

PRODUCTION, PURIFICATION, QUANTIFICATION, AND LABELING OF RECOMBINANT PROTEINS
AND SOLID STATE NUCLEAR MAGNETIC RESONANCE STUDIES
IN MEMBRANES AND CELLULAR MATERIALS

By

Erica Paige Vogel

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Chemistry and Quantitative Biology

2012

ABSTRACT

PRODUCTION, PURIFICATION, QUANTIFICATION, AND LABELING OF RECOMBINANT PROTEINS AND SOLID STATE NUCLEAR MAGNETIC RESONANCE STUDIES IN MEMBRANES AND CELLULAR MATERIALS

By

Erica Paige Vogel

Solid state nuclear magnetic resonance (SSNMR) spectroscopy provides the opportunity to obtain high resolution data regarding the chemical environment of NMR active nuclei in solid and semi-solid samples. Of particular interest for study by SSNMR are biological molecules like proteins, as NMR provides a way to determine properties of these molecules such as secondary structure, internuclear distances, and dynamics. My dissertation project consisted of several different applications of SSNMR to study biological systems, as well as the preparation of these systems for study.

gp41 is a protein present on the surface of virions of the human immunodeficiency virus (HIV). The protein gp41 is a glycoprotein which aids in the process of viral entry into the human host T cells by catalyzing the process of membrane fusion between the viral membrane and the T cell plasma membrane. Due to its implication in this process, it has been an attractive target for anti-HIV drug development. I produced in *E. coli* and purified an ectodomain construct of the gp41 protein called Fgp41 which included the catalytic fusion peptide. Structural analyses by circular dichroism spectroscopy and rotational echo double resonance (REDOR) SSNMR indicated that the protein was folded into the post-fusion low energy six helix bundle conformation. This was further supported by functional assays that showed little lipid-mixing ability of the protein. REDOR SSNMR was used to obtain high resolution structural information

about the protein while associated with lipid membranes. This is the first example of atomic resolution structural data of the fusion peptide embedded into lipid membranes in the context of the protein.

Human proinsulin is the biological precursor to the insulin hormone, which has therapeutic effects for people with the metabolic disease diabetes mellitus. Synthetic insulin is produced in many ways, including through recombinant protein expression in *E. coli* as the precursor protein proinsulin. It is documented that proinsulin is sequestered within inclusion bodies after recombinant expression, and drastic measures are taken to denature and refold the protein to produce bioactive insulin. By utilizing SSNMR, the REDOR pulse sequence, and selective isotopic labeling schemes, I was able to probe the secondary structure of human proinsulin within bacterial inclusion bodies. Both helical and β -strand conformations of the protein were observed in the A and B chains, while C chain (which is cleaved during the processing to form insulin) exhibited primarily neither helical nor β -strand chemical shifts.

Recombinant expression in *E. coli* is a major way of producing protein for structural and functional studies. Different proteins express to different levels within *E. coli*, and for proteins that are difficult to solubilize, it is often difficult to determine whether they are expressing at all. By utilizing SSNMR, REDOR, and isotopically labeled whole *E. coli* cells I was able to detect the level of recombinant protein expressed. The NMR spectrum is simplified if the sample preparation includes a step to remove all soluble proteins. By comparison to a standard curve, I was able to determine the level of recombinant protein expression in mg protein produced per liter of bacterial cell culture for several different protein constructs. This is the first method of recombinant protein expression quantification in whole cells or insoluble cell pellets.

ACKNOWLEDGEMENTS

I would like to first and foremost thank my advisor, David Weliky, for his support and encouragement throughout the course of my Ph.D. work. He allowed me the freedom to explore research projects that I was interested in, and encouraged me to focus when it was important. He taught me the importance of looking at data and making my own interpretations, and not just taking another's conclusions at face value. I also learned the importance of running control experiments while in his group, and perhaps became a bit too cautious with presenting results as fact. David has helped me to become a careful and thorough scientist, and I truly appreciate this.

I would also like to acknowledge former Weliky group members that helped me. Jaime Curtis-Fisk taught me much of the E. coli culture and protein purification skills that I needed. Wei Qiang, Yan Sun, and Scott Schmick were a huge help in teaching me about solid-state NMR. Matt Nethercott was helpful with MALDI and gel filtration experiments. I also had the opportunity to work with two very talented undergraduate research students. Kaitlin Young was a huge help with all of the Fgp41 sample prep, and Ryan Spencer helped get the human proinsulin project going.

Current Weliky group members have been very helpful over the past few years. Kelly Sackett taught me how to perform lipid mixing assays, and has been great for bouncing research ideas off of. Charles Gabrys has been helpful with ideas in research, data processing, and life. Li Xie was very helpful in teaching me some of the theory behind SSNMR. Koyeli Banerjee has taught me better ways to perform PCR and been very helpful in coming up with ideas on protein purifications when I've gotten stuck. Punsisi Ratnayake has been fun to teach

in lab, has shared with me what she has learned about cloning and other things. Ujjayini Ghosh has been a great help with math when I have gotten stuck. The whole Weliky group has been very helpful and a joy to be around.

I would also like to thank my husband, Paul Vogel, for his support during my entire Ph.D. career, and for coming to grad school against his better judgment. The support from Paul and my friends and family has made this whole process much easier than it would have been on my own.

.....

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xiii
LIST OF ABBREVIATIONS.....	xxii
Chapter 1 – Introduction	1
<i>Nuclear Magnetic Resonance</i>	1
<i>NMR Theory</i>	1
<i>Zeeman Splitting.....</i>	2
<i>The effect of radiofrequency (RF) pulses</i>	5
<i>Biomolecular NMR and sensitivity</i>	6
<i>Isotopic enrichment</i>	7
<i>Cross Polarization (CP)</i>	7
<i>Solid State Nuclear Magnetic Resonance (SSNMR).....</i>	10
<i>Magic Angle Spinning (MAS) NMR.....</i>	11
<i>Dipolar Coupling (DC)</i>	12
<i>Chemical shift anisotropy (CSA)</i>	13
<i>Rotational Echo Double Resonance (REDOR) NMR.....</i>	15
<i>The REDOR S_0 experiment.....</i>	16
<i>The REDOR S_1 experiment.....</i>	16
<i>Applications of REDOR NMR</i>	19
<i>Human Immunodeficiency Virus (HIV) Fusion Protein gp41</i>	22
<i>gp41 fusion peptide (FP)</i>	23
<i>Fgp41 – an ectodomain construct of gp41</i>	25
<i>Bacterial Inclusion Bodies</i>	26
<i>Utilization of inclusion bodies</i>	27
<i>Quantitative detection of protein in inclusion bodies</i>	28
<i>Diabetes and the prehormone human proinsulin</i>	28
<i>Synthetic production of insulin.....</i>	28
<i>Structural studies of human proinsulin.....</i>	29
REFERENCES	30
Chapter 2 – Studies of Fgp41, an ectodomain construct of HIV fusion protein gp41	36
<i>Introduction</i>	36
<i>Fgp41 Construct Information</i>	37
<i>Source of Fgp41.....</i>	37
<i>DNA Sequence of Fgp41.....</i>	37
<i>Protein Sequence of Fgp41</i>	37
<i>Fgp41 Expression Optimization</i>	38
<i>Fgp41 Purification Protocol Development</i>	39

Circular Dichroism Spectroscopy of Fgp41	43
Fluorescence Based Lipid-Mixing Assays for Activity of Fgp41	44
Experimental Details	45
Solid-State NMR Analysis of Membrane Associated Fgp41	47
Membrane Reconstitution	47
SSNMR Experimental Parameters	47
SSNMR Experimental Results	48
Discussion of Results of Fgp41 studies	58
Expanded Studies of Fgp41	61
Mutations to Fgp41 to Enhance Solubility	61
Expression and Purification of Fgp41 _{noCys}	64
Future Work	66
REFERENCES	68
Chapter 3 – Development of a quantitative method of recombinant protein expression in whole E. coli cells and bacterial inclusion bodies	72
Introduction	72
Protein Construct Information	74
Sample Preparation	76
Protein Expression	76
NMR Sample Preparation for Insoluble Cell Pellet Experiments	80
NMR Sample Preparation for Whole Cell Experiments	80
NMR Experimental Parameters	80
Whole Cell SSNMR Spectroscopy	81
Analysis of the SSNMR spectra of lyophilized whole cells	87
Conclusions from Whole Cell NMR Experiments	92
Insoluble Cell Pellet SSNMR Spectroscopy	93
Quantitative Detection of Recombinant Protein Expression	97
Calculation of Expression Levels	100
Conclusions from ICP NMR Experiments	107
A Possible Alternate Method of Calculating Expression Levels	108
Future Work	112
REFERENCES	114
Chapter 4 – Structural analysis of human proinsulin within bacterial inclusion bodies by solid state NMR	117
Introduction	117
Human Proinsulin Construct Information	118
Source of Human Proinsulin	118
DNA Sequence of Human Proinsulin	118
Amino Acid Sequence of Human Proinsulin	118
Human proinsulin expression	118
NMR sample preparation	119

