are: 2,2', 33.00-37.00; 3,3', 24.00-26.30; 8,11, 26.50-28.20; 9,10, 128.00-131.00; 16,14', 31.50-33.00; 17,15', 21.50-23.60; *, 28.30-31.50; *1, 30.24-31.50; *2, 30.04-30.24; *3, 28.30-30.04 (Figure B2). The ^{13}C sites that make the largest contributions to the *1, *2, and *3 integration ranges are listed between the brackets. The % Mn^{2+} = (mole bound Mn^{2+})/(mole lipid) \times 100. The % Fp is calculated using the same type of expression.

Table B4. Site-specific ^{13}C transverse relaxation rates of acyl chains of POPC:POPG (4:1) membrane with 3% Fp and without Mn^{2+} , with fitted values for replicate samples (uncertainties in parentheses)^a.

^{13}C	R_2 (s^{-1})	
	Samp. 1	Samp.2
2,2'	28.1(1.2)	37.3(1.5)
3,3'	21.3(2.0)	20.3(1.2)
8,11	13.6(1.6)	10.7(1.0)
9,10	8.7(1.0)	11.6(1.4)
16,14'	13.8(0.8)	19.9(1.0)
17,15'	8.6(0.8)	9.3(1.4)
* [4-7, 12-15, 4'-13']	17.8(0.7)	18.4(0.5)
*1 [6'-9']	25.4(0.5)	32.5(0.6)
*2 [7,10',11']	16.0(0.7)	16.4(0.4)
*3 [4-6,12-15, 4',5',12',13']	15.8(1.2)	13.9(0.5)

^a Each ^{13}C transverse relaxation rate (R_2) was determined from best-fitting the integrated NMR peak intensity S vs. delay time τ using $S(\tau) = A \times \exp(-R_2 \times \tau)$ where A and R_2 are fitting parameters. The fitting uncertainty of R_2 is given in parentheses. The R_2 's are given for fitting of data from replicate samples, Samp. 1 and Samp. 2, with Samp. 1 values reported in Table 2 in the main manuscript. Typical ppm integration ranges for peaks are: 2,2', 33.00-37.00; 3,3', 24.00-26.30; 8,11, 26.50-28.20; 9,10, 128.00-131.00; 16,14', 31.50-33.00; 17,15', 21.50-23.60; *, 28.30-31.50; *1, 30.24-31.50; *2, 30.04-30.24; *3, 28.30-30.04 (Figure B2). The ^{13}C sites that make the largest contributions to the *1, *2, and *3 integration ranges are listed between the brackets.

APPENDIX C SSNMR SPECTRA AND THS FLUORESCENCE DATA FOR HM

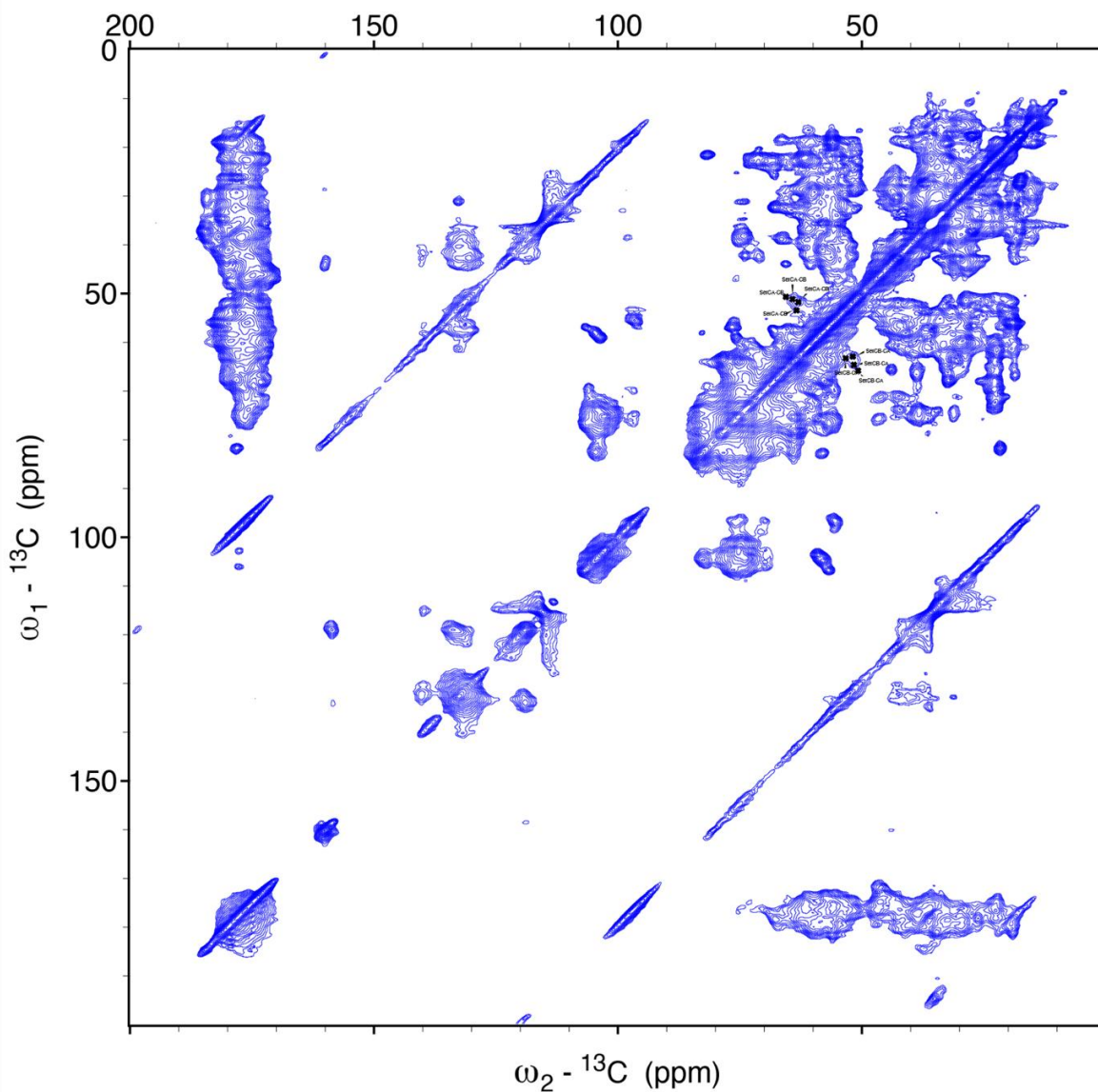
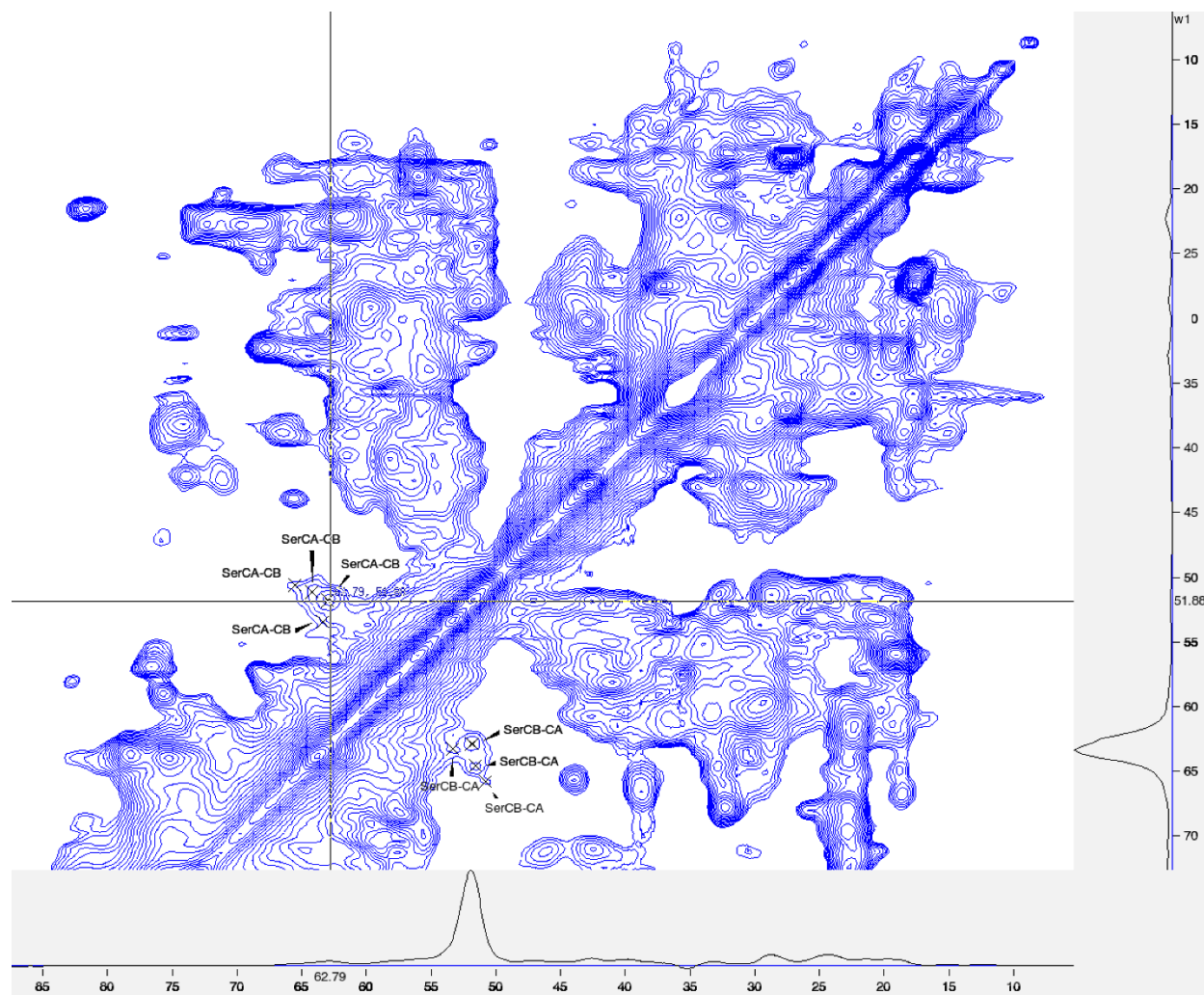


Figure C1. Homonuclear ^{13}C - ^{13}C correlation spectrum of Leu-Rev-HM using DARR with mixing time 30 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate and 128 scans. The spectrum was processed with the window function QSINE, SSB = 2 for both dimensions.

Figure C1 (cont'd)



To evaluate the linewidth of each CB peak, each row spectrum of crosspeak CA-CB was extracted from the 2D spectrum and then peak width at FWHM was determined. The chemical shift and linewidth of each assigned peaks is summarized in Table. The spectra used for linewidth determination is provided as Figure C2.

Table C1. The chemical shift of Ser CB-CA peak and the Linewidth of CB peaks.

	CA	CB	Linewidth/ppm
Ser	50.81	65.10	2.81
	51.54	64.72	2.36
	51.90	62.88	2.81
	53.31	63.34	1.73

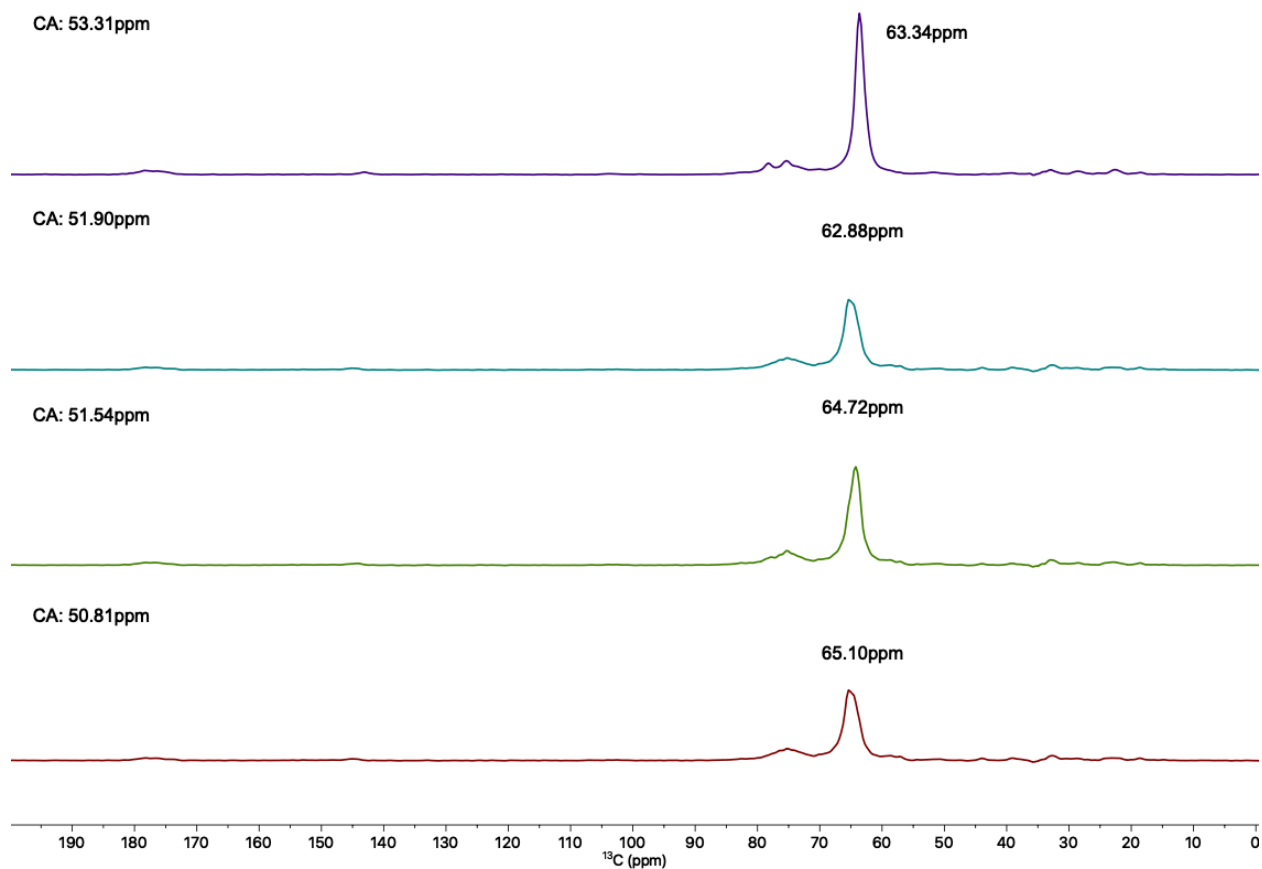


Figure C2. The extracted spectra columns of CA-CB region to evaluate Ser CB peak linewidth.