<i>Isotopic Labeling Considerations</i>	120
<i>Summary of NMR Labeling Schemes</i>	120
NMR Experimental Parameters	121
Experimental Results	122
1- ¹³ C Leu Labeling Schemes.....	123
1- ¹³ C Leu Labeling Schemes.....	123
1- ¹³ C Ala Labeling Schemes.....	126
1- ¹³ C Gly Labeling Schemes.....	129
Summary of experimental results	132
REFERENCES.....	138
APPENDIX A	141
The Entire Ectodomain of gp41 – Fgp41:Fragment2	141
REFERENCES.....	150
APPENDIX B.....	152
Studies of FHA2 – dependence of secondary structure within membranes on sample pH and the presence of cholesterol.....	152
REFERENCES.....	168
APPENDIX C.....	170
Locations of NMR Files	170
Chapter 2 Figures	171
Chapter 3 Figures	171
Chapter 4 Figures	172
Appendix A Figures.....	172
Appendix B Figures	172

LIST OF TABLES

Table 1-1: Gyromagnetic ratios and spin quantum numbers for select biologically important nuclei. This table was adapted from reference (1).....	3
Table 2-1: Analysis and deconvolution of S_0 SSNMR spectra of membrane reconstituted Fgp41. ^a Spectral deconvolution was conducted with three Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For all cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 2-7. Deconvolution was not meaningful for the 1- ¹³ C Ala and 1- ¹³ C Gly samples because the S_0 spectra were broad and relatively featureless, resulting in deconvolutions that were dominated by a line shape with ~7 ppm line width. ^b The conformations designated are assigned based on RefDB(9). ^c Full width at half-maximal line width.....	51
Table 2-2: Comparison between experimental and calculated REDOR dephasing for membrane reconstituted Fgp41.	52
Table 3-1: Protein construct information. The name of the protein construct, plasmid type, and <i>E. coli</i> cell type used are listed for each protein.....	75
Table 3-2: Integrated signal intensities in 15 ppm regions from spectra corresponding to either whole bacterial cells induced to express FHA2 that had been 1- ¹³ C Ala, ¹⁵ N Val labeled with glycerol present as the only additional carbon source in the growth medium, or whole bacterial cells induced to express FHA2 that had been 1- ¹³ C Ala, ¹⁵ N Val labeled with glycerol and all other unlabeled amino acids present in the growth medium. The Ala-Val sequential pair of amino acids does not appear within the FHA2 protein sequence.	79
Table 3-3: Deconvolution of spectra of lyophilized cells induced to produce Fgp41. ^a Spectral deconvolution was done with three Gaussian line shapes whose peak shifts, linewidths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and experimental line shape, see Figure 3-5. ^b The reasons for assignment of peaks to specific conformations are provided in the main text. ^c Full-width at half-maximum linewidth.	89
Table 3-4: Best fit deconvolution of Figure 3-8 spectra. The parameters are for the best-fit Gaussian lineshape of the dominant spectral peak. The integrated signal intensity was obtained by integrating the peak in the difference spectrum that appears between 170 ppm to 185 ppm.	

The uncertainty in integrated signal intensity was calculated using the RMSD integrated intensity of 5 ppm regions without signal.....97

Table 3-5: Information obtained from REDOR S_0 spectra of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu/talc samples. The error in integrated signal intensity was obtained by integrating regions of noise in the S_0 spectrum for 0.5 mg $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu containing sample. This sample was used because all spectra showed apodization of the signal, and this had the least amount. The noise should be the same in all spectra as the same conditions were used for the experiments.98

Table 3-6: Integrated signal intensities from the S_0 spectrum for each ICP sample and the calculated scaling factors. The scaling factor was $[1000/(\text{integrated signal intensity in the } 0 \text{ to } 90 \text{ ppm region})]$103

Table 3-7: Calculated normalized carbonyl signal = $aA - bB$ and expression level for each ICP sample. The # of Leu $\equiv$ number of Leu residues in the recombinant protein sequence. The sample-to-sample variation in recombinant protein expression level is $\sim 10\%$ based on the analysis for the three $1\text{-}^{13}\text{C}$ labeled Leu HPI samples.104

Table 3-8: Data obtained from $^{13}\text{C}\text{-}^{15}\text{N}$ REDOR ΔS spectra of ICP samples processed without line broadening and with a 5^{th} order polynomial baseline correction. Each ΔS spectrum was the result of 50,000 S_0 scans – 50,000 S_1 scans. Line width reported is the Full Width at Half Maximal value, and was measured from the spectra.....111

Table 3-9: Calculated recombinant protein expression levels using the ΔS spectra for the samples mentioned in Table 3-8.....111

Table 4-1: Analysis and deconvolution of ΔS SSNMR spectra of human proinsulin labeled with $1\text{-}^{13}\text{C}$ Leu (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Spectral deconvolution was conducted for Leu_{11,17} and Leu_{15,78} with two Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 4-7. Deconvolution was not meaningful for the Leu₄₄ and Leu₅₆ samples because the ΔS spectra were broad and relatively featureless. The conformations designated are assigned based on characteristic ^{13}CO chemical shifts for different Leu secondary structures which have Gaussian distributions as follows: coil = 176.9 ± 1.7 ppm, helical = 178.5 ± 1.3 ppm, β strand = 175.7 ± 1.5 ppm (4). In refDB, “helical” is defined as $[-120^\circ < \phi < -34^\circ \text{ AND } -80^\circ < \psi < 6^\circ]$. “beta” or β as presented in the table is defined as $[-180^\circ < \phi < -40^\circ \text{ OR } 160^\circ < \phi \leq 180^\circ] \text{ AND } [70^\circ < \psi < 180^\circ \text{ OR } -180^\circ < \psi < -170^\circ]$. “coil” is defined as “everything else”(5).132

Table 4-2: Analysis and deconvolution of ΔS SSNMR spectra of human proinsulin labeled with $1\text{-}^{13}\text{C}$ Ala (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Spectral deconvolution was conducted for Ala_{14,57} and Ala₅₀ with two Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 4-8. The conformations designated are assigned based on characteristic ^{13}C CO chemical shifts for different Ala secondary structures which have Gaussian distributions as follows: coil = 177.7 ± 1.6 ppm, helical = 179.4 ± 1.3 ppm, β strand = 176.1 ± 1.5 ppm (4). Please see the caption for Table 4-1 for an explanation of helical, β strand, and coil in terms of dihedral angles.....134

Table 4-3: Analysis of ΔS SSNMR spectra of human proinsulin labeled with $1\text{-}^{13}\text{C}$ Gly (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Deconvolution was not meaningful for the spectra as the peaks are relatively featureless. The conformations designated are assigned based on characteristic ^{13}C CO chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (4). Please see the caption for Table 4-1 for an explanation of helical, β strand, and coil in terms of dihedral angles.....135

Table A-1: Numerical data obtained from the REDOR S_0 spectra of $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41_{noCys} and Fgp41:Fragment2_{noCys} insoluble cell pellets. To calculate the “scaling factor”, the integrated signal intensity in the 0 to 90 ppm region of the spectrum was divided by 1000. This number was then multiplied by the integrated signal intensity in the carbonyl region and the value from the same process for pET24a(+) sample was subtracted to yield the “reduced carbonyl signal”. The reduced carbonyl signal was divided by the number of Leu residues present in the protein constructs to give the “normalized signal”.149

Table B1: Information obtained from analysis of ΔS spectra observing Phe3 of FHA2 in membranes. The ΔS spectra are shown in Figure B1. Deconvolution was not meaningful because the ΔS spectra were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C CO chemical shifts for different Phe secondary structures which have Gaussian distributions as follows: coil = 175.6 ± 1.6 ppm, helical = 177.3 ± 1.4 ppm, β strand = 174.3 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.164

Table B2: Information obtained from analysis of ΔS spectra observing Gly4 of FHA2 in membranes. The ΔS spectra are shown in Figure B2. Deconvolution of the pH 5 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the other spectra because they were

relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.164

Table B3: Information obtained from analysis of ΔS spectra observing Ala7 of FHA2 in membranes. The ΔS spectra are shown in Figure B3. Deconvolution of the pH 7.4 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the other spectra because they were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Ala secondary structures which have Gaussian distributions as follows: coil = 177.7 ± 1.6 ppm, helical = 179.4 ± 1.3 ppm, β strand = 176.1 ± 1.5 ppm (8). The peak width reported is the full width at half maximal value.165

Table B4: Information obtained from analysis of ΔS spectrum observing Gly16 of FHA2 in membranes. The ΔS spectrum is shown in Figure B4. Deconvolution was not meaningful for the spectrum because it was relatively featureless. The conformation designated is assigned based on characteristic ^{13}C O chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.165

Table B5: Information obtained from analysis of ΔS spectra observing Phe70 of FHA2 in membranes. The ΔS spectra are shown in Figure B5. Deconvolution of the pH 5.0 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the pH 7.4 sample because the ΔS spectrum was relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Phe secondary structures which have Gaussian distributions as follows: coil = 175.6 ± 1.6 ppm, helical = 177.3 ± 1.4 ppm, β strand = 174.3 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.166

Table B6: Information obtained from analysis of ΔS spectra observing Leu98 of FHA2 in membranes. The ΔS spectra are shown in Figure B6. Deconvolution was not meaningful because the ΔS spectra were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Leu secondary structures which have Gaussian distributions as follows: coil = 176.9 ± 1.7 ppm, helical = 178.5 ± 1.3 ppm, β strand = 175.7 ± 1.5 ppm (8). The peak width reported is the full width at half maximal value.166

LIST OF FIGURES

Figure 1-1: Larmor precession of a nucleus in a magnetic field. The static magnetic field $\mathbf{B}_0$ (green) is along the z axis, and thus the nuclear magnetic moment (depicted in blue) rotates around the z axis with a frequency $\omega_0 = \gamma B_0$. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.2

Figure 1-2: Depiction of the breakdown of angles in MAS experiments. The C-N internuclear vector at angle θ to the external magnetic field (depicted in green) can be broken into two components. One component is a vector along the axis of rotation (angle $\theta_{MA}=54.7^\circ$ to the external magnetic field). The other component is 90° to the axis of rotation. If we consider one rotor period, the contribution along the rotor axis will remain unchanged, and the contribution perpendicular to the axis of rotation will average to zero. This is only shown for one internuclear vector direction, but is true for an internuclear vector in any orientation.12

Figure 1-3: Depiction of the principle axes $\delta_{11}, \delta_{22}, \delta_{33}$ with respect the external magnetic field $\mathbf{B}_0$. The angles $\theta_{11}, \theta_{22}, \theta_{33}$ are the angles between the axes and $\mathbf{B}_0$14

Figure 1-4: Simplified model of how the ^{13}C nuclear magnetic moment vector, local field induced by ^{15}N nuclei onto ^{13}C nuclei, and the dipolar coupling energy evolve with time under Magic Angle Spinning conditions in REDOR. The dipolar interaction energy is averaged to zero over each rotor period as shown for the S_0 experiment. As a result of ^{13}C and ^{15}N π pulses, the dipolar interaction energy during the S_1 experiment is nonzero when an average is taken over rotor periods.18

Figure 1-5: Conceptual representation of gp41 structural states with time increasing from left to right. In the middle and right panels, the region shown in blue represents the fusion peptide, red represents the C-terminal helix, and green represents the N-terminal helix. In the right panel of the figure, the red and green helices are antiparallel to one another.22

Figure 2-1: Representative SDS-PAGE gel of soluble cell lysates produced using buffers with different detergents or urea. For each buffer, the left and right lanes respectively correspond to 2 and 5 μL aliquots of lysate. The ~ 19 kDa band apparent in some lanes is assigned to Fgp41. One example is circled in red in lane 4 for lysis in SDS.40

Figure 2-2: Representative SDS-PAGE gel of lysates of soluble cell lysates produced using buffers containing different concentrations of SDS. The ~ 19 kDa band was assigned to Fgp41 and is most apparent in the lane corresponding to 1% SDS lysis buffer.41

Figure 2-3: (a) SDS-PAGE gels of (lane 1) an elution aliquot of Fgp41 in buffer containing 250 mM imidazole and (lane 2) molecular weight standards. (b) SDS-PAGE gel of (lane 1) an aliquot

of the proteoliposome complexes formed during membrane reconstitution of Fgp41 and (lane 2) molecular weight standards. The samples were boiled prior to loading on the gel.42

Figure 2-4: (a) CD spectra of Fgp41 at 25 °C. The black trace is for a sample that has not been heated, and the red trace was obtained after the sample had been heated to 100 °C with subsequent cooling to 25 °C. Each trace is the difference between the CD spectrum of Fgp41 with buffer and the spectrum of buffer alone. Fgp41 samples were prepared by precipitation of excess SDS, subsequent dialysis in HEPES/MES buffer (pH 7.4), and addition of DTT at two times the molar concentration of Fgp41 to inhibit disulfide bond formation. For these spectra, the concentration of Fgp41 was 20 μM. Spectra for other Fgp41 samples were similar with minima near 208 and 222 nm that were diagnostic of α-helical structure. In some spectra, θ_{222} could be as low as $-15000 \text{ deg cm}^2 \text{ dmol}^{-1}$. (b) Plot of CD θ_{222} vs temperature for Fgp41. No unfolding transition is apparent for temperatures up to 100 °C. Sample conditions were the same as those described in (a).43

Figure 2-5: Vesicle fusion assayed by fluorescence. An aliquot of either Fgp41 with buffer (black trace) or buffer alone (red trace) was added to a vesicle solution at 350 s. Fgp41-induced vesicle fusion was evidenced by the fluorescence increase (ΔF_{Fgp41}) of the black trace. In either trace, Triton X-100 was added at 750 s and solubilized the vesicles, resulting in maximal fluorescence and fluorescence increase (ΔF_{max}). The spikes at 350 and 750 s were artifacts caused by transient exposure to stray light. Assay parameters included vesicles with 4:1 POPC:POPG composition, Fgp41:total lipid molar ratio of 1:50, pH 7.5, 37 °C.46

Figure 2-6: REDOR ^{13}C CO NMR spectra of Fgp41 reconstituted in membranes. The labeled amino acids in the expression medium are shown. The left panels display S_0 (blue) and S_1 (red) spectra; the middle panels display the best-fit Gaussian deconvolutions of the S_0 spectra, and the right panels display $\Delta S \equiv S_0 - S_1$ spectra. The REDOR dephasing time was either (a) 1 or (b-f) 2 ms, and the dominant contribution to each ΔS spectrum was from residues labeled with ^{13}C that were directly bonded to labeled ^{15}N atoms. The major contribution to each ΔS spectrum is indicated. Each S_0 or S_1 spectrum was processed with 100 Hz Gaussian line broadening, and each ΔS spectrum was processed with (a and b) 100 or (c-f) 200 Hz line broadening. Polynomial baseline correction (typically fifth order) was applied to each spectrum. Each S_0 or S_1 spectrum was the sum of (a) 93424, (b) 115610, (c) 109504, (d) 110736, (e) 165216, or (f) 103717 scans.50

Figure 2-7: The fittings of S_0 deconvolutions for membrane associated Fgp41 samples are displayed. The labeling present in each sample is indicated. The experiment is shown in orange, the best-fit deconvolution sum is shown in green, and the difference is shown in purple. The best-fit deconvolution sum is the sum of the Gaussian curves shown previously in Figure 2-6. .56

Figure 2-8: Deconvolutions of ΔS spectra are displayed. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.....57

Figure 2-9: The top sequence which is underlined is the Fgp41 sequence, not including the eight non-native residues at the C-terminus. The bottom sequence is the sequence of the HXB2 laboratory isolated strain. The center sequence shows the agreement between the pair.....62

Figure 2-10: An initial attempt at purification of Fgp41_{noCys} involved solubilizing the protein in a buffer containing no detergent. There was no detectable band in earlier attempts to solubilize Fgp41 under the same conditions (data not shown).....65

Figure 2-11: Purification of the insoluble fraction of protein using urea resulted in ~95% pure Fgp41_{noCys} in elution fractions. The yield of this particular purification was estimated as ~1.5 mg pure protein per 5 grams of cells.66

Figure 3-1: ΔS spectra for a) 1- ^{13}C Ala, ^{15}N Val labeled dry whole E. coli cells induced to produce FHA2 with glycerol as the only other carbon source, and b) 1- ^{13}C Ala, ^{15}N Val labeled dry whole E. coli cells induced to produce FHA2 where the growth medium was supplemented with all unlabeled amino acids as well as glycerol. Each ΔS spectrum was the result of a) 46652 ($S_0 - S_1$) scans and b) 43647 ($S_0 - S_1$) scans. The spectra were processed with no line broadening and a) 5th order and b) 7th order polynomial baseline corrections.....78

Figure 3-2: Amino acid sequence of the Fgp41 protein construct. The LL pairs targeted with the 1- ^{13}C , ^{15}N Leu labeling are bolded in the sequence. The fusion peptide region is shown in blue, the N helix and C helix in red and green, respectively. All LL pairs are located either within or right at the end of the helical regions of the protein.82

Figure 3-3: REDOR ^{13}C NMR spectra of whole bacterial cells induced to produce Fgp41 by sequential steps: (1) growth in rich medium, (2) growth in minimal medium, (3) addition of labeled or unlabeled amino acids, (4) induction of Fgp41 expression, (5) centrifugation. The induction temperature and duration were either (a-c) 23 °C and ~2 hr or (d) 37 °C and ~5hr. The left panels display S_0 (blue) and S_1 (red) spectra and the right panels display ΔS spectra. The REDOR dephasing time was either (a-c) 1 ms or (d) 2 ms. For panels a, b, and d, the dominant contribution to each ΔS spectrum was from residues with labeled ^{13}C CO groups that were directly bonded to ^{15}N atoms. These residues were (a and b) L33, L44, L54, L81, L134, and L149 of the LL sequential pairs of Fgp41 and (d) G10 of the G10-F11 unique sequential pair. Each S_0 or S_1 spectrum was processed with 100 Hz Gaussian line broadening, and each ΔS spectrum was processed with either (a and d) 200 or (b and c) 100 Hz line broadening. Polynomial

baseline correction (typically fifth order) was applied to each spectrum. Each S_0 or S_1 was the sum of (a) 100000, (b) 100000, (c) 127222, or (d) 48448 scans.....83