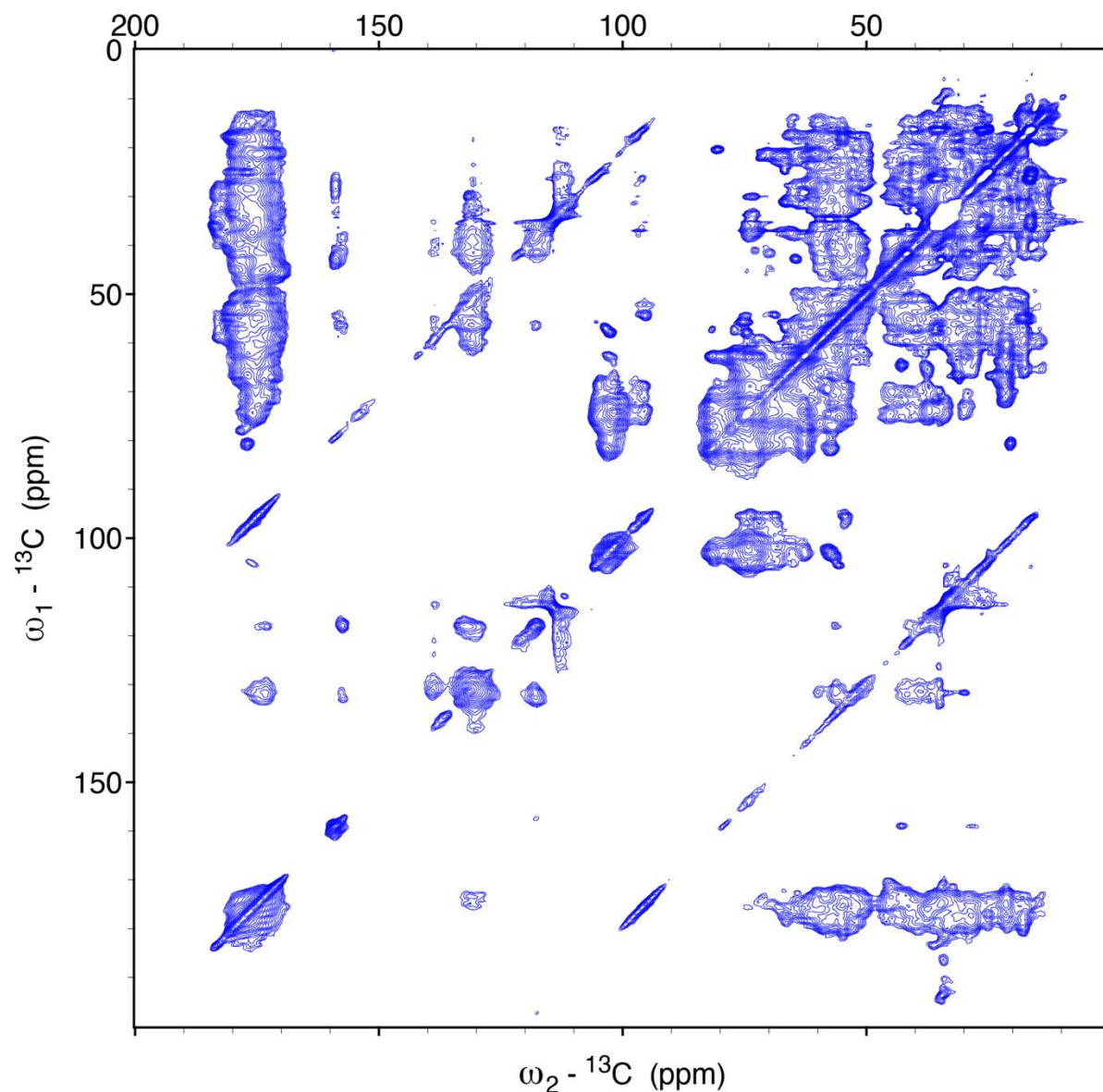


Figure C3. Homonuclear ^{13}C - ^{13}C correlation spectrum of Leu-Rev-HM using DARR with mixing time 100 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate and 128 scans. The spectrum was processed with the window function QSINE, SSB = 2 for both dimensions.

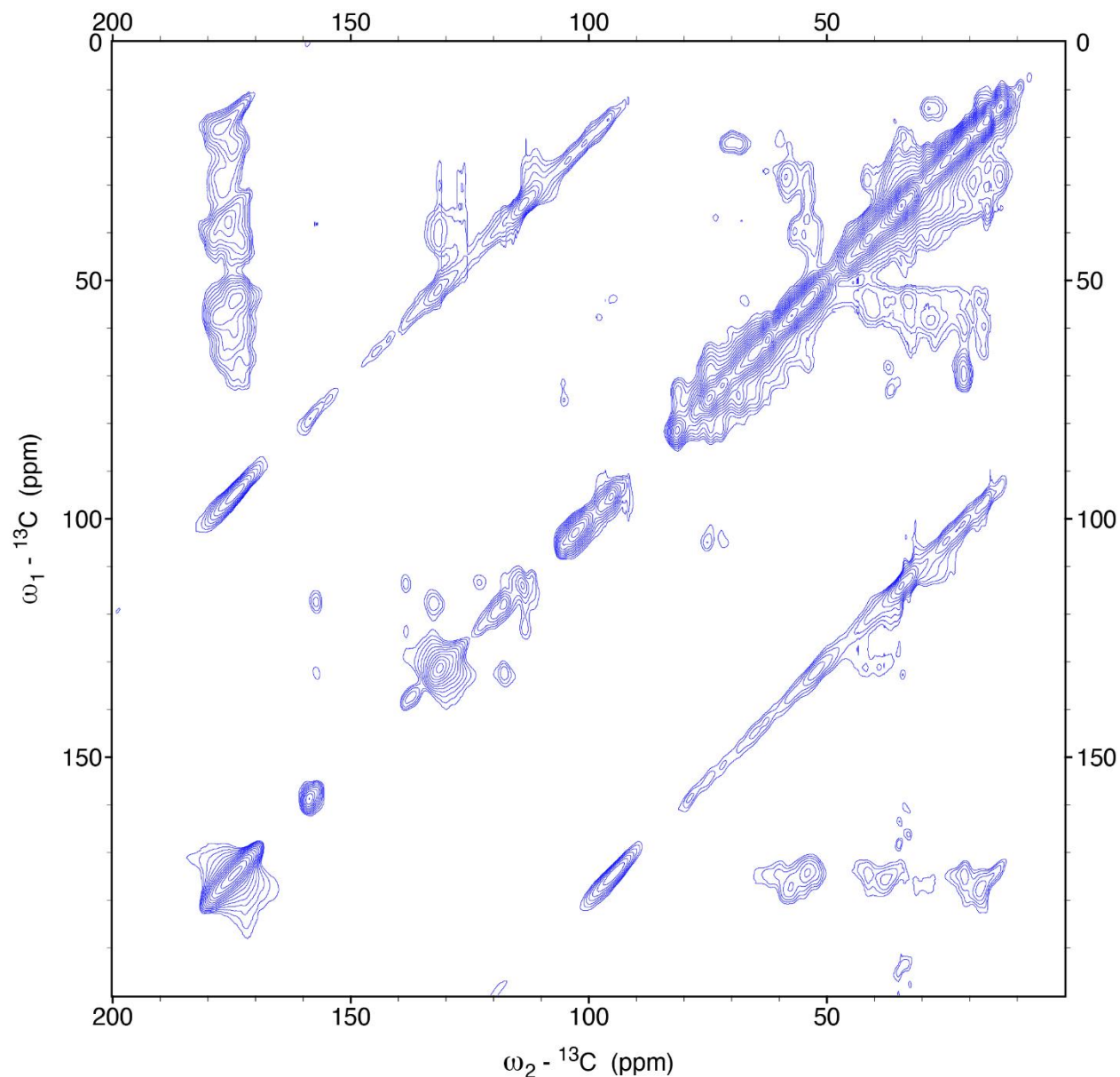


Figure C4. Homonuclear ^{13}C - ^{13}C correlation spectrum of 1,3- ^{13}C -Glyc-HM using DARR with mixing time 30 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate. 50% NUS, 256 scans with T2 relaxation time 0.0005s was used for acquisition. The spectrum was reconstructed by CS and then was processed with the window function QSINE, SSB = 2 for both dimensions.

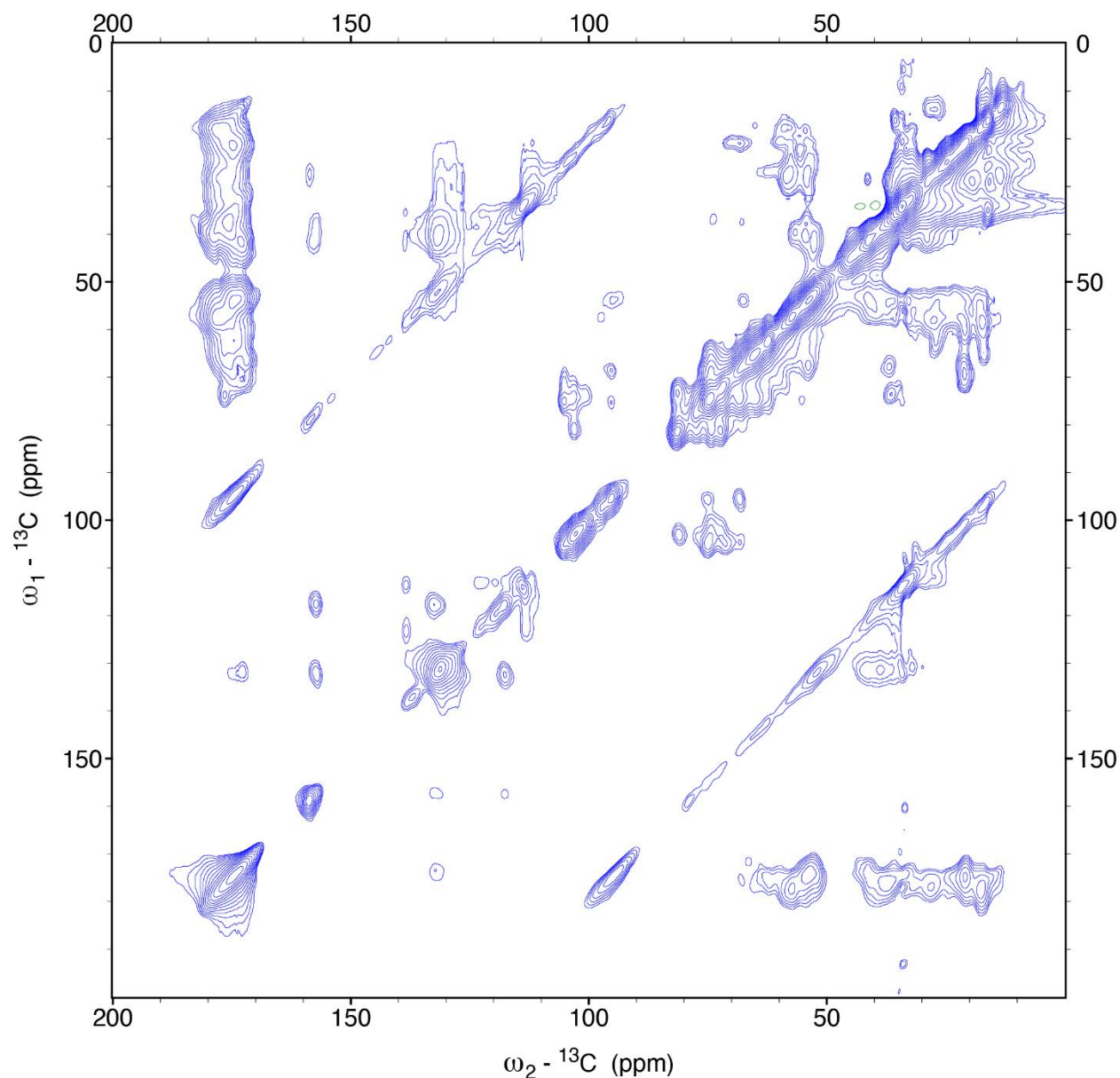


Figure C5. Homonuclear ^{13}C - ^{13}C correlation spectrum of 1,3- ^{13}C -Glyc-HM using DARR with mixing time 100 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate. 50% NUS, 256 scans with T2 relaxation time 0.0005s was used for acquisition. The spectrum was reconstructed by CS and then was processed with the window function QSINE, SSB = 2 for both dimensions.

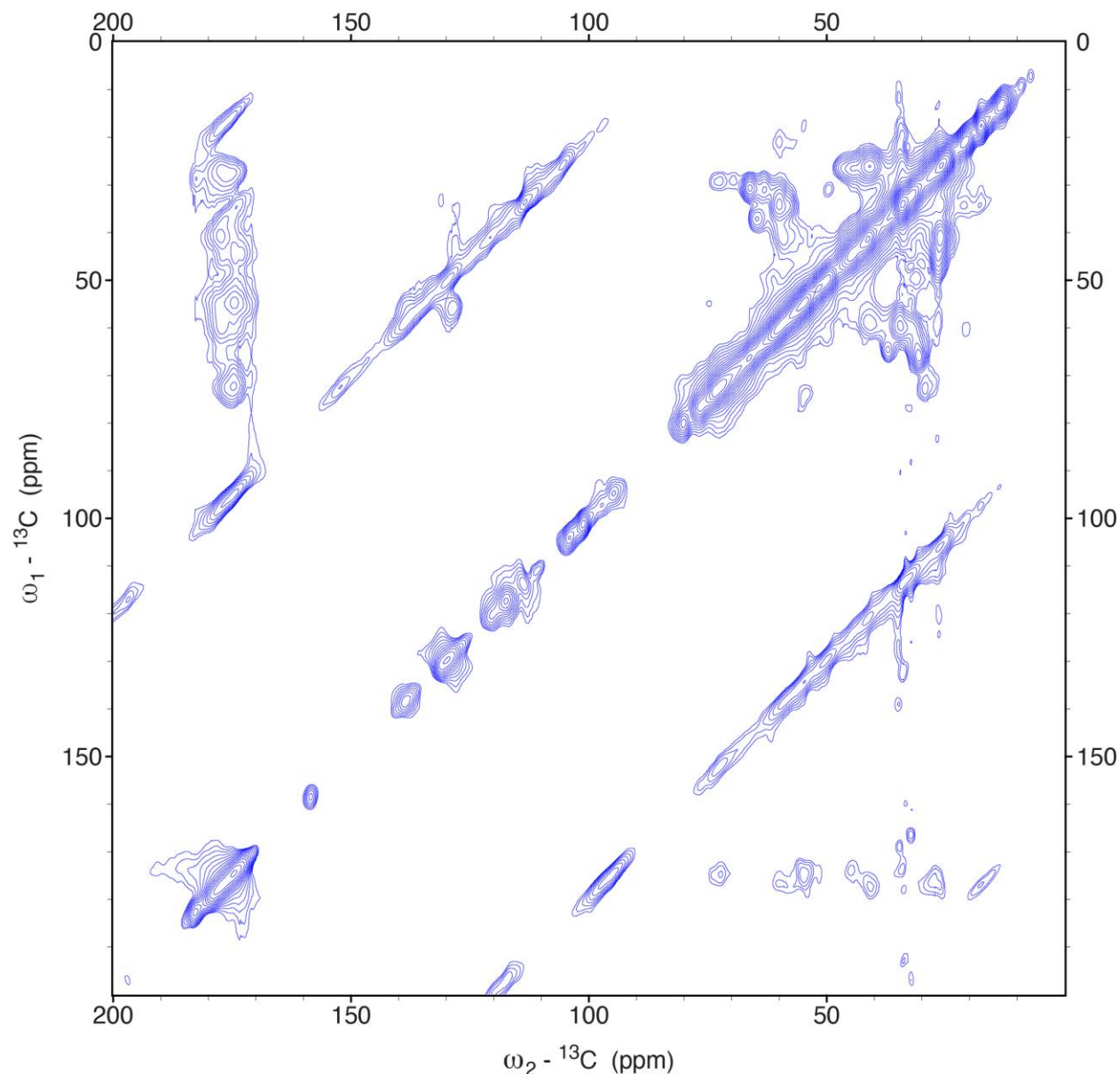


Figure C6. Homonuclear ^{13}C - ^{13}C correlation spectrum of 2- ^{13}C -Glyc-HM using DARR with mixing time 30 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate. 50% NUS, 256 scans with T2 relaxation time 0.0005s was used for acquisition. The spectrum was reconstructed by MDD and then was processed with the window function QSINE, SSB = 2 for both dimensions.