Figure 3-4: REDOR ^{13}C NMR spectra of lyophilized whole bacterial cells induced to produce Fgp41 with either $1\text{-}^{13}\text{C}, ^{15}\text{N}$ labeled Leu or unlabeled Leu. The cell production and NMR parameters are described in the legend of Figure 3-3. Panel a displays the S_0 spectra of the labeled (blue) and unlabeled (black) cells with the relative intensities adjusted to yield the best agreement in the 0 to 90 ppm region, as this region should be unaffected by labeling. The incorporation of the labeled Leu synthesized during the induction period is evidenced by the larger ^{13}C CO intensity for the labeled cell spectrum. Panel b displays the S_1 spectra of the labeled (red) and unlabeled (black) cells. Panel c displays the S_0 (blue) and S_1 (red) spectra processed from the difference NMR data: labeled cells – $0.75 \times$ unlabeled cells. The 0.75 factor reflects the ratio of the number of scans summed for the labeled cells relative to number for the unlabeled cells and resulted in a minimal signal in the 0 to 90 ppm region. The spectra in panel c are representative of the $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu incorporated into the cellular protein. Spectra were processed with no line broadening and a 5^{th} order polynomial baseline correction.....85

Figure 3-5: Deconvolutions are shown for (top) ΔS spectrum of lyophilized cells induced to produce Fgp41 and labeled with $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu, and (bottom) S_0 spectrum from [lyophilized cells induced to produce Fgp41 and labeled with $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu] – $0.75 \times$ [lyophilized cells induced to produce Fgp41 with no label].90

Figure 3-6: Difference spectra are displayed for (top) lyophilized whole cell samples that were induced to produce Fgp41 and labeled with $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu and (bottom) membrane reconstituted purified Fgp41 labeled with $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu. The similarity in line shape and chemical shift of the peak is indicative that Fgp41 is the primary labeled protein present in the lyophilized whole cell sample. The spectra were processed with 100 Hz Gaussian line broadening and a 3^{rd} order polynomial baseline correction.....91

Figure 3-7: S_0 (black) and S_1 (red) spectra for a) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled Fgp41 ICP sample, b) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled Fgp41 ICP spectrum minus $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled pET24a+ ICP spectrum, c) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled Fgp41 ICP spectrum minus unlabeled Fgp41 ICP spectrum. Spectra were processed with no line broadening and a 5^{th} order polynomial baseline correction.....94

Figure 3-8: $\Delta S \equiv S_0 - S_1$ spectra derived from ICP samples. For panel a, both S_0 and S_1 are from the same $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41 sample. For panels b and c, $S_0(S_1)$ is the difference between the individual $S_0(S_1)$ of two different samples: b) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41 sample minus $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu

pET24a+ sample; c) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41 sample minus unlabeled Fgp41 sample. Both the S_0 and the S_1 spectrum of each ICP sample was the sum of 50,000 scans. Spectra were processed with 100 Hz Gaussian line broadening and a 5th order polynomial baseline correction.....96

Figure 3-9: Plot of the integrated signal intensity in the carbonyl region of the ^{13}C spectrum (170 $\rightarrow$ 185 ppm) from 50,000 REDOR S_0 scans vs. the number of moles of label present. The samples measured to create this calibration curve were made of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu manually mixed with talc to create a uniform distribution of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu to fill the 4 mm MAS rotor. The line shown is a linear regression fit with a forced (0,0) intercept. The equation of linear regression is $y=2.12\times 10^8 x$, and $R^2 = 0.985$. The standard error associated with the slope is 1.3×10^7 . Numerical data corresponding to this plot is presented in Table 3-5. S_0 spectra are shown in Figure 3-10.98

Figure 3-10: REDOR S_0 spectra of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ labeled Leu mixed with talc. Blue = 25 mg $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu, pink = 5 mg $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu, and green = 0.5 mg $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu. The spectra are scaled such that the y axis of the spectra containing 0.5 mg : 5 mg : 25 mg of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu were multiplied by 50 : 10 : 1. This was done so that we may assess the linearity of the spectral intensities with respect to the amount of labeled material present. Each spectrum is the result of 50,000 S_0 scans. Spectra are processed with 200 Hz Gaussian line broadening and 5th order polynomial baseline correction.100

Figure 3-11: REDOR S_0 spectra for $1\text{-}^{13}\text{C}$ Leu labeled Human Proinsulin ICP samples. The labeling of each sample is indicated. Each spectrum is the sum of 50,000 S_0 scans. The spectra were processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra are scaled such that the signal in the 0 to 90 ppm region is the same, as this should be unaffected by isotopic labeling.....105

Figure 3-12: a) ^{13}C S_0 REDOR SSNMR spectra of ICP samples labeled with $1\text{-}^{13}\text{C}$ Leu. Each spectrum is the sum of 50000 scans. The spectral intensities are scaled to approximate equal values in the 0 to 90 ppm range. The intensity in this region should be least affected by protein synthesized with $1\text{-}^{13}\text{C}$ Leu in the medium. The spectra are all processed with 200 Hz of Gaussian line broadening and a 3rd order polynomial baseline correction. b) SDS-PAGE gel of insoluble cell pellets after boiling in SDS-containing sample buffer. The molecular weight standards are labeled in the right most lane in kDa and the band attributed to recombinant protein is circled in each sample lane. c,d) Recombinant protein (RP) expression levels calculated from the difference in ^{13}CO signal intensity between the cells with RP and cells without RP. These values were calculated based on analysis of the NMR data shown in panel A,

and the colors correspond. Numerical values from the NMR data can be found in Tables 3-6 and 3-7.106

Figure 3-13: ^{13}C - ^{15}N REDOR ΔS spectra of ICP samples processed without line broadening and with a 5th order polynomial baseline correction. Each ΔS spectrum was the result of 50,000 S_0 scans – 50,000 S_1 scans. The labeling and protein construct is indicated above the spectrum for each sample.110

Figure 4-1: 1- ^{13}C Leu S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from E. coli induced to express human proinsulin labeled with a) 1- ^{13}C Leu and ^{15}N Val, b) 1- ^{13}C Leu and ^{15}N Tyr, c) 1- ^{13}C Leu and ^{15}N Gly, and d) 1- ^{13}C Leu and ^{15}N Ala.....124

Figure 4-2: 1- ^{13}C Leu REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from E. coli induced to express human proinsulin labeled with a) 1- ^{13}C Leu and ^{15}N Val, b) 1- ^{13}C Leu and ^{15}N Tyr, c) 1- ^{13}C Leu and ^{15}N Gly, and d) 1- ^{13}C Leu and ^{15}N Ala.....125

Figure 4-3: 1- ^{13}C Ala S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from E. coli induced to express human proinsulin labeled with a) 1- ^{13}C Ala and ^{15}N Leu, and b) 1- ^{13}C Ala and ^{15}N Gly.....127

Figure 4-4: 1- ^{13}C Ala REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from E. coli induced to express human proinsulin labeled with a) 1- ^{13}C Ala and ^{15}N Leu, and b) 1- ^{13}C Ala and ^{15}N Gly.128

Figure 4-5: 1- ^{13}C Gly S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully

hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Phe, b) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala, and c) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ile.130

Figure 4-6: $1\text{-}^{13}\text{C}$ Gly REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Phe, b) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala, and c) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ile131

Figure 4-7: Deconvolutions of ΔS spectra are displayed for human proinsulin ICP samples labeled with $1\text{-}^{13}\text{C}$ Leu. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.133

Figure 4-8: Deconvolutions of ΔS spectra are displayed for human proinsulin ICP samples labeled with $1\text{-}^{13}\text{C}$ Ala. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.135

Figure A-1: Examination of the solubility of Fgp41:Fragment2_{noCys} under different conditions. The lanes are as follows: 1) proteins soluble in sodium phosphate buffer, 2) insoluble material after sonication in urea, 3) unbound protein in “flow through”, 4) protein eluted with wash buffer, 5) Broad Molecular Weight Standards with important mass markers on the right-hand side of the figure, 6) proteins present in an eluent from the purification of cells containing the empty pET24a+ plasmid as a control, 7) purified Fgp41_{noCys} (as shown in Figure 2-11), 8) protein eluted in 250 mM imidazole containing buffer. The darkest band corresponds to Fgp41:Fragment2_{noCys}.145

Figure A-2: Comparison of Fgp41_{noCys} and Fgp41:Fragment2_{noCys} both purified using urea. The lanes are as follows: 1) Fgp41_{noCys} elution fraction, 2) Spectra Molecular Weight Standards, 3) Fgp41:Fragment2_{noCys} elution fraction, and 4) Fgp41:Fragment2_{noCys} elution fraction. The gel shift due to the molecular weight difference is clearly observed in this gel. The band that corresponds to Fgp41:Fragment2_{noCys} can be seen most clearly in circled in Lane 4.146

Figure A-3: Results of the purification of Fgp41:Fragment2_{noCys} with guanidine HCl as the denaturant. Lane 1) Fgp41 purified with urea, Lane 2) Fgp41 purified with guanidine HCl, Lane 3) Spectra Molecular Weight Standards, 4) concentrated Fgp41:Fragment2_{noCys} elution fractions after dialysis into 8M urea, Lane 5) mixture of Fgp41:Fragment2_{noCys} and Fgp41_{noCys}

after dialysis into 8M urea. The large band between molecular weight markers 19 and 26 kDa can most likely be attributed to the chloramphenicol resistance protein.147

Figure A-4: REDOR S_0 spectra for $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41:Fragment2_{noCys} insoluble cell pellet (left) and $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41_{noCys} insoluble cell pellet (right). The spectra are each the sum of 50,000 REDOR S_0 scans and were both processed with 100 Hz Gaussian line broadening and a 5th order baseline correction. The spectra are scaled so that the intensity in the 0 to 90 ppm region is the same (as this should be unaffected by isotopic labeling and recombinant protein production).....148