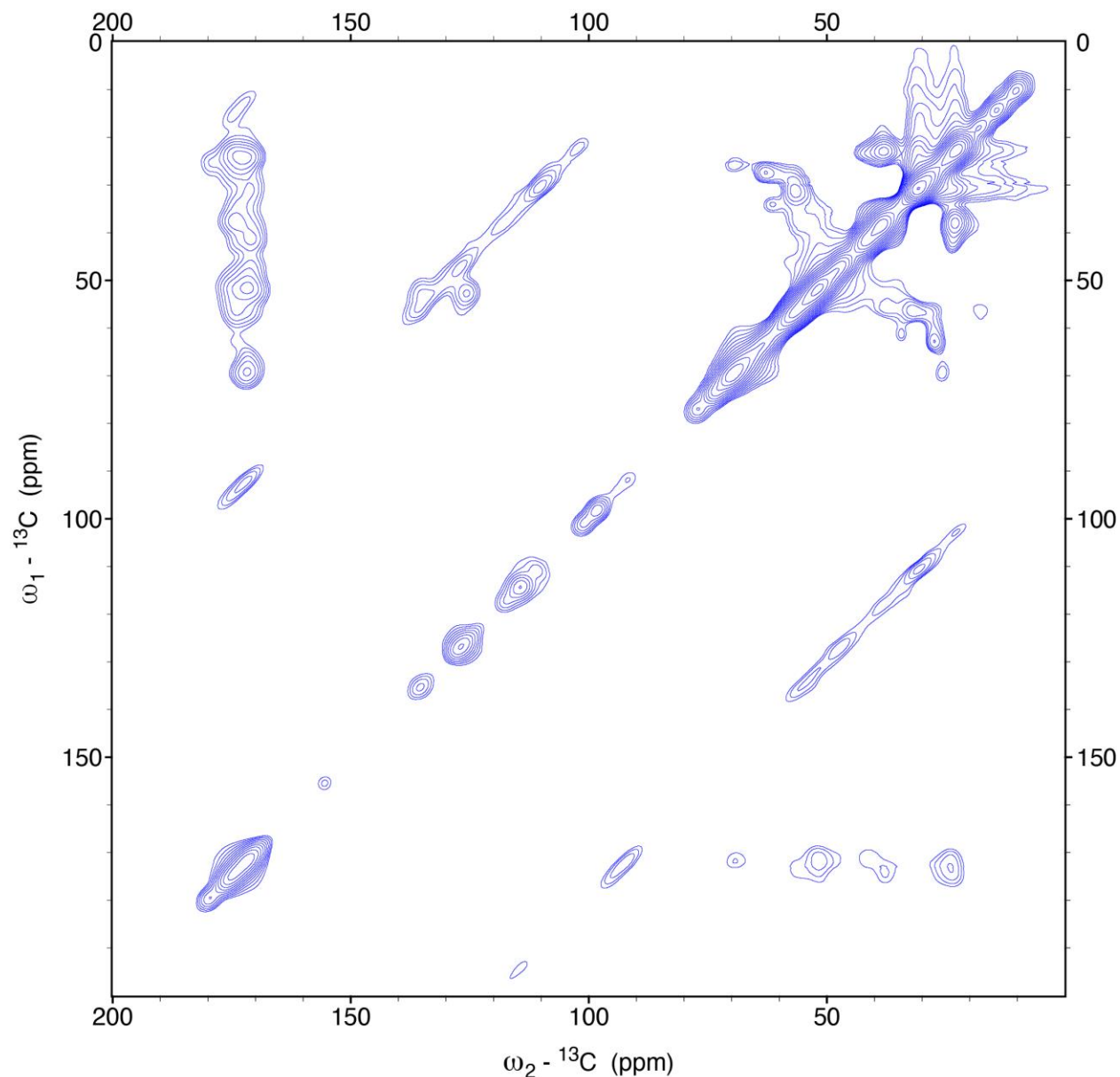


Figure C7. Homonuclear ^{13}C - ^{13}C correlation spectrum of 2- ^{13}C -Glyc-HM using DARR with mixing time 100 ms. Top: the spectrum at 0-200ppm range. Bottom: the expanded spectrum of the Ser peak region. The crosshairs peak was chosen to display the 1D slice on the bottom and the right side of one the of Ser CA-CB peaks. Parameters for data acquisition: 100 kHz for ^1H excitation pulse, $^1\text{H} \rightarrow ^{13}\text{C}$ CP contact time 0.5 ms with 40 kHz on ^{13}C and 60-72 kHz linear CP ramp on ^1H ; DARR mixing: 97 kHz for ^{13}C , mixing time 30 ms with 1.6 kHz recoupling field applied on ^1H . The decoupling field during evolution period and data acquisition was about 81 kHz. The number of points are 600 for direct ^{13}C dimension and 480 for indirect ^{13}C dimension. The increment for delay in the indirect dimension is 16.5625 μs . Both dimensions have spectral width as of 300 ppm and only 0-200 ppm is shown. The data was collected at 253 K with 16 kHz spinning rate. 50% NUS, 256 scans with T2 relaxation time 0.0005s was used for acquisition. The spectrum was reconstructed by MDD and then was processed with the window function QSINE, SSB = 2 for both dimensions.

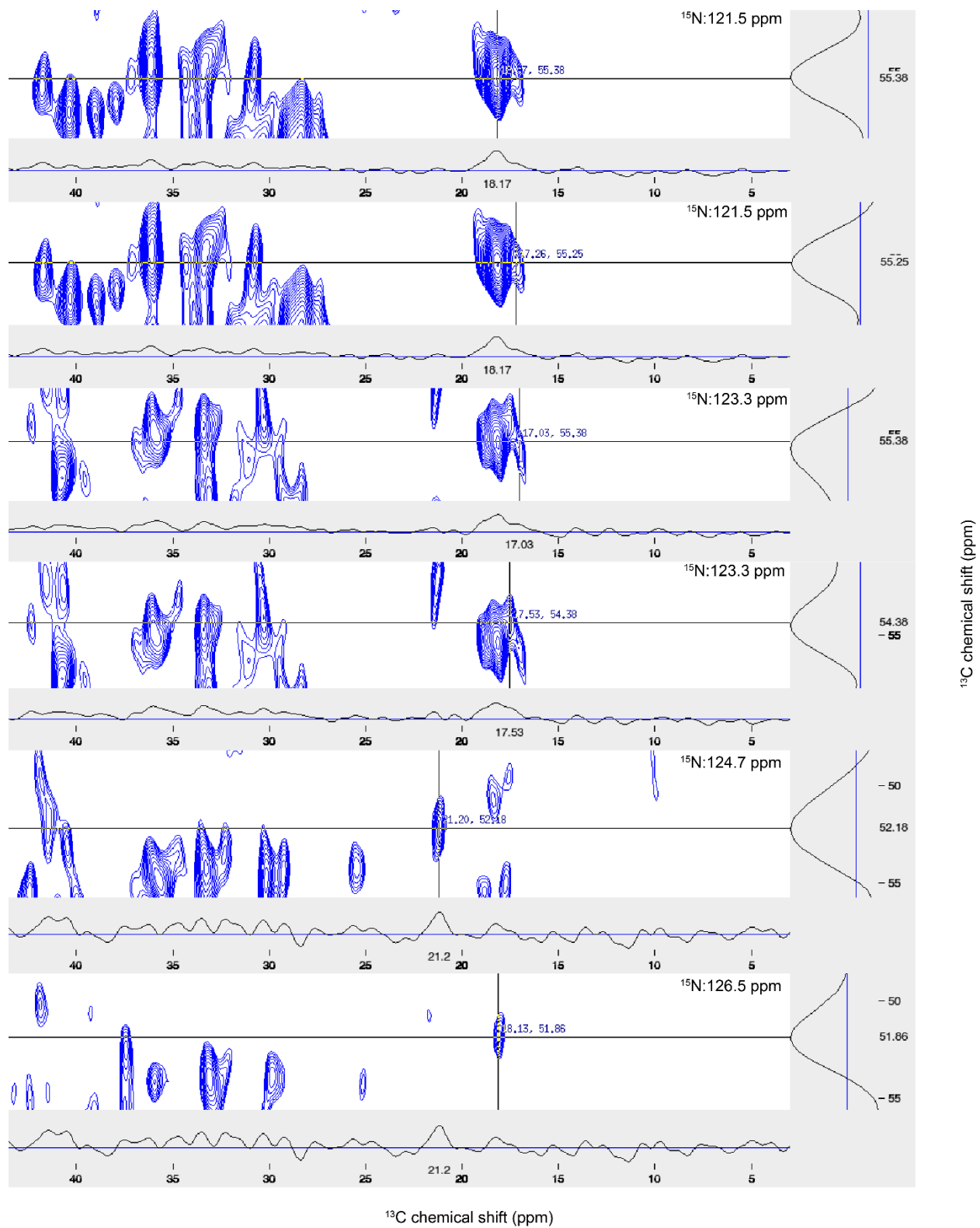


Figure C8. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to AlaN-CA-CB peak listed in Table 5.1.

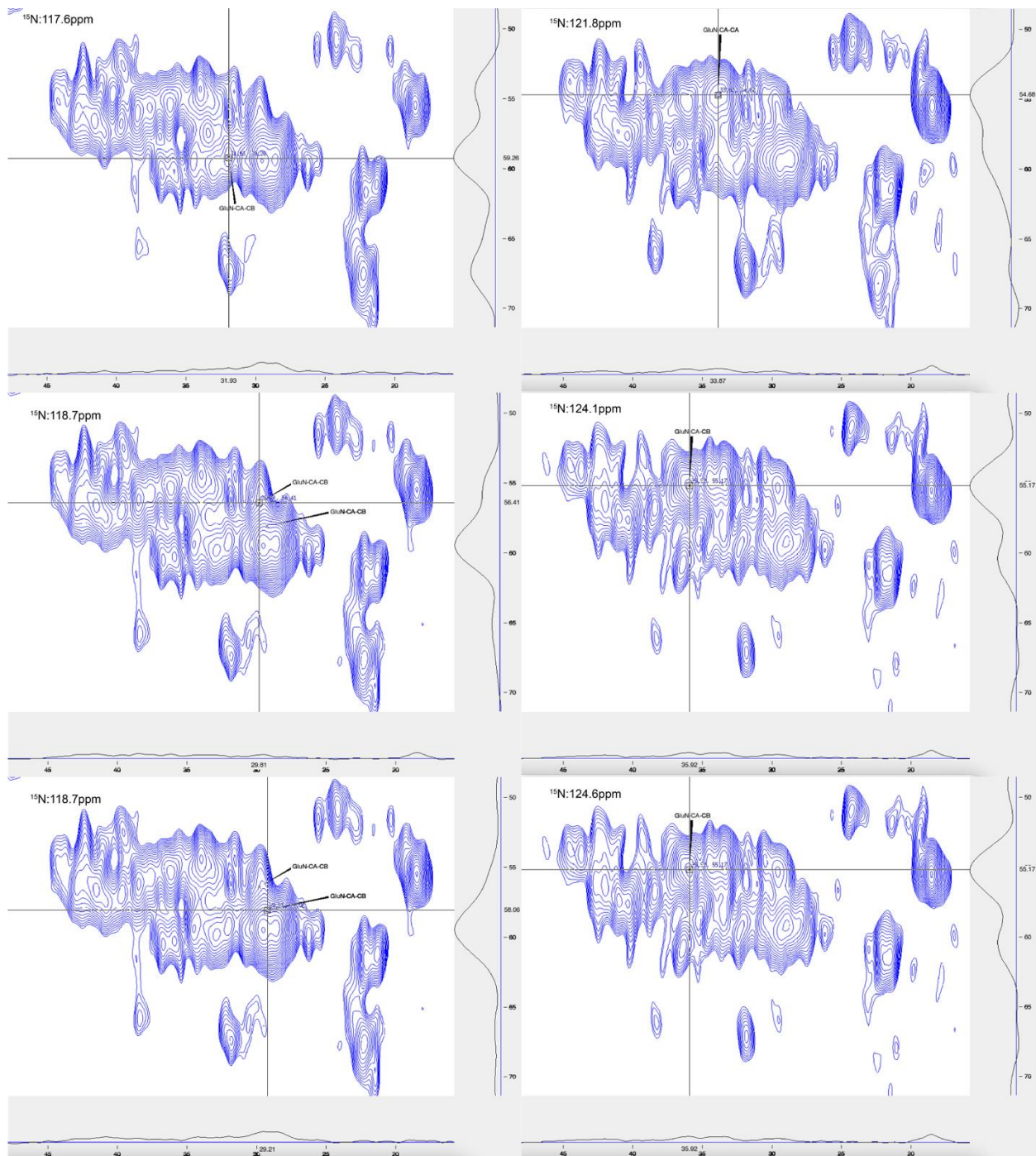


Figure C9. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to GluN-CA-CB peak listed in Table 5.1.

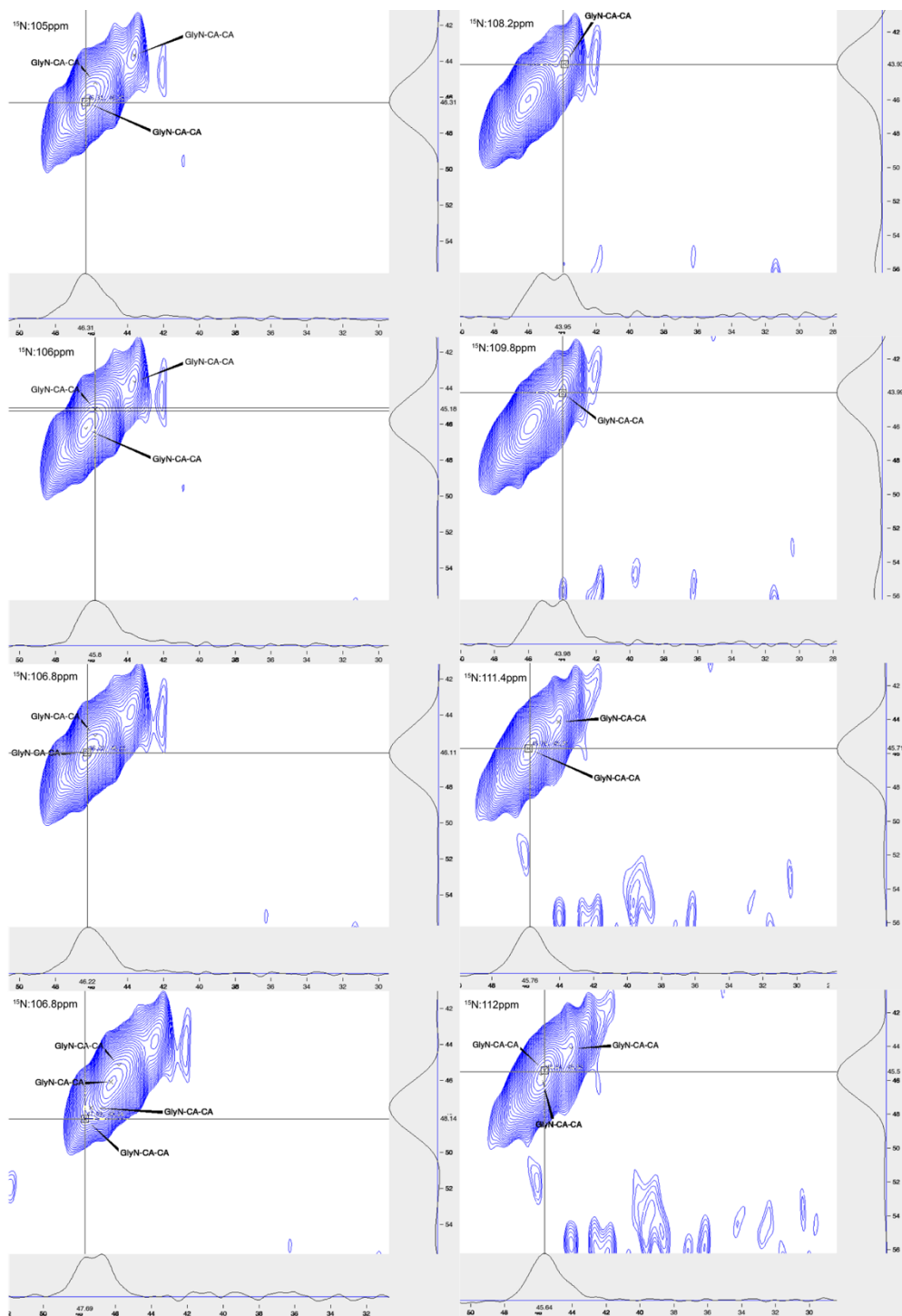


Figure C10. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to GlyN-CA-CB peak listed in Table 5.1.