Figure B1: ΔS spectra corresponding to labeling at Phe3 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Phe and ^{15}N Gly). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [52996 S_0 – 52996 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [58080 S_0 – 58080 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.158

Figure B2: ΔS spectra corresponding to labeling at Gly4 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [149040 S_0 – 149040 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [167904 S_0 – 167904 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [102928 S_0 – 102928 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.159

Figure B3: ΔS spectra corresponding to labeling at Ala7 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Ala and ^{15}N Gly). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [49328 S_0 – 49328 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [55408 S_0 – 55408 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [96240 S_0 – 96240 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.....160

Figure B4: ΔS spectrum corresponding to labeling at Gly16 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Gly and ^{15}N Met). A) Purified FHA2 protein was combined with a lipid film containing a 4:1

POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [139296 S_0 – 139296 S_1] scans. Spectrum was processed with 200 Hz Gaussian line broadening and 5th order baseline correction.161

Figure B5: ΔS spectra corresponding to labeling at Phe70 of FHA2. (FHA2 was labeled with 1-¹³C Phe and ¹⁵N Ser). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [172679 S_0 – 172679 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [200192 S_0 – 200192 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.161

Figure B6: ΔS spectra corresponding to labeling at Leu98 of FHA2. (FHA2 was labeled with 1-¹³C, ¹⁵N Leu). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [101408 S_0 – 101408 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [111552 S_0 – 111552 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [92800 S_0 – 92800 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.162

Figure B-7: Deconvolutions of ΔS are displayed for select samples of FHA2 in membranes. The position observed in FHA2 as well as the sample conditions are given in the figure. The fitting of each deconvolution is shown on the right, where orange represents the experimental data, green is the best-fit deconvolution sum, and purple is the difference between the two.....163

LIST OF ABBREVIATIONS

A ₂₈₀	absorbance at 280 nm
B ₀	external magnetic field
B ₁	radiofrequency magnetic field
CD	circular dichroism
Chol	cholesterol
CO	carbonyl
CP	cross polarization
CS	chemical shift
CSA	chemical shift anisotropy
Da	dalton
DC	dipolar coupling
E	energy
ΔS	S ₀ – S ₁ ; the filtered REDOR ¹³ C spectrum
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
Fgp41	ectodomain construct of gp41 including the 154 N-terminal amino acids
FHA2	ectodomain construct of HA2 including the 185 N-terminal amino acids
FID	free induction decay
FP	fusion peptide
FWHM	full width at half maximum

gp120	HIV receptor binding protein
gp41	HIV fusion protein
HA1	Influenza receptor binding protein
HA2	Influenza fusion protein
HEPES	N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid
HFP	HIV fusion peptide
HIV	human immunodeficiency virus
HPI	human proinsulin
HPLC	high performance liquid chromatography
$\underline{l}$	spin quantum number
IB	inclusion body
ICP	insoluble cell pellet
IPTG	isopropyl- β -D-1-thiogalactopyranoside
LB	Luria Bertani broth
LUV	large unilamellar vesicles
$\underline{m}$	spin state quantum number
MAS	magic angle spinning
MES	1-(N-morpholino)ethanesulfonic acid
MPER	membrane proximal external region
MW	molecular weight
MWCO	molecular weight cutoff
MWS	molecular weight standards

N70	gp41 fusion peptide and N-terminal helix
N_x	population of state x
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PHI	pre-hairpin intermediate
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-[phospho- <i>rac</i> -(1-glycerol)]
ppm	parts per million
REDOR	rotational echo double resonance
RF	radio frequency
rpm	rotations per minute
S_0	full ^{13}C spectrum from REDOR (without ^{13}C - ^{15}N dipolar interactions)
S_1	attenuated ^{13}C spectrum from REDOR (with ^{13}C - ^{15}N dipolar interactions)
SEDOR	spin echo double resonance
SHB	six helix bundle
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSNMR	solid state nuclear magnetic resonance
t	time

T_1	spin-lattice relaxation time
T_2	transverse relaxation time
T_m	melting temperature
T_r	rotor period
TPPM	two phase pulse modulation
π	180°
γ	gyromagnetic ratio
ω_{rf}	frequency of RF pulse
ω_0	Larmor frequency
ω_R	Rabi frequency
ν_R	Rabi frequency
$\mathbf{i}$	unit vector in the x direction
$\mathbf{k}$	unit vector in the z direction
$\hat{\mu}$	nuclear magnetic moment
$\hbar$	Planck's constant, $1.05 \times 10^{-34} \text{ J} \cdot \text{s} \cdot \text{rad}^{-1}$
$\sigma_{\text{isotropic}}$	isotropic chemical shift tensor
$\sigma_{11}, \sigma_{22}, \sigma_{33}$	principal components of the isotropic chemical shift tensor

Chapter 1 – Introduction

Nuclear Magnetic Resonance

NMR Theory

Nuclear Magnetic Resonance (NMR) spectroscopy investigates the transitions between energy levels of magnetic nuclei within a magnetic field. Not all nuclei are “NMR active”, or able to be probed by NMR spectroscopy. In order for a nucleus to be observed by NMR techniques, it must have a non-zero spin quantum number; the spin quantum number is usually given the designation $\underline{I}^1$. A nucleus that has a non-zero spin associated with it will interact with an applied magnetic field. As a result of the interaction with the magnetic field, the nuclear magnetic moment for each nucleus will precess about the applied magnetic field with a frequency of rotation described in equation 1.1 below

$$\omega_0 = \gamma B_0 \quad (1.1)$$

where γ is the gyromagnetic ratio of the nucleus and B_0 is the magnitude of the applied magnetic field. This frequency of precession is called the Larmor frequency.

¹ Letters that are underlined in Chapter 1 will represent quantum numbers. Letters or symbols that represent vectors will be represented in bold, and letters or symbols that represent quantum mechanical operators will each have a “^” above them.

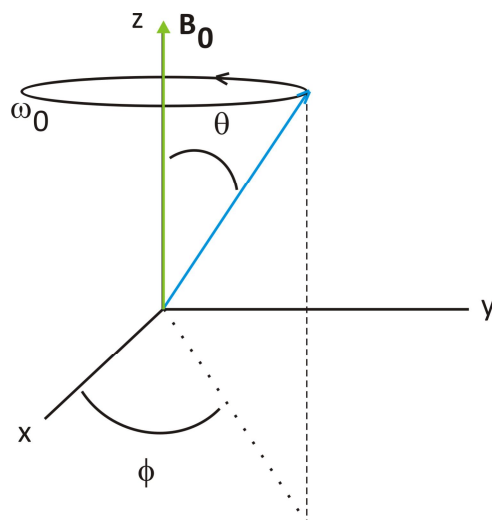


Figure 1-1: Larmor precession of a nucleus in a magnetic field. The static magnetic field $\mathbf{B}_0$ (green) is along the z axis, and thus the nuclear magnetic moment (depicted in blue) rotates around the z axis with a frequency $\omega_0 = \gamma B_0$. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Zeeman Splitting

Common nuclei observed by NMR include ^1H , ^{13}C , ^{15}N , and ^{31}P , all of which have a spin quantum number equal to one half, generally referred to as spin- $\frac{1}{2}$ nuclei. Spin- $\frac{1}{2}$ nuclei exhibit two spin states, calculated by: # of states = $(2I + 1)$. Each individual spin state has a magnetic quantum number, which is given the designation $\underline{m}$, and for a spin- $\frac{1}{2}$ nucleus, are either $\underline{m} = +\frac{1}{2}$ or $\underline{m} = -\frac{1}{2}$. Outside of a magnetic field, these spin states of the nucleus are degenerate in energy, however within a magnetic field the nuclei experience Zeeman splitting, where the energies of the two spin states are no longer degenerate. The magnetic quantum number $\underline{m}$ determines whether the nucleus is in the lower energy spin state (where the nuclear magnetic moment is aligned parallel the static magnetic field) or the higher energy spin state

(where the nuclear magnetic moment is aligned antiparallel to the static magnetic field). The Zeeman Hamiltonian $\hat{H}_{\text{Zeeman}}$ can be expressed in terms of the nuclear magnetic moment $\hat{\mu}$ and the applied magnetic field $\mathbf{B}_0$, which is directed along the z axis, as described in equation 1.2, where $\hat{\mu}$ is defined in terms of the nuclear spin operators $\hat{I}_x, \hat{I}_y$, and $\hat{I}_z$ in equation 1.3. The unit vectors in the x, y, and z directions are represented by $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$, respectively.

$$\hat{H}_{\text{Zeeman}} = -\hat{\mu} \cdot \mathbf{B}_0 = -\gamma \hbar \hat{I}_z B_0 \quad (1.2)$$

$$\hat{\mu} = \gamma \hbar \hat{\mathbf{I}} = \gamma \hbar [\mathbf{i} \hat{I}_x + \mathbf{j} \hat{I}_y + \mathbf{k} \hat{I}_z] \quad (1.3)$$

The associated energies of the different spin states of nuclei in a magnetic field are calculated by obtaining the eigenvalues of the Hamiltonian. The eigenvalue equation for the Zeeman Hamiltonian is displayed in equation 1.4.

$$\hat{H}_{\text{Zeeman}} |\underline{I}, \underline{m}\rangle = E_{\underline{I}, \underline{m}} |\underline{I}, \underline{m}\rangle = -\gamma \hbar B_0 \underline{m} |\underline{I}, \underline{m}\rangle \quad (1.4)$$

From the eigenvalue equation we can easily calculate the energies of the $\underline{m} = +\frac{1}{2}$ and $\underline{m} = -\frac{1}{2}$ eigenstates of the nuclei. Table 1-1 includes numerical data useful for the calculation of the energies of the different eigenstates of the Zeeman Hamiltonian.