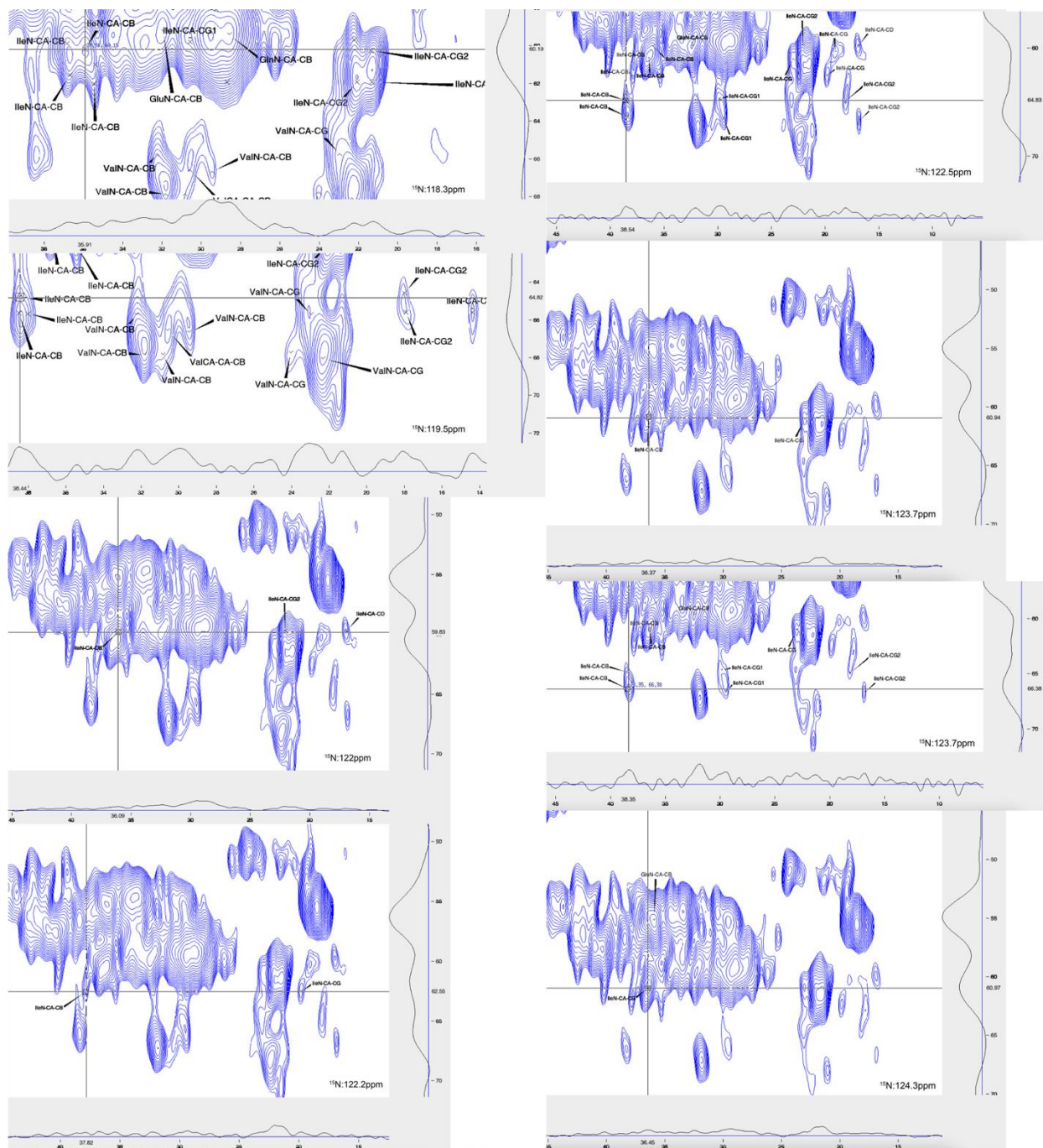


Figure C11. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to IleN-CA-CB peak listed in Table 5.1.

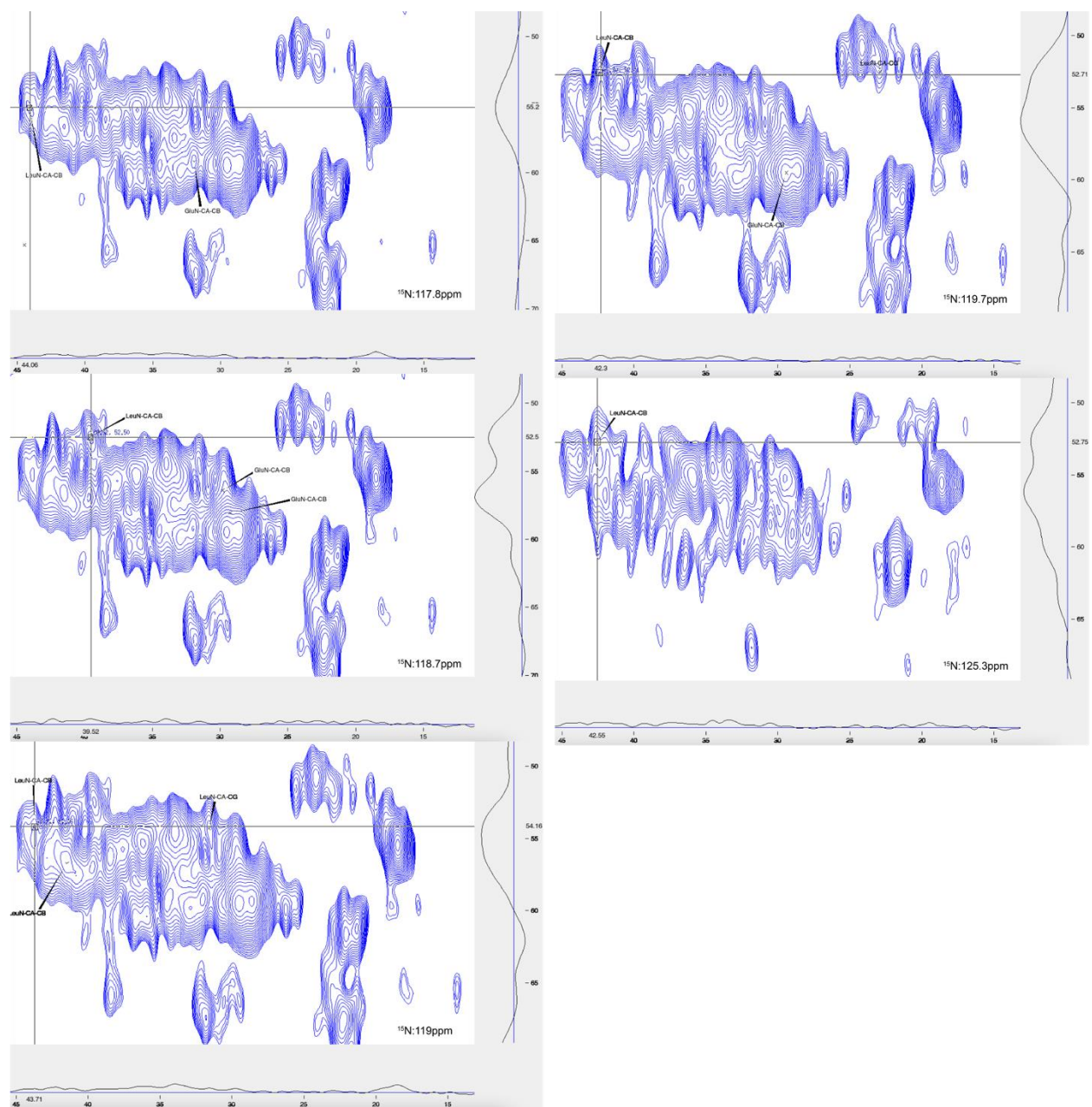


Figure C12. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to LeuN-CA-CB peak listed in Table 5.1.

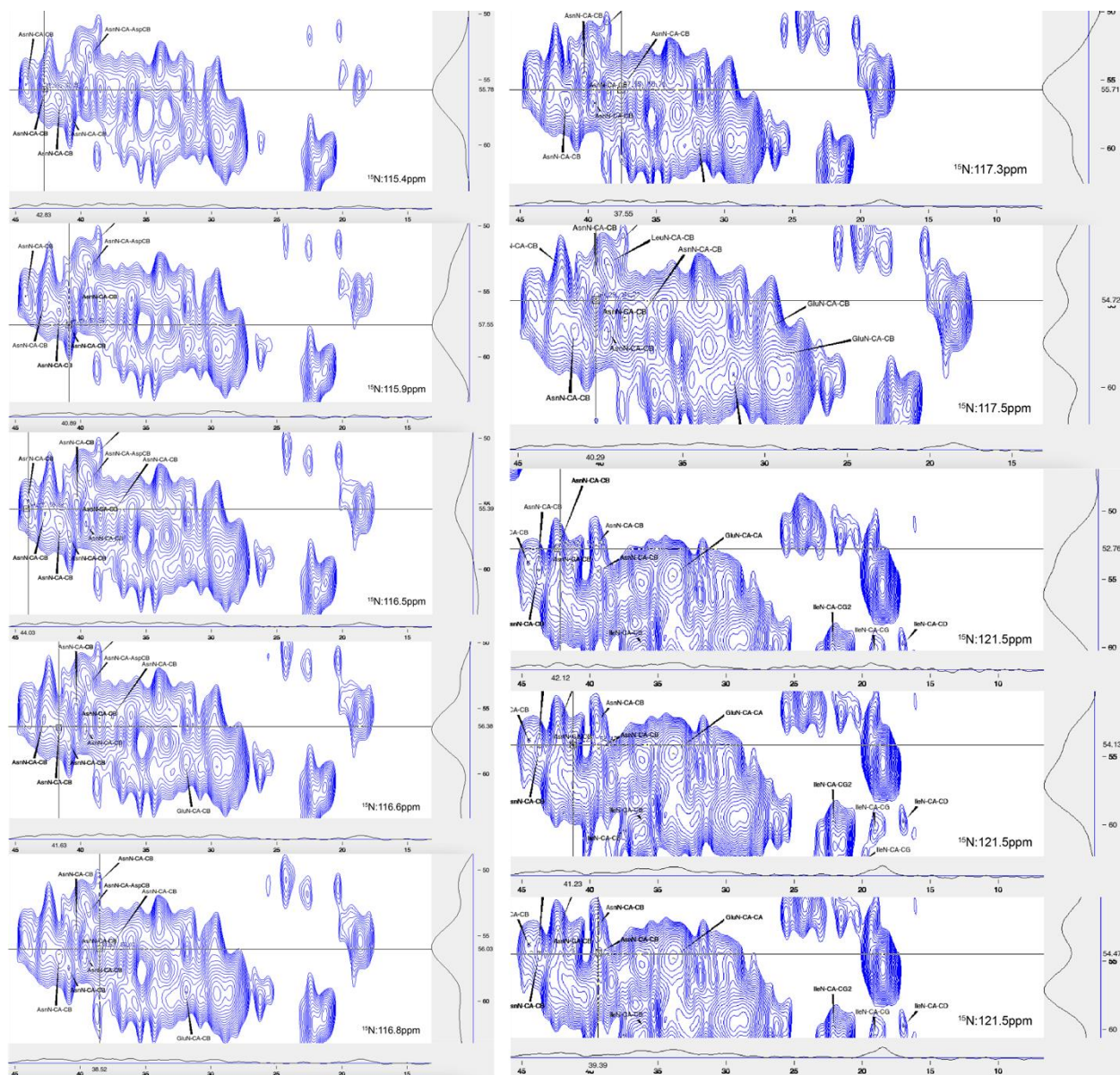
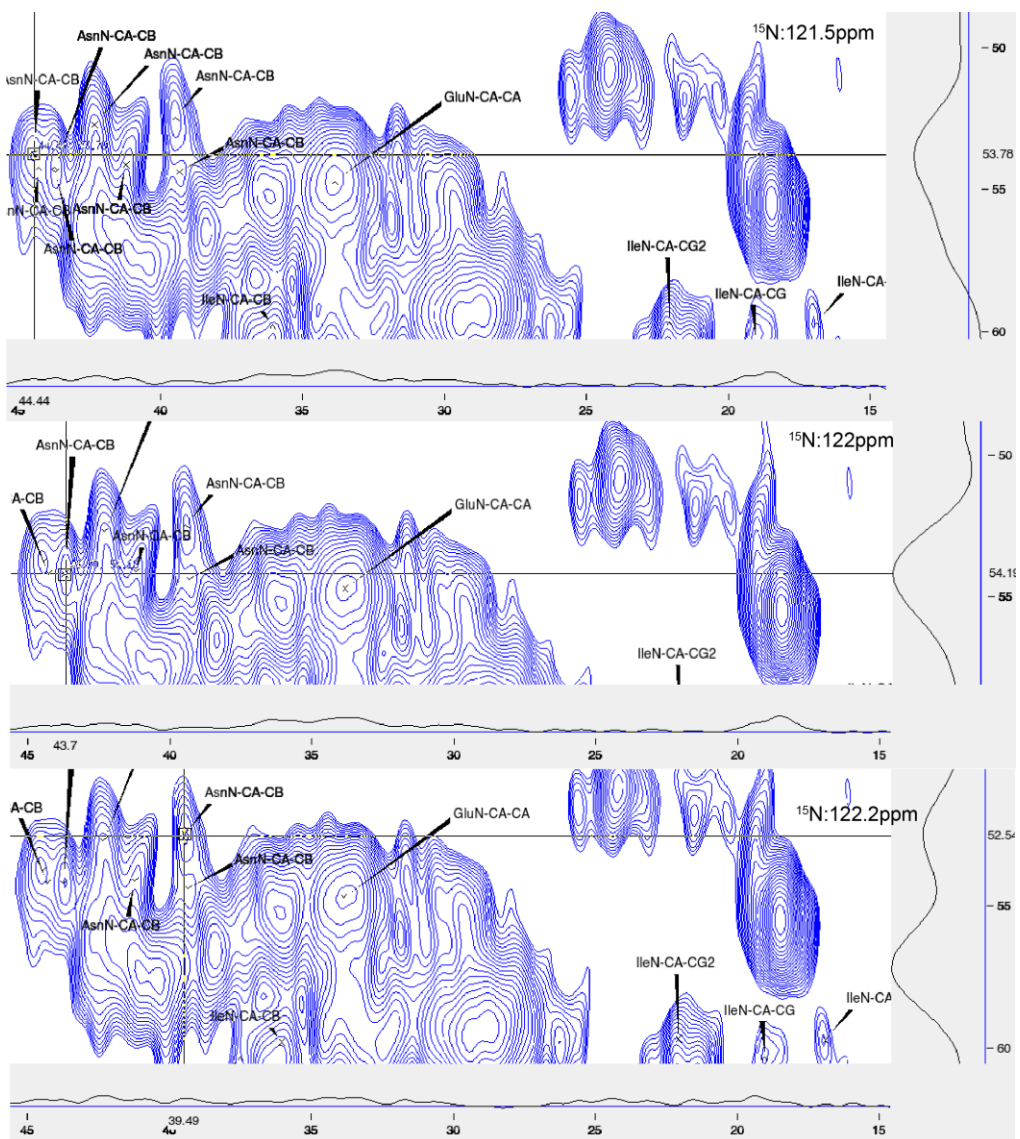


Figure C13. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to AsnN-CA-CB peak listed in Table 5.1.