Table 1-1: Gyromagnetic ratios and spin quantum numbers for select biologically important nuclei. This table was adapted from reference (1).

Nucleus	Spin (I)	γ ($\text{rad} \cdot \text{s}^{-1} \cdot \text{Tesla}^{-1}$)
^1H	$\frac{1}{2}$	26.7510×10^7
^{13}C	$\frac{1}{2}$	6.7263×10^7
^{15}N	$\frac{1}{2}$	-2.7116×10^7
^{31}P	$\frac{1}{2}$	10.8289×10^7

For ^1H nuclei in a 9.4 Tesla field (all of the NMR experimental data shown in this dissertation were acquired on a 9.4 Tesla spectrometer) we can calculate the energies of the eigenstates using equation 1.4:

$$E_{+1/2} = -\gamma\hbar B_0 \underline{m} = -(26.7510 \times 10^7 \frac{\text{rad}}{\text{s} \cdot \text{Tesla}})(1.05 \times 10^{-34} \frac{\text{J} \cdot \text{s}}{\text{rad}})(9.4 \text{ Tesla})(+1/2) = -1.326 \times 10^{-25} \text{ J}$$

$$E_{-1/2} = -\gamma\hbar B_0 \underline{m} = -(26.7510 \times 10^7 \frac{\text{rad}}{\text{s} \cdot \text{Tesla}})(1.05 \times 10^{-34} \frac{\text{J} \cdot \text{s}}{\text{rad}})(9.4 \text{ Tesla})(-1/2) = +1.326 \times 10^{-25} \text{ J}$$

The difference between the two energy levels is calculated by using equation 1.5.

$$\Delta E = E_{-1/2} - E_{+1/2} = \hbar\omega_0 = \hbar\gamma B_0 \quad (1.5)$$

For ^1H nuclei in a 9.4 Tesla field the difference in energy of the two Zeeman eigenstates is $2.652 \times 10^{-25} \text{ J}$. This is a very small energy gap relative to experimental RT values (RT at 298 K is $\sim 2480 \text{ J/mol}$, or $4.11 \times 10^{-21} \text{ J/nucleus}$), and thus Boltzmann statistics show a very small population difference between the $\underline{m} = +1/2$ and $\underline{m} = -1/2$ eigenstates at thermal equilibrium, with a slightly higher population in the lower energy state. The fractional populations of the two states can be calculated as shown in equation 1.6.

$$\frac{N_{-1/2}}{N_{+1/2}} = e^{(-\Delta E/kT)} = e^{(-2.652 \times 10^{-25} \text{ J} / ((1.38 \times 10^{-23} \text{ J} \cdot \text{K}^{-1})(298 \text{ K}))} = 0.999936 \quad (1.6)$$

From the value of the energy gap between the two states of ^1H nuclei, we can see that absorption of a photon with the energy calculated above ($2.652 \times 10^{-25} \text{ J}$) will cause a transition

from one energy level to the other. For other nuclei, absorption of a photon in the radiofrequency range (2×10^{-30} to 2×10^{-22} J/photon) that meets the resonance condition (i.e. the frequency of the photon is equal to the frequency of the spin) will cause a transition between the two energy levels.

The effect of radiofrequency (RF) pulses

By applying radiofrequency (RF) pulses to the system, an oscillating magnetic field is introduced. The magnetic field is time dependent and is generally denoted $\mathbf{B}_1$. If we consider the case where $\mathbf{B}_1$ oscillates along the x axis, the total field experienced by the nucleus is described in equation 1.7.

$$\mathbf{B}(t) = iB_1 \cos(\omega_{rf}t) + \mathbf{k}B_0 \quad (1.7)$$

In equation 1.7, $\mathbf{B}(t)$ is the total field experienced by the nucleus, i is a unit vector in the x direction, ω_{rf} is the angular frequency of the RF pulse, and $\mathbf{k}$ is a unit vector in the z direction. Previously, we had considered that a nucleus could either be in the higher or lower energy spin states. In the presence of RF pulses, a time dependence of the state of the system is introduced. The new, time dependent Hamiltonian when considering the addition of RF pulses to the system is displayed in equation 1.8, where $\hat{H}_{rf}$ is the Hamiltonian in the presence of RF pulses.

$$\hat{H}_{rf} = -\gamma\hbar(\hat{\mathbf{I}}_x B_1 \cos(\omega_{rf}t) + \hat{\mathbf{I}}_z B_0) \quad (1.8)$$

The spin state of the system in the presence of RF radiation can be described as a time-dependent linear combination of both spin states, displayed in equation 1.9.

$$|\underline{1}, \underline{m}(t)\rangle = \cos\left(\frac{1}{2}\omega_R t\right) \left|\frac{1}{2}, +\frac{1}{2}\right\rangle + i \sin\left(\frac{1}{2}\omega_R t\right) \left|\frac{1}{2}, -\frac{1}{2}\right\rangle \quad (1.9)$$

$$\omega_R = \frac{\gamma B_1}{\hbar} \quad (1.10)$$

Equation 1.10 displays the calculation of the Rabi frequency, which describes the frequency at which transitions between the energy states are occurring under B_1 radiation. A more useful form of the Rabi frequency is expressed in equation 1.11.

$$\nu_R = \omega_R / 2\pi = \gamma B_1 / h \quad (1.11)$$

In equations 1.10 and 1.11, B_1 is the strength of the RF field and h is Planck's constant. We may calculate the Rabi frequency experienced by a proton from the length of the RF pulse applied. For a 5 μ s pulse, the corresponding Rabi frequency for a ^1H nucleus can be calculated as shown in equation 1.12. The calculated frequency (50,000 Hz) tells us the rate at which a nucleus cycles through a spin-up orientation, to spin-down, and back again.

$$\nu_R = \frac{1}{4 \times (90^\circ \text{ pulse length})} = \frac{1}{4 \times 5\mu\text{s}} = 50,000\text{Hz} \quad (1.12)$$

Biomolecular NMR and sensitivity

In biological molecules, much of the NMR spectroscopy performed is aimed at gaining structural knowledge from studying protein backbone carbon and nitrogen atoms. The chemical shift of these nuclei is quite sensitive to the dihedral angles of the peptide bond planes, which change depending on the secondary structure of the protein. However, for rare spin- $\frac{1}{2}$ nuclei

such as these, (^{13}C is only 1.1% naturally abundant and ^{15}N is only 0.37% naturally abundant)

methods can be utilized to increase the sensitivity of the NMR experiment to these dilute spins.

Isotopic enrichment

One method that is often utilized to increase NMR experiment sensitivity for biomolecules is isotopic enrichment of rare nuclei. For synthetic peptide production, it is quite straightforward to include commercially available isotopically labeled amino acids into the synthesis reaction and this is done quite routinely(2-4). Amino acids are commercially available with many different labeling schemes, including ^{15}N or ^{13}CO backbone labeling, uniform labeling (where all carbons are ^{13}C labeled) and various $^{13}\text{C}_\alpha$ or side chain labels. For recombinant protein expression in bacteria, several methods are available to label the proteins. Growth of the bacteria in a minimal medium with controlled carbon and nitrogen sources allow for the *E. coli* to incorporate the selected labels into the proteins as they are being synthesized(5). Supplementing the growth medium of *E. coli* with labeled amino acids allows for labeling at specific positions, while ^{13}C labeled glucose/glycerol or $^{15}\text{NH}_4\text{Cl}$ allow for uniform labeling of the entire protein(6). By incorporating isotopic labels, one can increase the signal obtained per molecule of protein that is present and obtain structural information about the biomolecule in question.