Figure C13 (cont'd)



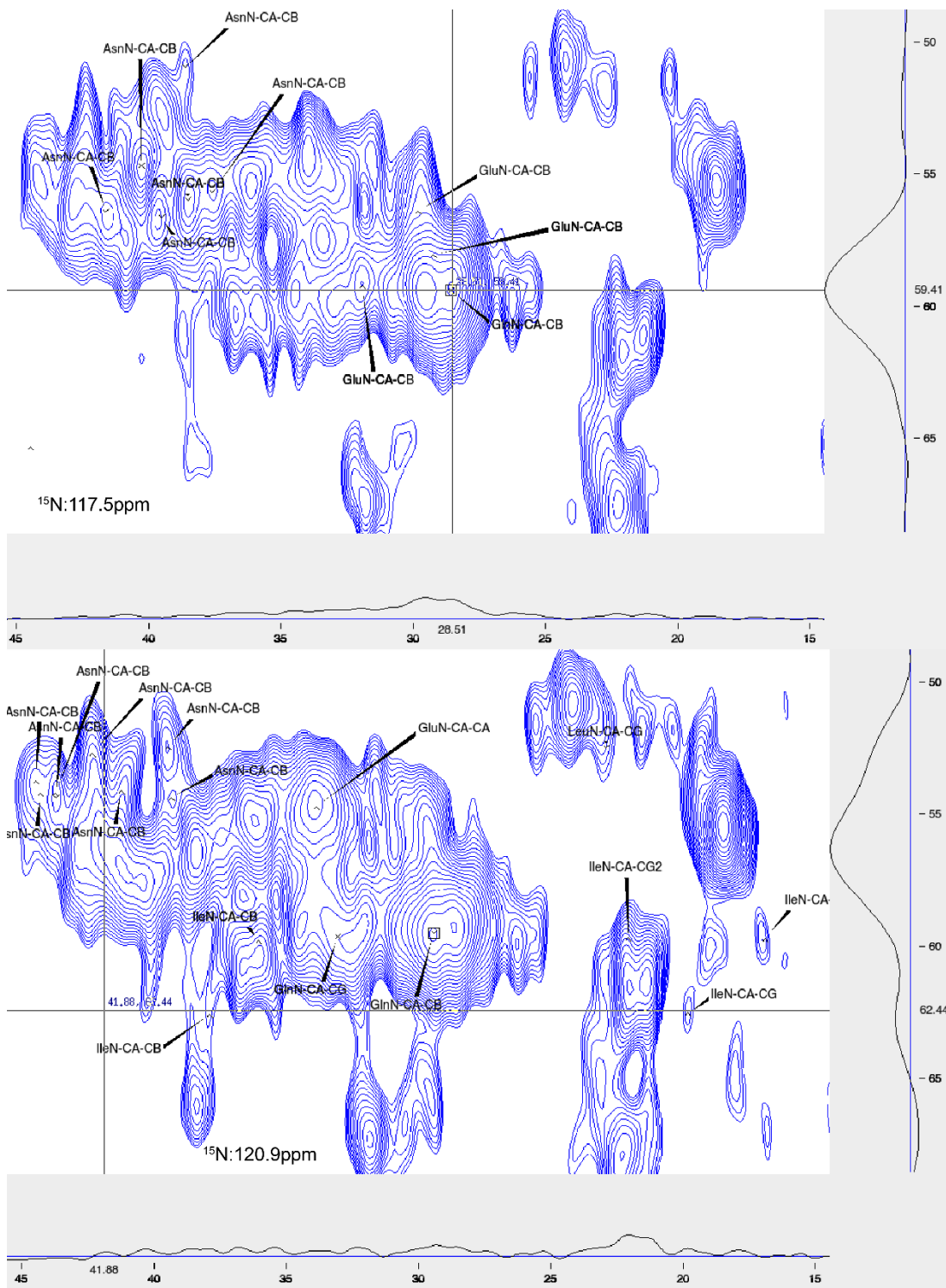


Figure C14. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to GlnN-CA-CB peak listed in Table 5.1.

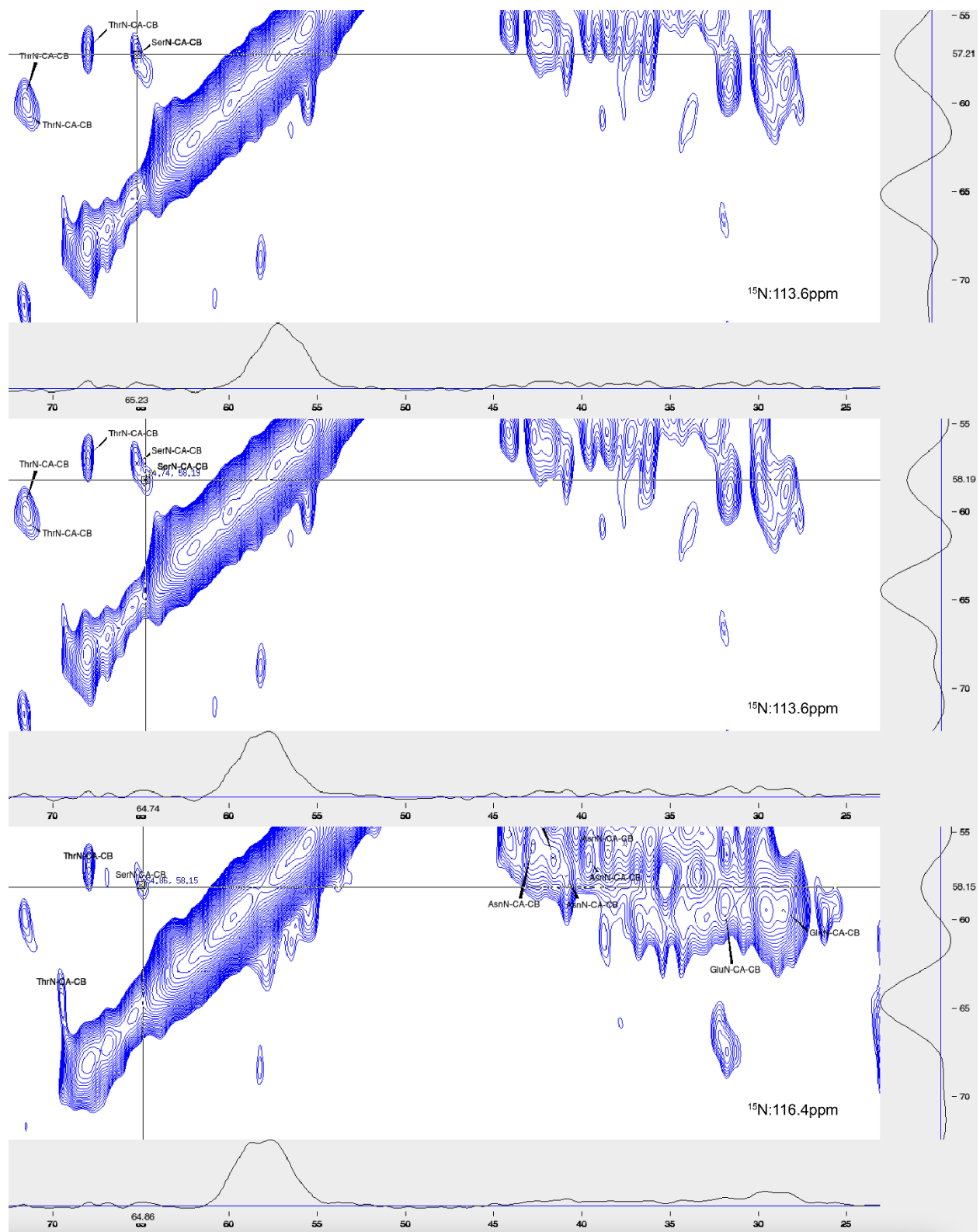


Figure C15. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to SerN-CA-CB peak listed in Table C1.

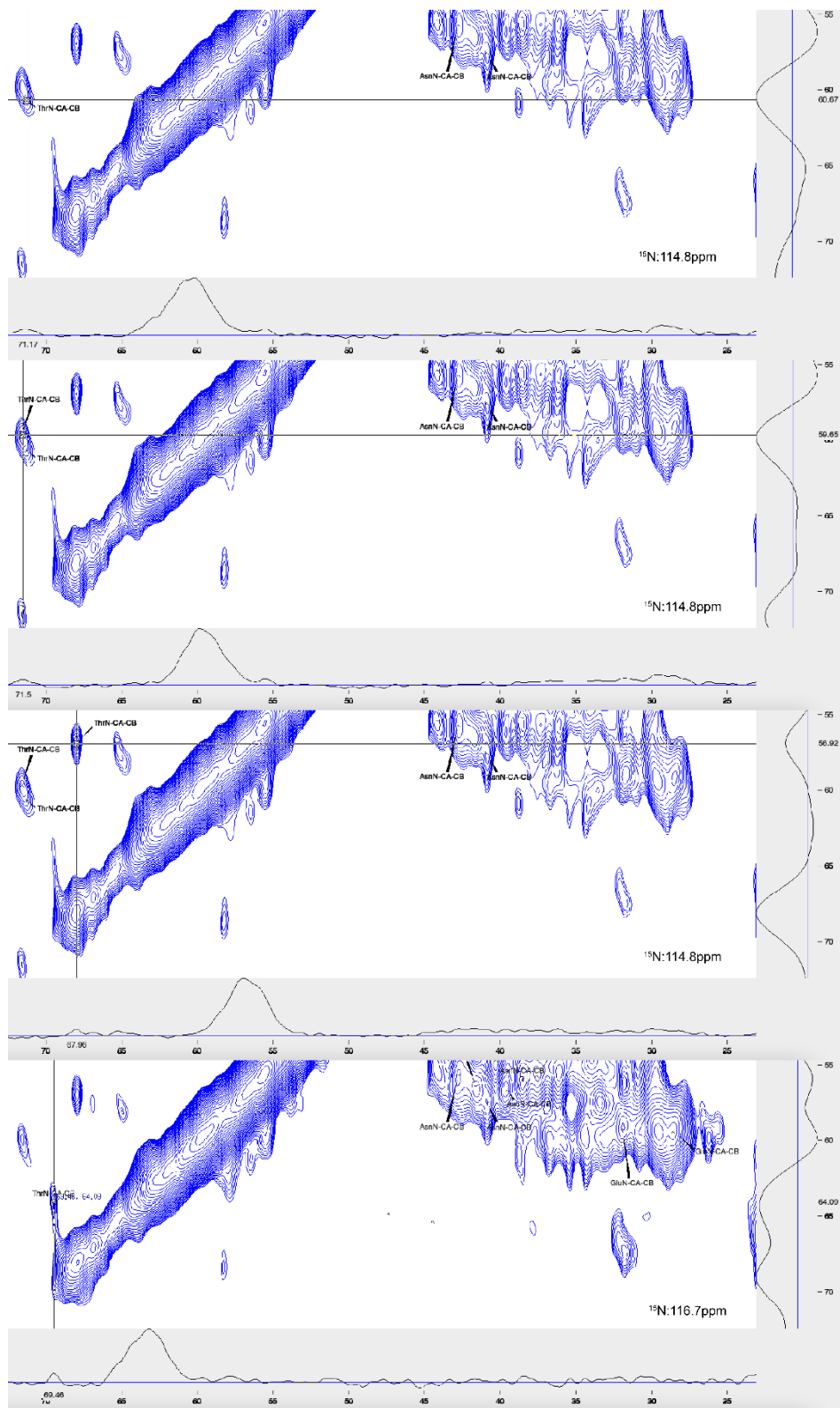


Figure C16. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to ThrN-CA-CB peak listed in Table 5.1.

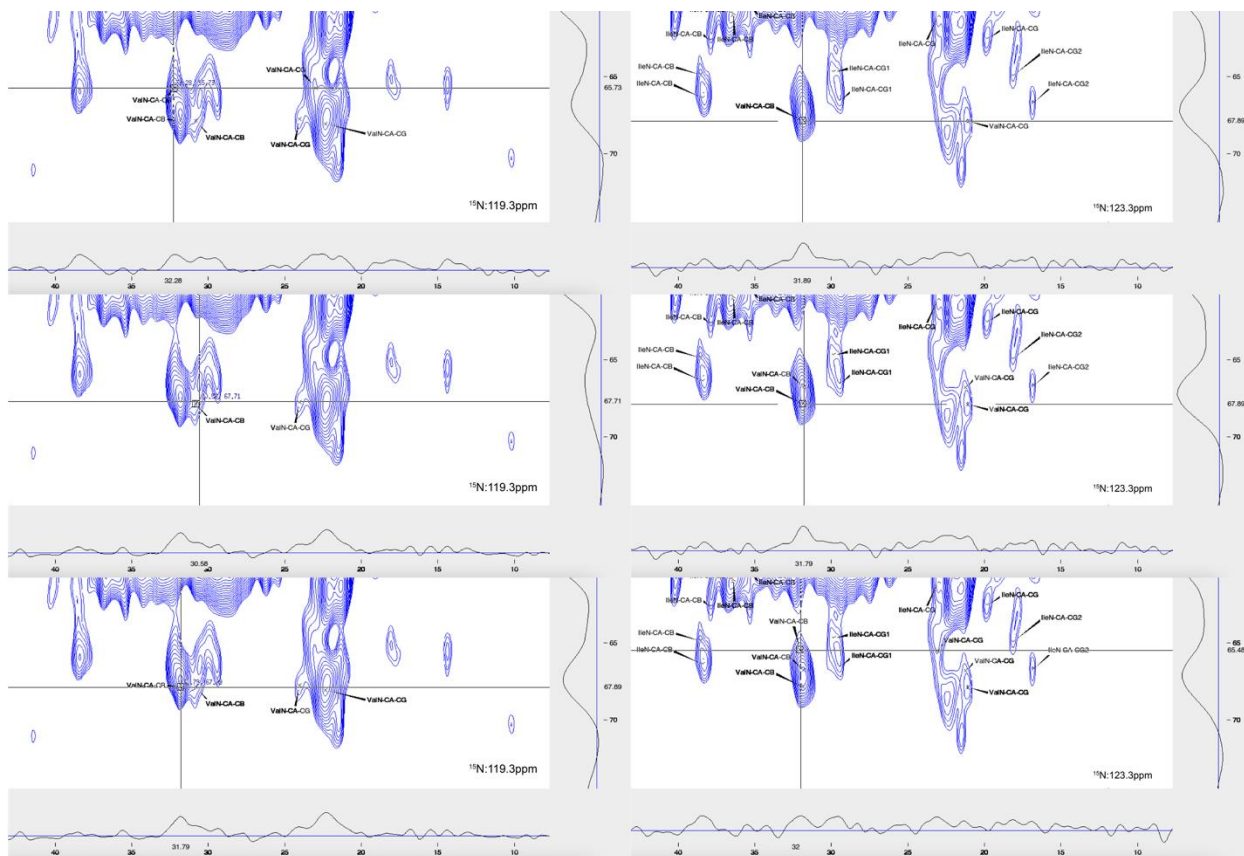


Figure C17. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to ValN-CA-CB peak listed in Table 5.1.

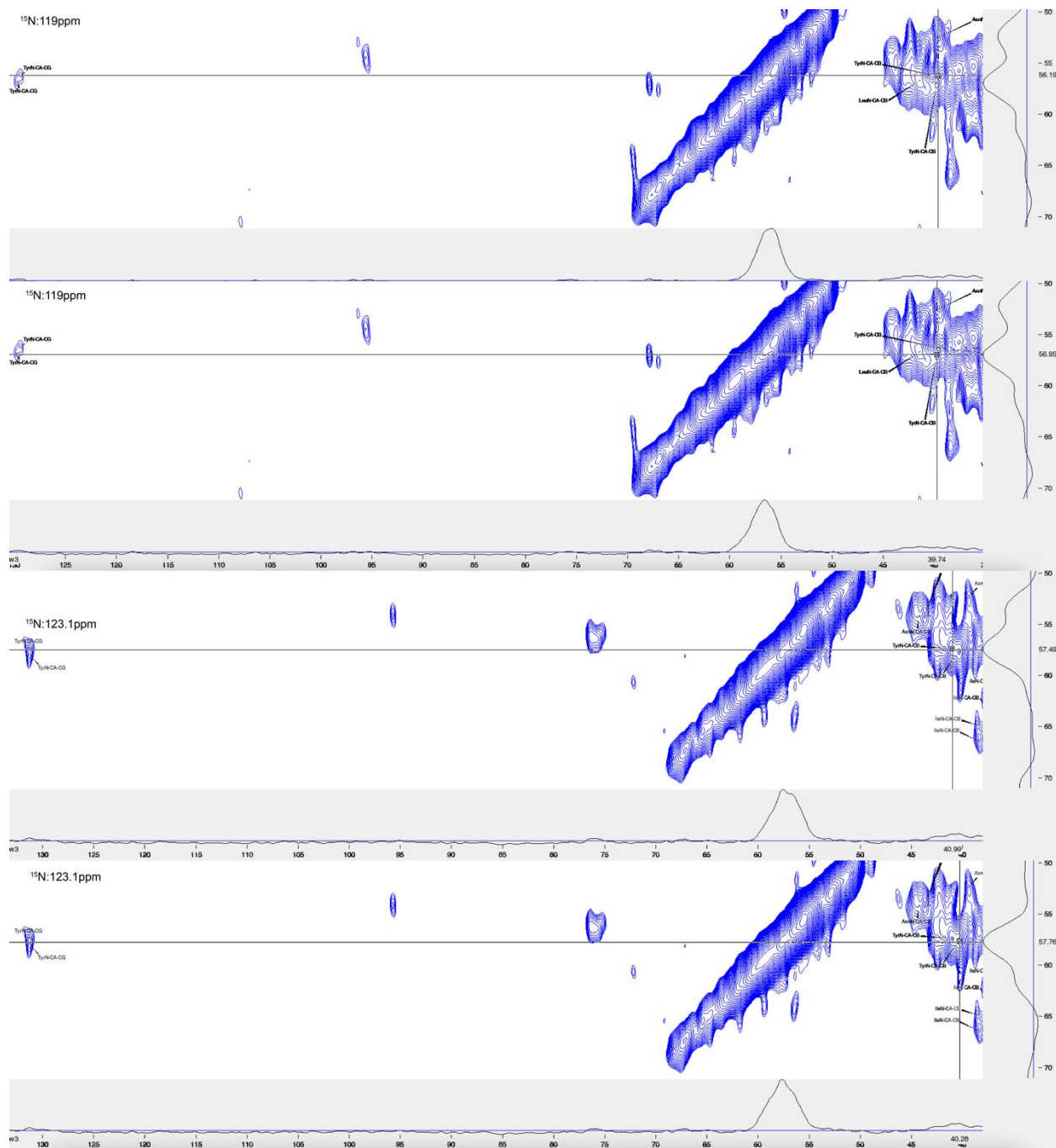
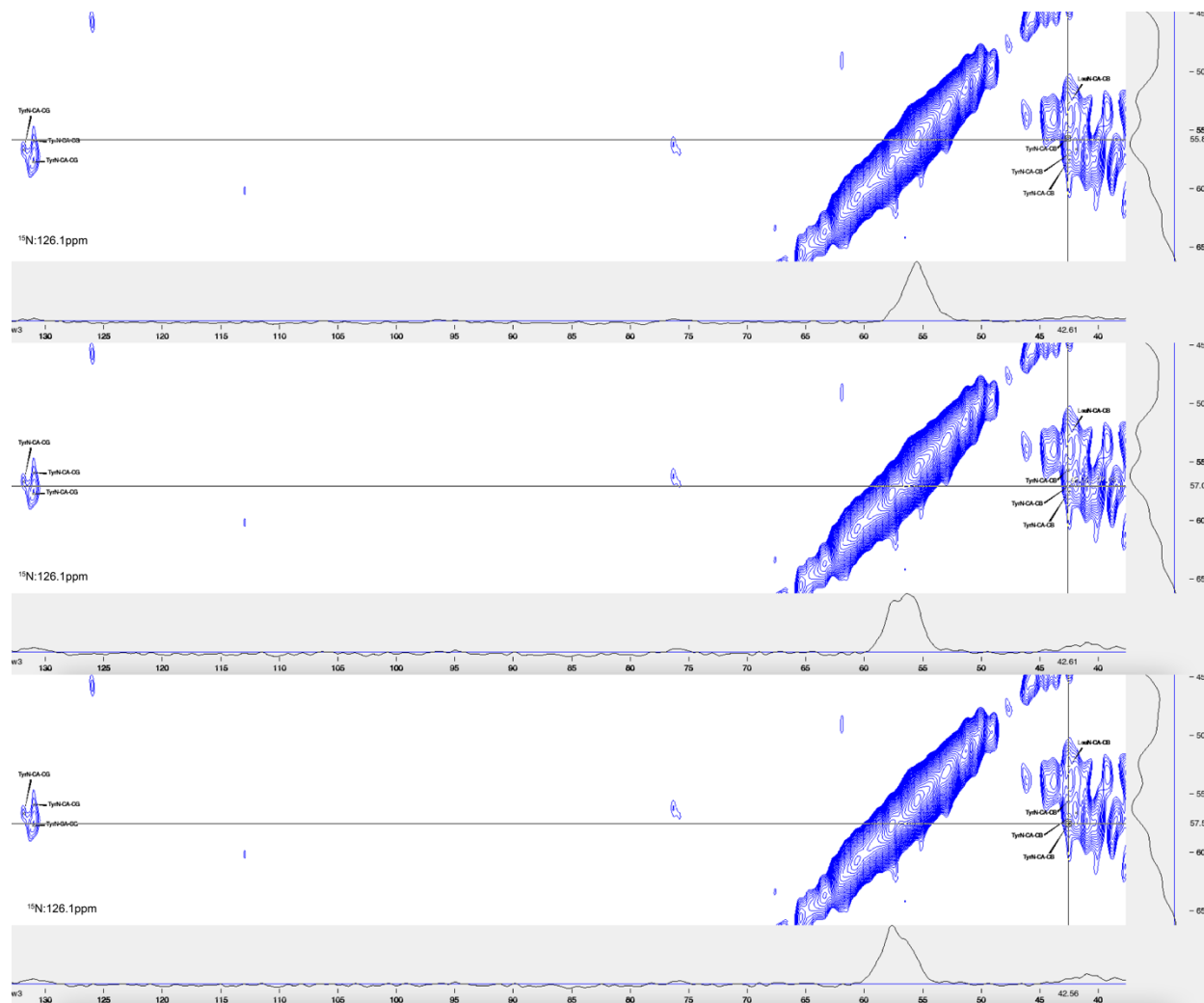


Figure C18. NCACX spectra of Leu-Rev-HM. The peak with crosshair corresponds to TyrN-CA-CB peak listed in Table 5.1.

Figure C18 (cont'd)



The presence of amyloid/ β sheet aggregates in the protein can be monitored by the strong fluorescence signal coming from thioflavin T (ThT) or thioflavin S (ThS) upon binding to amyloids. ThS has weak emission at 510 nm (excitation 450 nm) without binding to amyloid β sheet structure and the fluorescence would be significantly enhanced when amyloid protein present. The ThS experiments was carried out for HM proteins separated with either PBS buffer or PBS+wash buffer, denoted as PBS-IB and PW-IB in the following description.

Three samples were prepared with either PBS-IB, PW-IB, or p-tau (hyperphosphorylated tau protein), and [protein] = 30 μ M in each sample. The samples were all suspensions/solutions formed by light vortexing in either PBS (PBS-IB and PW-IB) or 20 mM Tris buffer at pH 7.4 with 100 mM NaCl (p-tau). The three samples were incubated overnight at 37 °C. Three fresh samples were also prepared the following morning. Each of the six samples was mixed in a well with an aliquot of ThS stock in Tris buffer so that final [protein] = 15 μ M and [ThS] = 20 μ M. The plate was placed in a SpectraMax M2 plate reader which measured ThS fluorescence at 510 nm with excitation at 450 nm. Fluorescence was measured in each well every 10 minutes for 220 minutes. Fluorescence data is shown in Table C2 and plot of fluorescence signal vs. time is displayed in Figure C19. The strong fluorescence of all three samples supports that they all contain β sheet structure before and after overnight incubation. The incubation increases fluorescence of PW-IB but weakens the signal of PBS-IB.

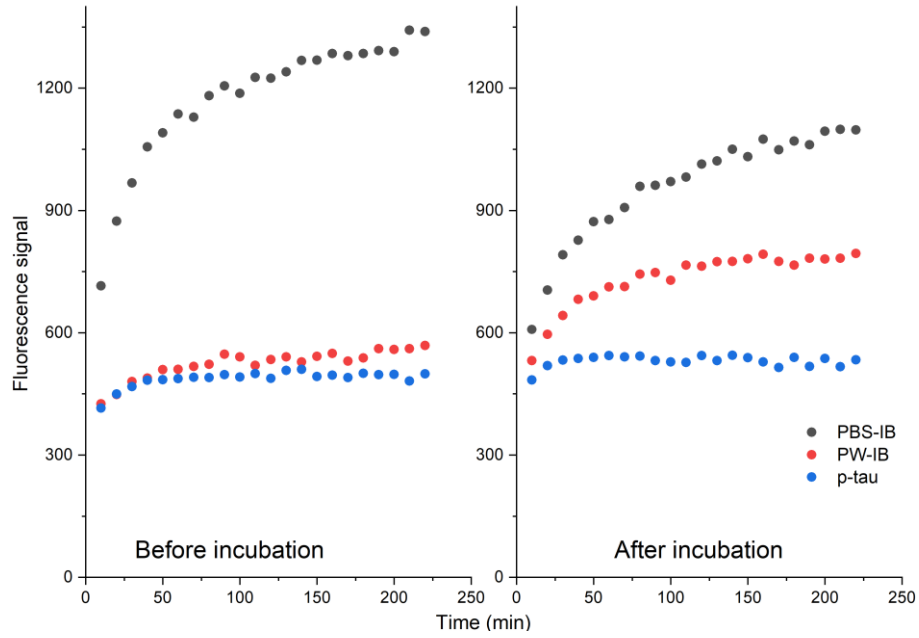


Figure C19. Fluorescence spectra for the ThS experiments of PBS-IB, PW-IB, and p-tau protein before and after incubation indicating the presence of β sheet structure.

Table C2. Fluorescence data for the ThS experiments of PBS-IB, PW-IB, and p-tau protein before and after overnight incubation at 37°C.

Time (min)	Before incubation			After incubation		
	PBS-IB	PW-IB	p-tau	PBS-IB	PW-IB	p-tau
10	715.33	425.33	415.00	608.00	531.33	484.00
20	874.00	448.00	449.33	705.00	596.00	519.00
30	967.67	480.33	468.00	791.33	642.00	532.67
40	1056.00	488.67	483.33	826.67	681.67	536.33
50	1090.00	509.33	484.33	872.33	690.67	539.00
60	1136.67	509.67	487.33	877.67	712.33	543.67
70	1128.67	517.00	490.33	906.67	713.33	540.33
80	1181.67	522.33	490.00	959.00	743.67	542.67
90	1205.67	547.00	497.00	961.33	747.33	531.67
100	1187.33	540.33	491.33	970.67	729.00	528.00
110	1226.00	519.67	499.33	981.67	765.67	527.00
120	1224.00	534.00	487.67	1013.67	763.33	544.00
130	1240.00	540.67	507.33	1021.33	774.00	531.67
140	1268.00	528.33	510.00	1050.00	775.00	544.33
150	1268.33	542.00	492.33	1032.00	781.67	538.67
160	1285.00	548.67	495.33	1075.00	792.67	528.00
170	1279.33	530.33	489.67	1049.00	775.00	514.67
180	1285.00	537.67	500.33	1070.33	766.00	539.33
190	1291.67	561.00	497.00	1061.00	783.00	517.00
200	1289.33	558.67	497.67	1094.00	781.00	536.67
210	1342.00	560.67	481.33	1099.00	782.67	516.67
220	1338.67	568.67	499.00	1097.33	794.33	533.33

APPENDIX D SUPPORTING INFORMATION FOR REDOR DATA ANALYSIS

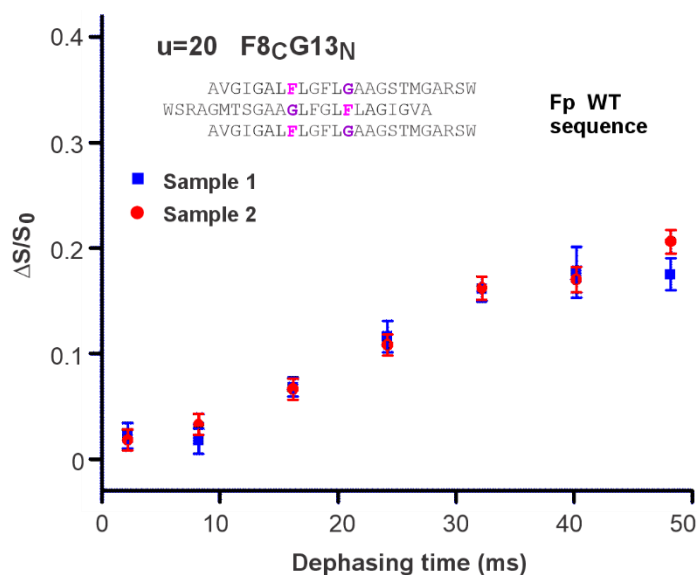


Figure D1. REDOR NMR $\Delta S/S_0$ vs. dephasing time (τ_k) for two separately prepared and measured WT F8C13N, u=20 samples^a.

^a Each sample was prepared from separately synthesized and purified batches of peptide and separately-prepared batches of membrane vesicles. The two samples exhibit similar $\Delta S/S_0$ values within uncertainties, which supports the reproducibility of the sample preparation and NMR measurements. The numerical values of $\Delta S/S_0$ and uncertainties for these and other replicate samples are presented in Table D2.

Table D1. Experimental REDOR $\Delta S/S_0$ and uncertainties for replicate sample with WT Fp's^a.

τ_k (ms)	u = 13		u = 16		u = 17		u = 20		
	Fp	Fp-dimer	Fp	Fp-dimer	Fp	Fp-dimer	Fp	Fp	Fp-dimer
2.2	0.010(5)	0.012(5)	0.012(7)	0.016(8)	0.004(9)	0.005(6)	0.022(12)	0.018(10)	0.024(8)
8.2	0.034(8)	0.039(7)	0.044(9)	0.049(7)	0.058(10)	0.047(7)	0.017(12)	0.033(10)	0.040(8)
16.2	0.067(12)	0.086(8)	0.090(10)	0.085(8)	0.099(7)	0.097(8)	0.068(9)	0.066(10)	0.066(11)
24.2	0.102(15)	0.125(8)	0.128(8)	0.138(8)	0.155(13)	0.149(11)	0.116(15)	0.108(10)	0.113(18)
32.2	0.172(14)	0.170(9)	0.179(11)	0.178(14)	0.192(11)	0.211(18)	0.161(12)	0.162(11)	0.151(13)
40.2	0.218(16)	0.212(15)	0.238(15)	0.232(13)	0.247(16)	0.216(13)	0.177(24)	0.170(12)	0.157(15)
48.2	0.256(25)	0.236(15)	0.253(15)	0.277(18)	0.275(21)	0.253(13)	0.175(15)	0.206(11)	0.201(20)

^a For each replicate sample, the Fp was separately -synthesized and -purified but had the same ¹³CO-labeled residue and the same ¹⁵N-labeled residue, Figure 6. 5. The Fp-dimer was synthesized by cross-linking in air AVGIGALFLGFLGAAGSTMGARSWKKKKKCA, with underlining for the 23 N-terminal residues of HIV gp41. The synthesis and purification procedures for Fp-dimer are described in *Biochemistry* (2009) 48, 289-301. The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. an entry of 0.058(10) means 0.058 ± 0.010 .

Table D2. SIMPSON-calculated values of $\gamma_t(\tau_k)^{lb,lb}$ used in constrained fittings, Eq.6.12, and values of $\gamma_{t_1,t_2}(\tau_k)^{lb,lb}$ used in unconstrained fittings, Eq. 6.18. The first row in the table is the t values for constrained fittings and the second row is the t_1 and t_2 values for unconstrained fittings.