Cross Polarization (CP)

Another method of increasing sensitivity in biomolecular NMR is to take advantage of the properties of abundant nuclei. The most useful abundant spin $\frac{1}{2}$ nucleus may be ^1H ,

because of its large gyromagnetic ratio and large natural abundance (over 99.9% of all hydrogen is ^1H). In a cross polarization (CP) experiment, magnetization is transferred from the abundant ^1H to the more dilute ^{13}C nuclei, allowing for a larger population difference in ^{13}C spin states. Several steps are required for this transfer. First, a 90° pulse is used to rotate the net ^1H magnetization into the transverse plane. The magnetization of one type of nucleus can be described in terms of the nuclear magnetic moments as shown in equation 1.13.

$$\mathbf{M} = \sum_i \hat{\boldsymbol{\mu}}_i \quad (1.13)$$

If the initial net magnetization $\mathbf{M}$ of ^1H is along the z axis (as a result of the static magnetic field $\mathbf{B}_0$), and the $\mathbf{B}_1$ field is along the x axis, a 90° pulse will rotate the net magnetization into the transverse plane along the y axis according to equation 1.14, where “ $\times$ ” denotes the cross product between the two vectors and $\mathbf{M}(t)$ is the magnetization at time t. The cross product of two vectors produces a vector perpendicular to the two initial vectors with its direction determined by using the right hand rule (start with fingers pointing up (along the z axis) and curl towards the x axis, thumb will point in the direction of the net magnetization immediately following application of the 90° pulse, i.e. along the y axis); in this case the resulting vector ($\mathbf{M}$ after the pulse is applied) will be along the y axis.

$$\frac{d\mathbf{M}}{dt} = \mathbf{M}(t) \times \mathbf{B}_1 \quad (1.14)$$

At this point, a spin-locking field (contact pulse on the y axis) is applied to ^1H , where a constant amplitude irradiates the sample to maintain the ^1H magnetization along the y axis. In the presence of a constantly applied $\mathbf{B}_1$ field, the effect of $\mathbf{B}_0$ on the sample is null. The ^1H magnetization is transferred to ^{13}C via dipolar coupling. This is achieved through the Hartmann-Hahn matching condition, where the energy of a photon emitted from ^1H can be absorbed by ^{13}C and vice versa because the gap between the upper and lower spin states for ^1H and ^{13}C is equal due to the set amplitudes of $\mathbf{B}_1$ radiation (as the $\mathbf{B}_0$ field does not affect the nuclei under constant $\mathbf{B}_1$ radiation. This is described in equation 1.15, where $B_1(\text{H})$ and $B_1(\text{C})$ are the RF field strengths applied to ^1H and ^{13}C nuclei.

$$\gamma_{\text{H}}B_1(\text{H})=\gamma_{\text{C}}B_1(\text{C}) \quad (1.15)$$

Experimentally, we achieve this condition by simultaneously irradiating ^1H and ^{13}C nuclei. The ^1H frequency is irradiated with constant amplitude, and the ^{13}C frequency is irradiated with ramped amplitude to achieve the greatest amount of transfer by meeting the matching condition for as many nuclei as possible. As the chemical shielding (and therefore energy) of nuclei of the same isotope differ within a magnetic field, there are a variety of matching conditions to meet for these different nuclei. After the transfer of magnetization, high power decoupling is applied to the ^1H channel to prevent recoupling of ^1H nuclei to the ^{13}C nuclei,

subsequent increased T_2 relaxation, and the associated line-broadening that is not desirable in NMR spectra.

There are two distinct advantages to employing cross-polarizations in NMR experiments. The first is the increase in sensitivity of the experiment which was already discussed. The second advantage to employing a cross-polarization from ^1H to ^{13}C is that the recycle delay between pulse trains can be much shorter than for an experiment without cross-polarization. This is due to the much faster spin-lattice (T_1) relaxation times of ^1H nuclei compared to ^{13}C nuclei. Spin-lattice relaxation can occur as a result of dipolar couplings; homonuclear dipolar couplings between ^1H nuclei are strong due to the large abundance of ^1H nuclei in samples (which leads to ^1H nuclei pairs with smaller internuclear distances) as well as the large gyromagnetic ratio of these nuclei. The faster T_1 relaxation rates of ^1H allow for more spectra to be acquired during the same amount of time if CP is used and thus requires less overall signal averaging time to achieve the same experimental signal to noise ratios.

Solid State Nuclear Magnetic Resonance (SSNMR)

High resolution liquid state NMR is dependent on molecules tumbling rapidly enough in solution to average out anisotropic contributions to spectra. For biomolecular samples such as proteins or peptides embedded in lipid membranes, the tumbling is too slow to average out orientation dependent effects on the spectra such as dipolar coupling (DC) and chemical shift anisotropy (CSA) with time.

Magic Angle Spinning (MAS) NMR

Magic Angle Spinning (MAS) is a technique used to increase resolution in NMR spectra of solid and semi-solid samples. Contributions to line broadening and therefore loss of resolution in solid-state NMR spectra come from chemical shift anisotropy and dipolar coupling contributions, both of which can be resolved by fast rotation of the sample at the magic angle. The “magic angle” is the angle θ which satisfies the equation 1.16. This particular angle can be defined as $\theta_{MA} = 54.7^\circ$.

$$(3\cos^2\theta - 1) = 0 \quad (1.16)$$

Both CSA and DC exhibit a proportionality to a term $(3\cos^2\theta - 1)$, where θ is the angle between an internuclear vector and the external magnetic field. When this expression is equal to zero, CSA and DC contributions to a spectrum are removed.

Consider Figure 1-2, where a sample is spun about the “rotor axis”. The rotor axis in this case will be at the magic angle. A single C-N internuclear vector is pictured, and this vector can be broken down into a sum of two vector components. One component will be considered to be aligned with the rotor axis, (with angle to the external magnetic field θ_{MA}) and the other component will be perpendicular to the rotor axis (with angle $\theta_{\perp MA}$). Over one rotation of the sample about the rotor axis (the time it takes for this rotation will be termed T_r , a rotor period), the magnitude of the vector component along the rotor axis remains unchanged, while the perpendicular vector will be averaged to zero. This, with spinning about the rotor axis, we can approximate that each internuclear vector (regardless of its orientation with respect to the

external magnetic field) will be reduced to its contribution along the axis of sample rotation, as long as integer multiples of rotor periods are used in an experiment.

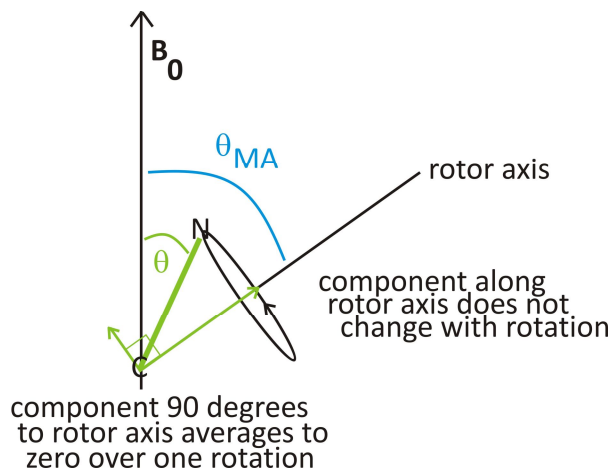


Figure 1-2: Depiction of the breakdown of angles in MAS experiments. The C-N internuclear vector at angle θ to the external magnetic field (depicted in green) can be broken into two components. One component is a vector along the axis of rotation (angle $\theta_{MA}=54.7^\circ$ to the external magnetic field). The other component is 90° to the axis of rotation. If we consider one rotor period, the contribution along the rotor axis will remain unchanged, and the contribution perpendicular to the axis of rotation will average to zero. This is only shown for one internuclear vector direction, but is true for an internuclear vector in any orientation.

Dipolar Coupling (DC)

The heteronuclear dipolar coupling Hamiltonian describing interactions between two nuclei can be expressed as shown in equation 1.17, where μ_0 is the permeability constant, γ_I and γ_S are the gyromagnetic ratios for nuclei I and S, r is the internuclear distance between nuclei I and S, $\hat{I}_z$ and $\hat{S}_z$ are the spin operators for nuclei I and S, and θ is the angle between the external magnetic field and the internuclear vector.

$$\hat{H}_{\text{DipolarCoupling}} = -\left(\frac{\mu_0 \hbar}{4\pi}\right) \frac{\gamma_I \gamma_S}{(r)^3} \hat{I}_z \hat{S}_z (3\cos^2 \theta - 1) \quad (1.17)$$

The largest value of dipolar coupling between two nuclei will be observed when the internuclear vector is either parallel or antiparallel to the magnetic field. Zero dipolar coupling will be observed in instances where the angle between the internuclear bond vector and the magnetic field satisfies equation 1.16. In macroscopic samples there are large ensembles of spins, with internuclear vectors oriented in many different directions, and it is quite unlikely that many of the internuclear vectors in the samples will align with the magic angle. However, by spinning the sample about the magic angle, we can reduce the internuclear vectors to their vector contributions along the magic angle, as discussed previously, which will average dipolar couplings in the sample to zero over each rotor period.

Chemical shift anisotropy (CSA)

The chemical shift observed for a nucleus in an NMR spectrum is dependent on its chemical shielding σ , which affects what magnitude of the applied magnetic is experienced by the nucleus according to equation 1.18.

$$\mathbf{B}_{\text{total}} = \mathbf{B}_0(1 - \sigma) \quad (1.18)$$

From equation 1.18 we can see that a nucleus with more shielding will experience a smaller total magnetic field. Chemical shielding of a nucleus arises from interaction with the electronic fields of nearby electrons (such as in bonds) which are the result of $\mathbf{B}_0$ induced electronic currents. In an external magnetic field, electrons have an induced electronic magnetic dipole moment which lies antiparallel to $\mathbf{B}_0$ and will decrease the magnitude of $\mathbf{B}_0$ experienced by the nucleus. As most nuclei (especially those in biomolecular samples) are not in a completely

symmetric chemical (i.e. different bonding) environment, the orientation of a molecule within the external magnetic field will affect the extent of chemical shielding and therefore the magnitude of the external magnetic field that is experienced by the nucleus. The chemical shift δ of a nucleus is defined in equation 1.19, where γ is the gyromagnetic ratio of the nucleus, B_0 is the strength of the external magnetic field, σ_{ref} is the chemical shielding of a reference compound, σ is the shielding of the nucleus, and ν_{RF} is the frequency of the spectrometer.

$$\delta = \left(\frac{\gamma}{2\pi} \right) B_0 \left(\frac{\sigma_{\text{ref}} - \sigma}{\nu_{\text{RF}}} \right) \quad (1.19)$$