	$t = u$	$t = u \pm 1$	$t = u \pm 2$							
	$t_1=u$	$t_1 = u \pm 1$	$t_1=u$	$t_1=u$	$t_1 = u \pm 1$	$t_1=X$	$t_1 = u \pm 1$	$t_1 = u \pm 1$	$t_1 = X$	
	$t_2=u$	$t_2 = u \pm 1$	$t_2 = u \pm 1$	$t_2=X$	$t_2=u$	$t_2=u$	$t_2 = u -1$	$t_2 = X$	$t_2 = u \pm 1$	
τ_k										
(ms)										
2.2	0.9917	0.9984	0.9998	0.9921	0.9928	0.9980	0.9989	0.9984	0.9991	0.9992
8.2	0.8938	0.9785	0.9971	0.8974	0.9064	0.9737	0.9850	0.9785	0.9885	0.9898
16.2	0.6453	0.9186	0.9890	0.6476	0.6710	0.9012	0.9427	0.9191	0.9560	0.9608
24.2	0.3786	0.8263	0.9755	0.3565	0.3778	0.7914	0.8754	0.8293	0.9039	0.9143
32.2	0.1964	0.7103	0.9569	0.1288	0.1236	0.6571	0.7870	0.7200	0.8348	0.8521
40.2	0.1186	0.5814	0.9334	0.0156	-0.0230	0.5132	0.6824	0.6037	0.7520	0.7769
48.2	0.0939	0.4501	0.9052	0.0011	-0.0470	0.3739	0.5675	0.4926	0.6592	0.6917

Table D3. WT $\Delta S/S_0$ values, $u=8-24$, from experiment and from unconstrained and constrained fittings with $b = 0.98^a$.

τ_k	u = 8			u = 9			u = 10			u = 11			u = 12		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.006(7)	0.0127	0.0126	0.012(6)	0.0128	0.013	0.015(10)	0.0134	0.014	0.014(5)	0.0143	0.0163	0.011(9)	0.0164	0.0174
8.2	0.017(6)	0.0247	0.0244	0.009(6)	0.0253	0.025	0.022(7)	0.026	0.0263	0.026(6)	0.0295	0.031	0.016(9)	0.0349	0.0364
16.2	0.030(6)	0.0309	0.0299	0.032(9)	0.0328	0.031	0.033(12)	0.0341	0.0333	0.046(9)	0.0449	0.0447	0.060(9)	0.0593	0.0613
24.2	0.038(9)	0.0379	0.0361	0.033(9)	0.0414	0.0378	0.046(16)	0.0441	0.0418	0.066(12)	0.0645	0.0613	0.095(12)	0.0917	0.0918
32.2	0.037(11)	0.0441	0.0413	0.047(11)	0.0491	0.0435	0.039(19)	0.0541	0.05	0.081(11)	0.0839	0.0771	0.113(11)	0.1257	0.1202
40.2	0.052(12)	0.0482	0.0443	0.068(12)	0.0541	0.0468	0.062(17)	0.0626	0.0565	0.097(17)	0.0998	0.09	0.17(16)	0.1562	0.1429

τ_k	u = 13			u = 14			u = 15			u = 16			u = 17		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.010(5)	0.0171	0.0197	0.003(6)	0.0178	0.0205	0.008(8)	0.0178	0.0224	0.012(7)	0.0192	0.0221	0.004(9)	0.0193	0.0218
8.2	0.034(8)	0.04	0.0433	0.033(9)	0.0372	0.0403	0.043(8)	0.0413	0.046	0.044(9)	0.0426	0.0461	0.058(10)	0.0452	0.049
16.2	0.067(12)	0.0763	0.0799	0.088(12)	0.0646	0.0681	0.093(12)	0.0794	0.0831	0.090(10)	0.081	0.0846	0.099(7)	0.0903	0.0955
24.2	0.102(15)	0.1228	0.1223	0.109(13)	0.102	0.1041	0.123(11)	0.1287	0.1274	0.128(8)	0.1321	0.1327	0.155(13)	0.1487	0.1505
32.2	0.172(14)	0.1678	0.1579	0.138(14)	0.1428	0.141	0.173(14)	0.1772	0.1669	0.179(11)	0.1853	0.179	0.192(11)	0.2066	0.1988
40.2	0.218(16)	0.2033	0.1824	0.171(11)	0.1817	0.1746	0.215(19)	0.2169	0.1971	0.238(15)	0.2329	0.2184	0.247(16)	0.2547	0.2348

τ_k	u = 18			u = 19			u = 20			u = 21			u = 22		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.011(7)	0.0179	0.0218	0.010(7)	0.0172	0.0197	0.022(12)	0.0158	0.0178	0.010(9)	0.015	0.0156	0.011(9)	0.0134	0.0153
8.2	0.055(11)	0.0407	0.0453	0.022(5)	0.0347	0.0366	0.017(12)	0.0377	0.0405	0.005(10)	0.029	0.0296	0.042(9)	0.027	0.0279
16.2	0.085(11)	0.0772	0.0819	0.064(6)	0.057	0.0573	0.068(9)	0.0709	0.0742	0.028(11)	0.0412	0.0419	0.041(9)	0.0376	0.0358
24.2	0.126(10)	0.1248	0.1259	0.082(8)	0.0878	0.0856	0.116(15)	0.1126	0.1111	0.052(13)	0.058	0.0584	0.021(12)	0.0507	0.0455
32.2	0.174(12)	0.1724	0.1656	0.131(11)	0.1224	0.1174	0.161(12)	0.1513	0.1387	0.074(13)	0.0774	0.0768	0.070(11)	0.0632	0.055
40.2	0.188(20)	0.2126	0.1965	0.145(13)	0.1567	0.1491	0.177(24)	0.1791	0.1533	0.072(16)	0.0972	0.0944	0.084(16)	0.0727	0.0628

Table D3 (cont'd)

τ_k (ms)	u = 23		u = 24		u = 28	
	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Eq. 5
2.2	0.026(10)	0.0135 0.0131	0.006(5)	0.013 0.013	0.016(9)	0.01257
8.2	0.014(14)	0.0276 0.0267	0.016(7)	0.0252 0.0251	0.017(10)	0.02433
16.2	0.049(11)	0.0397 0.0366	0.031(6)	0.0319 0.0316	0.021(11)	0.02974
24.2	0.059(18)	0.0545 0.0475	0.024(8)	0.0399 0.0393	0.032(13)	0.03571
32.2	0.057(16)	0.0684 0.056	0.015(19)	0.0477 0.0464	0.045(13)	0.04071
40.2	0.089(19)	0.0784 0.0605	0.050(14)	0.0537 0.0515	0.043(17)	0.04349

^a For u=28, the values are from experiment and from calculation with Eqs. 1-5. The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. 0.015(10) means 0.015 ± 0.010 .

Table D4. V2E $\Delta S/S_0$ values from experiment and from unconstrained and constrained fittings with $b = 0.98^a$.

τ_k	u = 8			u = 9			u = 10			u = 11			u = 12		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.009(10)	0.0126	0.0126	0.014(10)	0.0126	0.0126	0.024(10)	0.0126	0.0126	-0.01(1)	0.0127	0.0126	0.001(10)	0.0128	0.0126
8.2	0.026(10)	0.0243	0.0243	0.005(10)	0.0243	0.0243	0.018(10)	0.0243	0.0243	0.006(11)	0.0246	0.0243	0.012(10)	0.0252	0.0243
16.2	0.028(10)	0.0297	0.0297	0.027(10)	0.0297	0.0297	0.024(10)	0.0297	0.0297	0.031(10)	0.0304	0.0297	0.022(10)	0.0325	0.0297
24.2	0.018(10)	0.0357	0.0357	0.032(10)	0.0357	0.0357	0.041(12)	0.0357	0.0357	0.040(12)	0.0371	0.0357	0.030(14)	0.0409	0.0357
32.2	0.053(14)	0.0407	0.0407	0.047(10)	0.0407	0.0407	0.047(12)	0.0407	0.0407	0.050(11)	0.043	0.0407	0.059(10)	0.0482	0.0407
40.2	0.052(10)	0.0435	0.0435	0.058(12)	0.0435	0.0435	0.050(13)	0.0435	0.0435	0.053(17)	0.0468	0.0435	0.086(10)	0.0527	0.0435

τ_k	u = 13			u = 14			u = 15			u = 16			u = 17		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.002(13)	0.0128	0.0138	0.007(17)	0.0139	0.0155	0.004(13)	0.0161	0.0179	0.019(10)	0.0179	0.0212	0.019(12)	0.02	0.0214
8.2	0.044(17)	0.0248	0.0257	0.043(13)	0.0273	0.0287	0.039(10)	0.0352	0.0372	0.034(12)	0.0405	0.0442	0.035(20)	0.0444	0.0468
16.2	0.038(10)	0.0309	0.0316	0.034(10)	0.0375	0.0382	0.046(16)	0.0613	0.0628	0.073(10)	0.0764	0.0798	0.078(20)	0.0856	0.0888
24.2	0.047(16)	0.038	0.0383	0.061(10)	0.0507	0.0507	0.095(10)	0.0952	0.0931	0.129(12)	0.1233	0.1231	0.175(19)	0.1406	0.1407
32.2	0.063(20)	0.0445	0.0444	0.073(12)	0.0646	0.064	0.143(10)	0.1292	0.1198	0.197(15)	0.1706	0.1629	0.238(16)	0.1981	0.1896
40.2	0.078(25)	0.0491	0.0486	0.114(14)	0.0773	0.0762	0.175(11)	0.1577	0.1397	0.245(14)	0.2107	0.1946	0.271(14)	0.2497	0.23

τ_k	u = 18			u = 19			u = 20			u = 21			u = 22		
	(ms)	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98	Expt.	Uncons. Cons. b=0.98 b=0.98		
2.2	0.025(17)	0.0192	0.0269	0.013(11)	0.0223	0.0248	0.009(10)	0.0213	0.0248	0.007(10)	0.0192	0.0203	0.005(10)	0.014	0.0198
8.2	0.057(13)	0.048	0.0557	0.046(10)	0.0448	0.0463	0.056(10)	0.06	0.0676	0.043(10)	0.0392	0.0393	0.022(11)	0.027	0.0331
16.2	0.113(12)	0.1004	0.106	0.069(14)	0.0818	0.0808	0.146(10)	0.1381	0.1521	0.075(10)	0.0689	0.066	0.029(10)	0.0364	0.0432
24.2	0.194(12)	0.1667	0.1646	0.144(10)	0.1347	0.1313	0.262(10)	0.2337	0.2441	0.130(10)	0.1104	0.1037	0.060(12)	0.0488	0.0567
32.2	0.254(19)	0.2295	0.2143	0.213(10)	0.1963	0.1917	0.33(10)	0.3181	0.31	0.179(10)	0.1574	0.1471	0.075(12)	0.0622	0.0717
40.2	0.302(18)	0.2777	0.2498	0.280(12)	0.2599	0.2558	0.379(14)	0.376	0.3422	0.198(15)	0.2045	0.1914	0.103(14)	0.075	0.0864

Table D4 (cont'd)

τ_k (ms)	u = 23			u = 24		
	Expt.	Uncons. Cons. b=0.98 b=0.98		Expt.	Uncons. Cons. b=0.98 b=0.98	
2.2	0.011(10)	0.0135	0.0135	0.001(13)	0.0133	0.0132
8.2	0.016(11)	0.0287	0.0279	0.014(11)	0.0257	0.0254
16.2	0.041(10)	0.0434	0.04	0.029(10)	0.0332	0.0324
24.2	0.077(10)	0.0614	0.0531	0.054(13)	0.0425	0.0408
32.2	0.087(11)	0.0777	0.0631	0.058(19)	0.0519	0.0489
40.2	0.103(10)	0.0887	0.0682	0.090(15)	0.06	0.0551

^a The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. 0.015(10) means 0.015 ± 0.010 .

Table D5. The $f(t)$ for unconstrained and constrained fittings with different b values, and based on $k=1-6$, $\tau_k=2.2-40.2$ ms, or $k=1-7$, $\tau_k=2.2-48.2$ ms data from $u=8-24$ samples. For all WT fittings, the average value of $\langle t \rangle$ and RMSD is $\langle \langle t \rangle^{WT} \rangle = 16.132 \pm 0.048$. For all V2E fittings, $\langle \langle t \rangle^{V2E} \rangle = 18.475 \pm 0.028$.

	WT	WT	WT	WT	WT	WT	WT	V2E	V2E	V2E	V2E	V2E	V2E	V2E	V2E
	Uncons	Uncons	Uncons	Uncons	Cons.	Cons.	Cons.	Uncons	Uncons	Uncons	Uncons	Cons.	Cons.	Cons.	Cons.
	b=0.98	b=0.98	b=1	b=1	b=0.98	b=1	b=1	b=0.98	b=0.98	b=1	b=1	b=0.98	b=1	b=1	b=0.9641 b=0.9554
	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-6	k=1-7	k=1-6 k=1-7
	$\chi^2=107$	$\chi^2=130$	$\chi^2=97$	$\chi^2=123$	$\chi^2=131$	$\chi^2=117$	$\chi^2=163$	$\chi^2=145$	$\chi^2=221$	$\chi^2=168$	$\chi^2=250$	$\chi^2=231$	$\chi^2=277$	$\chi^2=422$	$\chi^2=220 \chi^2=333$
	$\langle t \rangle=16.18$	$\langle t \rangle=16.08$	$\langle t \rangle=16.19$	$\langle t \rangle=16.09$	$\langle t \rangle=16.17$	$\langle t \rangle=16.11$	$\langle t \rangle=16.09$	$\langle t \rangle=18.46$	$\langle t \rangle=18.42$	$\langle t \rangle=18.49$	$\langle t \rangle=18.45$	$\langle t \rangle=18.50$	$\langle t \rangle=18.49$	$\langle t \rangle=18.49$	$\langle t \rangle=18.50 \langle t \rangle=18.49$
t															
8	0.0015	0.0029	0	0.0009	0	0	0	0	0	0	0	0	0	0	0
9	0.0090	0.0066	0.0090	0.0073	0.0027	0	0	0	0	0	0	0	0	0	0
10	0.0032	0.0130	0	0.0100	0	0	0	0	0	0	0	0	0	0	0
11	0.0355	0.0379	0.0346	0.0375	0.0247	0.0230	0.0293	0	0	0	0	0	0	0	0
12	0.0579	0.0608	0.0598	0.0626	0.0672	0.0681	0.0688	0.0092	0.0095	0.0067	0.0073	0	0	0	0
13	0.1306	0.1335	0.1294	0.1321	0.1384	0.1369	0.1400	0	0	0	0	0	0	0	0
14	0.0524	0.0543	0.0555	0.0575	0.0545	0.0577	0.0586	0.0065	0.0242	0.0038	0.0223	0	0	0	0.0017
15	0.1297	0.1285	0.1285	0.1268	0.1266	0.1301	0.1282	0.0769	0.0743	0.0776	0.0756	0.0806	0.0762	0.0950	0.0843 0.1013
16	0.1035	0.0990	0.1073	0.1025	0.1069	0.1113	0.1095	0.1113	0.1098	0.1118	0.1096	0.1114	0.1199	0.1140	0.1047 0.0976
17	0.1514	0.1509	0.1536	0.1524	0.1678	0.1720	0.1712	0.1106	0.1061	0.1104	0.1063	0.1254	0.1083	0.0972	0.1389 0.1322
18	0.1159	0.1071	0.1152	0.1060	0.1206	0.1182	0.1064	0.2054	0.1903	0.2014	0.1861	0.1993	0.2108	0.2020	0.1895 0.1781
19	0.0290	0.0260	0.0342	0.0302	0.0138	0.0280	0.0255	0.0351	0.0444	0.0434	0.0523	0.0058	0.0104	0.0123	0.0020 0.0034
20	0.1325	0.1240	0.1302	0.1218	0.1490	0.1384	0.1341	0.3564	0.3437	0.3549	0.3409	0.4275	0.4263	0.4225	0.4291 0.4286
21	0	0	0	0	0	0	0	0.0425	0.0413	0.0463	0.0465	0.0137	0.0183	0.0203	0.0091 0.00076
22	0.0193	0.0245	0.0184	0.0242	0.0033	0.0070	0.0152	0	0.0107	0	0.0081	0	0	0	0
23	0.0285	0.0309	0.0244	0.0282	0.0244	0.0094	0.0131	0.0460	0.0455	0.0436	0.0450	0.0364	0.0298	0.0367	0.0424 0.0495
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table D6. Comparison between $f(t)^{\text{WT}}$ and values from fitting to Eq. 19 with contributions to $G(t)^{\text{a}}$.

t	$f(t)^{\text{WT}}$	Fitted $f(t)$	$(t-11) \times G_{\beta}^{\text{WT}}$	$L(t) \times G_{\text{Leu}}^{\text{WT}}$	$g^{\text{WT}} \times G(t)_{\text{sc}}^{\text{WT}}$
	kcal/mole				
11	0.0355	0.0221	0	0	-0.387
12	0.0579	0.0429	-0.113	0	-0.671
13	0.1306	0.1371	-0.226	-0.350	-0.905
14	0.0524	0.0638	-0.339	0	-0.683
15	0.1297	0.1380	-0.452	-0.350	-0.683
16	0.1035	0.0930	-0.565	0	-0.683
17	0.1514	0.1390	-0.679	-0.350	-0.461
18	0.1159	0.1132	-0.792	-0.350	-0.225
19	0.0290	0.0520	-0.905	0	0.005
20	0.1325	0.1325	-1.018	-0.350	-0.093

^a The $f(t)^{\text{WT}}$ are from the unconstrained model with $b=0.98$ and the $k=1-6$, $\tau_k=2.2-40.2$ ms data of samples $u = 8-24$, Figure 6.6 and Table 6.3. The $G_{\beta}^{\text{WT}} = -0.113$ kcal/mole, $G_{\text{Leu}}^{\text{WT}} = -0.350$ kcal/mole, and $g^{\text{WT}} = 0.129$. Each $G(t)_{\text{sc}}^{\text{WT}}$ is the sum of free energies of membrane insertion of sidechains for residues between V2 and $t-1$ with sidechain energy relative to Ala, *Proc. Natl. Acad. Sci. U.S.A.* (2011) 108, 10174-10177. The $f(t)^{\text{WT}}$ and the free energy contributions are displayed as a bar plot in Figure 6.7a.

Table D7. Comparison between $f(t)^{V2E}$ and values from fitting to Eq. 20 with contributions to $G(t)^a$.

	$f(t)^{WT}$	Fitted $f(t)$	$(t-15) \times G_{\beta}^{V2E}$	$L(t) \times G_{Leu}^{V2E}$
t			kcal/mole	
15	0.0769	0.0703	0	-1.404
17	0.1106	0.1348	-0.391	-1.404
18	0.2054	0.1868	-0.586	-1.404
19	0.0351	0.0249	-0.782	0
20	0.3564	0.3583	-0.977	-1.404
21	0.0425	0.0478	-1.173	0

^a The $f(t)^{V2E}$ are from the unconstrained model with $b=0.98$ and the $k=1-6$, $\tau_k=2.2-40.2$ ms data of samples $u = 8-24$, Figure 6.6 and Table 6.3. The $G_{\beta}^{V2E} = -0.195$ kcal/mole and $G_{Leu}^{V2E} = -1.40$ kcal/mole. The $f(t)^{V2E}$ and the free energy contributions are displayed as a bar plot in Figure 6.7b.

Example Python code for data fitting:

Experimental REDOR and sigma data of WT-HFP:

```
exp8 = 0.006, 0.017, 0.030, 0.038, 0.037, 0.052, 0.057
exp9 = 0.012, 0.009, 0.032, 0.033, 0.047, 0.068, 0.063
exp10 = 0.015, 0.022, 0.033, 0.046, 0.039, 0.062, 0.111
exp11 = 0.014, 0.026, 0.046, 0.066, 0.081, 0.097, 0.151
exp12 = 0.011, 0.016, 0.060, 0.095, 0.113, 0.170, 0.215
exp13 = 0.010, 0.034, 0.067, 0.102, 0.172, 0.218, 0.256
exp14 = 0.003, 0.033, 0.088, 0.109, 0.138, 0.171, 0.235
exp15 = 0.008, 0.043, 0.093, 0.123, 0.173, 0.215, 0.244
exp16 = 0.012, 0.044, 0.090, 0.128, 0.179, 0.238, 0.253
exp17 = 0.004, 0.058, 0.099, 0.155, 0.192, 0.247, 0.275
exp18 = 0.011, 0.055, 0.085, 0.126, 0.174, 0.188, 0.201
exp19 = 0.010, 0.022, 0.064, 0.082, 0.131, 0.145, 0.157
exp20 = 0.022, 0.017, 0.068, 0.116, 0.161, 0.177, 0.175
exp21 = 0.010, 0.005, 0.028, 0.052, 0.074, 0.072, 0.112
exp22 = 0.011, 0.042, 0.041, 0.021, 0.070, 0.084, 0.096
exp23 = 0.026, 0.014, 0.049, 0.059, 0.057, 0.089, 0.113
exp24 = 0.006, 0.016, 0.031, 0.024, 0.015, 0.050, 0.046
sigma8 = 0.007, 0.006, 0.006, 0.009, 0.011, 0.012, 0.016
sigma9 = 0.006, 0.006, 0.009, 0.009, 0.011, 0.012, 0.018
sigma10 = 0.010, 0.007, 0.012, 0.016, 0.019, 0.017, 0.021
sigma11 = 0.005, 0.006, 0.009, 0.012, 0.011, 0.017, 0.022
sigma12 = 0.009, 0.009, 0.009, 0.012, 0.011, 0.016, 0.023
sigma13 = 0.005, 0.008, 0.012, 0.015, 0.014, 0.016, 0.025
sigma14 = 0.006, 0.009, 0.012, 0.013, 0.014, 0.011, 0.021
sigma15 = 0.008, 0.008, 0.012, 0.011, 0.014, 0.019, 0.019
sigma16 = 0.007, 0.009, 0.010, 0.008, 0.011, 0.015, 0.015
sigma17 = 0.009, 0.010, 0.007, 0.013, 0.011, 0.016, 0.021
sigma18 = 0.007, 0.011, 0.011, 0.010, 0.012, 0.020, 0.021
sigma19 = 0.007, 0.005, 0.006, 0.008, 0.011, 0.013, 0.013
```

```

sigma20 = 0.012, 0.012, 0.009, 0.015, 0.012, 0.024, 0.015
sigma21 = 0.009, 0.010, 0.011, 0.013, 0.013, 0.016, 0.016
sigma22 = 0.009, 0.009, 0.009, 0.012, 0.011, 0.016, 0.013
sigma23 = 0.010, 0.014, 0.011, 0.018, 0.016, 0.019, 0.019
sigma24 = 0.005, 0.007, 0.006, 0.008, 0.019, 0.014, 0.014
1. WT_DualAnnealing_5registries_6data_B0.98
import numpy as np
from scipy.optimize import dual_annealing
# Input data from files
fname_exp = '/Users/yijinzhang/Desktop/REDOR/python_WT/exp.txt'
fh_exp = open(fname_exp)
fname_sig = '/Users/yijinzhang/Desktop/REDOR/python_WT/sigma.txt'
fh_sig = open(fname_sig)
lst_exp = []
for line in fh_exp:
    line = line.rstrip()
    lst_exp.append(line)
exp_0 = {}
for i in range(0,len(lst_exp)):
    line = lst_exp[i].split('=')[1]
    exp0 = line.split(',')
    new_lst=[]
    for j in range(0,len(exp0)-1):
        k= float(exp0[j])
        new_lst.append(k)
    exp_0['exp'+str(i+8)] = np.array(new_lst)
dephasing = []
for i in range (8,25):
    item = sum([exp_0['exp'+str(i)][j] for j in range(6)])
    dephasing.append(item)
def dephasing_time(time):

```

```

if time == 2.2:
    x = 0
if time == 8.2:
    x = 1
if time == 16.2:
    x = 2
if time == 24.2:
    x = 3
if time == 32.2:
    x = 4
if time == 40.2:
    x = 5
return [exp_0[exp'+str(i+8)][x] for i in range(17)]
lst_sig = []
for line in fh_sig:
    line = line.rstrip()
    lst_sig.append(line)
sig = { }
for i in range(0,len(lst_sig)):
    line = lst_sig[i].split('=')[1]
    sig0 = line.split(',')
    new_lst=[]
    for j in range(0,len(sig0)-1):
        k= float(sig0[j])
        new_lst.append(k)
    sig['sigma'+str(i+8)] = np.array(new_lst)
gamma_of_na = np.array((0.7156, 0.449825000000000003, 0.32737499999999997,
                        0.192225, 0.07915000000000001, 0.01639999999999998))
gamma_nad = gamma_of_na * 0.0588 + 0.286
gon = np.array((0.991709633,0.893821899,0.645290442,0.378629871,
                0.196398598,0.118554369))

```

```

goff = np.array((0.998385086,0.978470252,0.918588976,0.826254485,
                 0.710329562,0.581357776))
goff_5registry= np.array((0.9998,0.9971,0.9890,0.9755,0.9569,0.9334))
S0 = 1.33
Slab = 0.9852
#### objective function
def chi_square_all(data):
    def ind_chi(i):
        j = data[i-2]
        x = data[i-1]
        y = data[i]
        z = data[i+1]
        k = data[i+2]
        t = i + 6
        gon_def = gon
        goff_def = goff
        goff_def_2 = goff_5registry
        gamma_naddef = gamma_nad
        exp_def = exp_0['exp'+str(t)]
        sig_def = sig['sigma'+str(t)]
        chi_square_ind = np.array(np.sum(pow((S0-
Slab/sum(data)*(0.98*gon_def*y+(z+x)*0.98*goff_def+(j+k) * 0.98*goff_def_2
                + sum(data) -x-z-y-j-k)-gamma_naddef)/S0
                -exp_def,2)/pow(sig_def,2))))
        return chi_square_ind
    chi_square_sum = np.sum([ind_chi(i) for i in range (2, len(data)-2)])
    return chi_square_sum
another_bounds = [[0,1.0]]*21
result_1 = dual_annealing(chi_square_all, another_bounds, maxiter=1000, accept=-5)
2. WT_DA_3reg_b_Uncon_6data (b1 = 0.98 b2 = 0.99 b3 = 1.0)
import numpy as np

```

```

from scipy.optimize import dual_annealing
# Input data from files
fname_exp = '/Users/yijinzhang/Desktop/REDOR/python_WT/exp.txt'
fh_exp = open(fname_exp)
fname_sig = '/Users/yijinzhang/Desktop/REDOR/python_WT/sigma.txt'
fh_sig = open(fname_sig)
lst_exp = []
for line in fh_exp:
    line = line.rstrip()
    lst_exp.append(line)
exp_0 = {}
for i in range(len(lst_exp)):
    line = lst_exp[i].split('=')[1]
    exp0 = line.split(',')
    new_lst=[]
    for j in range(len(exp0)-1):
        k= float(exp0[j])
        new_lst.append(k)
    exp_0['exp'+str(i+8)] = np.array(new_lst)
lst_sig = []
for line in fh_sig:
    line = line.rstrip()
    lst_sig.append(line)
sig = {}
for i in range(0,len(lst_sig)):
    line = lst_sig[i].split('=')[1]
    sig0 = line.split(',')
    new_lst=[]
    for j in range(0,len(sig0)-1):
        k= float(sig0[j])
        new_lst.append(k)

```

```

sig['sigma'+str(i+8)] = np.array(new_lst)
gamma_of_na = np.array((0.7156, 0.44982500000000003, 0.32737499999999997,
                        0.192225, 0.07915000000000001, 0.01639999999999998))
gamma_nad = gamma_of_na * 0.0588 + 0.286
guu = np.array((0.9917,0.8938,0.6453,0.3786,
                0.1964,0.1186))
# gamma for t_top=u-1 t_bottom=u
gum1u = np.array((0.9980, 0.9737, 0.9012, 0.7914, 0.6571, 0.5132))
# gamma for t_top=u+1 t_bottom=u
gup1u = np.array((0.9980, 0.9737, 0.9012, 0.7914, 0.6571, 0.5132))
# gamma for t_top=x t_bottom=u
gxu = np.array((0.9989, 0.985, 0.9427, 0.8754, 0.787, 0.6824))
# gamma for t_top=u t_bottom=u-1
guum1 = np.array((0.9921, 0.8974, 0.6476, 0.3565, 0.1288, 0.0156))
# gamma for t_top=u-1 t_bottom=u-1
gumlum1 = np.array((0.9984,0.9785,0.9186,0.8263,
                    0.7103,0.5814))
# gamma for t_top=u+1 t_bottom=u-1
guplum1 = np.array((0.9984, 0.9785, 0.9191, 0.8293, 0.7200, 0.6037))
# gamma for t_top=x, t_bottom=u-1
gxum1 = np.array((0.9992, 0.9898, 0.9608, 0.9143, 0.8521, 0.7769))
# gamma for t_top=u t_bottom=u+1
guup1 = np.array((0.9921, 0.8974, 0.6476, 0.3565, 0.1288, 0.0156))
# gamma for t_top=u-1 t_bottom=u+1
gumlup1 = np.array((0.9984, 0.9785, 0.9191, 0.8293, 0.7200, 0.6037))
# gamma for t_top=u+1 t_bottom=u+1
guplup1 = np.array((0.9984,0.9785,0.9186,0.8263,
                    0.7103,0.5814))
# gamma for t_top=x t_bottom=u+1
gxup1 = np.array((0.9992, 0.9898, 0.9608, 0.9143, 0.8521, 0.7769))
# gamma for t_top=u t_bottom=x

```



```

gux = np.array((0.9928, 0.9064, 0.671, 0.3778, 0.1236, -0.023))
# gamma for t_top=u-1 t_bottom=x
gum1x = np.array((0.9991, 0.9885, 0.956, 0.9039, 0.8348, 0.752))
# gamma for t_top=u+1 t_bottom=x
gup1x = np.array((0.9991, 0.9885, 0.956, 0.9039, 0.8348, 0.752))
S0 = 1.33
Slab = 0.9852
##### objective function
def chi_square_all(data):
    def ind_chi(i):
        x = data[i-1]
        y = data[i]
        z = data[i+1]
        rest = sum(data)-x-y-z
        t = i + 7
        # gamma_naddef = gamma_nad
        exp_def = exp_0['exp'+str(t)]
        sig_def = sig['sigma'+str(t)]
        # scaling factor b1(t_top=u,u±1;t_bottom=u,u±1) = 0.98
        # b2(t_top=u,u±1; t_bottom=x) = b3 (t_top=x;t_bottom=u,u±1) = 0.99
        # b3 (t_top=x;t_bottom=x)=1.0
        b1 = 0.98
        b2 = 0.99
        b3 = 1.0
        cal = (b1*guu*y*y+b1*gum1u*x*y+b1*gup1u*z*y+b2*gxu*rest*y+
        b1*guum1*y*x+b1*gum1um1*x*x+b1*gup1um1*z*x+b2*gxum1*rest*x+
        b1*guup1*y*z+b1*gum1up1*x*z+b1*gup1up1*z*z+b2*gxup1*rest*z+
        b2*gux*y*rest+b2*gum1x*x*rest+b2*gup1x*z*rest+
        b3*rest*rest)
        chi_square_ind = np.array(np.sum(pow(((S0-Slab/pow(sum(data),2)*cal
        -gamma_nad)/S0

```

```
        -exp_def),2)/pow(sig_def,2)))  
    return chi_square_ind  
    chi_square_sum = np.sum([ind_chi(i) for i in range (1, len(data)-1)])  
    return chi_square_sum  
another_bounds = [[0,0.00001]]+ [[0,1.0]]*17+[[0,0.00001]  
result = dual_annealing(chi_square_all, another_bounds, maxiter=1000, accept=-5)
```