The chemical shift δ can be expressed in terms to show the orientation dependence and this is shown in equation 1.20, where $\delta_{11}, \delta_{22}, \delta_{33}$ are the principle values and $\theta_{11}, \theta_{22}, \theta_{33}$ are the angles between the principle values axes and the external magnetic field $\mathbf{B}_0$, as defined in Figure 1-3(7).

$$\delta = \delta_{11} \cos^2 \theta_{11} + \delta_{22} \cos^2 \theta_{22} + \delta_{33} \cos^2 \theta_{33} \quad (1.20)$$

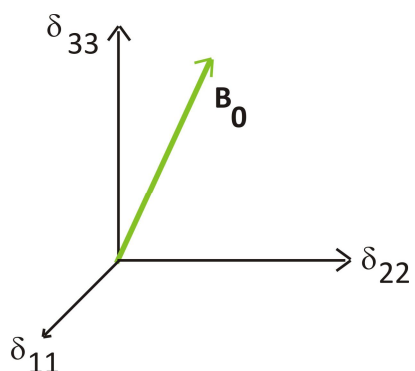


Figure 1-3: Depiction of the principle axes $\delta_{11}, \delta_{22}, \delta_{33}$ with respect the external magnetic field $\mathbf{B}_0$. The angles $\theta_{11}, \theta_{22}, \theta_{33}$ are the angles between the axes and $\mathbf{B}_0$.

The relationship between $\delta_{11}, \delta_{22}, \delta_{33}$ (which are the values of the three principal components of the isotropic chemical shift tensor) and $\delta_{\text{isotropic}}$ (which is observed for molecules in solution when rapid tumbling averages the shielding over time) is displayed in equation 1.21.

$$\delta_{\text{isotropic}} = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33}) \quad (1.21)$$

By spinning a sample at the magic angle at high frequencies, only the isotropic chemical shift will be observed. If the sample is not spun fast enough about the magic angle, then peaks will appear at integral multiples of the spinning frequency in the spectrum, centered around the isotropic chemical shift. These peaks are called spinning sidebands, and can be attenuated by spinning faster.

Rotational Echo Double Resonance (REDOR) NMR

The REDOR pulse sequence was developed in the lab of Jacob Schaefer in the late 1980s with the goal of measuring dipolar couplings between nuclei in solid samples to extract information such as internuclear distances (8). REDOR was modeled after the SEDOR (spin-echo double-resonance) NMR experiment which is performed on a static solid-state sample to measure dipolar couplings between nuclei(9).

REDOR is a SSNMR magic angle spinning experiment. Both homonuclear and heteronuclear dipolar couplings between nuclei are averaged to zero over each rotor period when a sample is subjected to rapid spinning at 54.7° (these are averaged more quickly as the sample is spun with a greater spinning rate) as explained by equation 1.17 and Figure 1-2. The

REDOR pulse sequence utilizes rotor synchronized pulses to selectively reintroduce heteronuclear dipolar couplings between nuclei. In every REDOR experiment, two types of spectra are acquired. The first can be thought of as a reference spectrum with all dipolar couplings removed, and the second spectrum includes some contributions from dipolar coupling.

The REDOR S_0 experiment

In the first experiment, generally referred to as the S_0 experiment, following CP from ^1H to ^{13}C , high power decoupling is applied to ^1H during the remainder of the experiment, and a π pulse is applied on the ^{13}C channel at the end of every rotor period except the last of the sequence. The π pulses serve to refocus the magnetization that has been dephased due to differences in the isotropic chemical shifts of the nuclei. The spectrum is acquired immediately at the end of the last rotor period of the pulse sequence. The spectrum acquired for this experiment corresponds to the signal from all ^{13}C present in the sample.

The REDOR S_1 experiment

The second experiment that is performed in REDOR is termed the S_1 experiment, and contains a second set of pulses. The CP and pulses on the ^1H and ^{13}C channel are exactly the same as during the S_0 experiment. During the S_1 experiment, π pulses are applied on the ^{15}N channel halfway through each rotor period to reintroduce dipolar coupling between ^{13}C and nearby ^{15}N nuclei. This causes the local field felt by nearby ^{13}C nuclei to “flip”, and the nuclei

begin to precess in the opposite direction. Due to nuclei experiencing different local fields, the rate of precession and thus the evolution angle is different between these nuclei. There is a net loss of magnetization due to the different rates of precession, termed dephasing. The longer the spins are allowed to precess, the more net dephasing will be observed in the acquired spectrum.

In terms of the dipolar coupling Hamiltonian, we can simplify the expression to see that it is an interaction of the nuclear magnetic moment $\hat{\mu}$ (for ^{13}C , in my example) with the local field $\mathbf{B}_{\text{local}}$ induced by ^{15}N nuclei. This is presented in equation 1.22.

$$\hat{H}_{\text{DipolarCoupling}} = -\hat{\mu} \cdot \mathbf{B}_{\text{local}} \quad (1.22)$$

The local field induced by ^{15}N nuclei is modulated over each rotor period in the absence of ^{15}N π pulses. Thus, over each rotor period in the S_0 experiment, the local dipolar field averages to zero as I discussed earlier. During the S_1 experiment, the direction of the dipolar field due to the ^{15}N nuclei is changed by ^{15}N π pulses halfway through each rotor period. This results in a net positive dipolar field during the first rotor period, and a net negative dipolar field during the second rotor period, and so on for even and odd rotor periods. The direction (or sign) of the nuclear magnetic moment vector changes with the application of ^{13}C π pulses at the end of each rotor period.

By combining the ideas regarding how the local field due to ^{15}N nuclei and the direction of the nuclear magnetic moment vector change during the experiments, we can gain an

understanding of how the dipolar coupling energy changes during the experiments. This is portrayed simplistically in Figure 1-4.

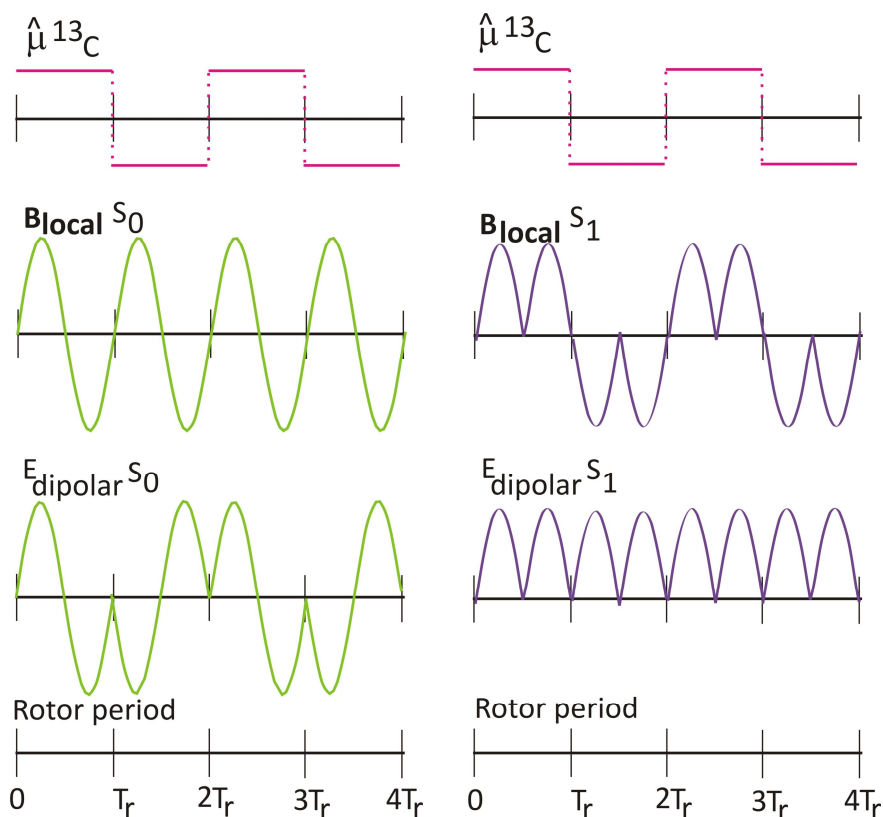


Figure 1-4: Simplified model of how the ^{13}C nuclear magnetic moment vector, local field induced by ^{15}N nuclei onto ^{13}C nuclei, and the dipolar coupling energy evolve with time under Magic Angle Spinning conditions in REDOR. The dipolar interaction energy is averaged to zero over each rotor period as shown for the S_0 experiment. As a result of ^{13}C and ^{15}N π pulses, the dipolar interaction energy during the S_1 experiment is nonzero when an average is taken over rotor periods.

By comparing spectra in which there is no net dipolar coupling interaction observed (S_0) and spectra for which dipolar coupling has been reintroduced (S_1) we can directly see the effect of the dipolar coupling on the spectra, and this is observed as a decrease in signal when there is

dipolar coupling present. This is generally referred to as dephasing, and is often expressed as a percentage of the S_0 signal. The percentage dephasing is calculated by equation 1.23.

$$\% \text{ dephasing} = \frac{S_0 - S_1}{S_0} \times 100 \quad (1.23)$$

Applications of REDOR NMR

Since the REDOR pulse sequence was introduced in 1989, it has been applied to many different systems. Initial proof-of-concept experiments were performed on ^{13}C and ^{15}N labeled alanine and mixtures of the molecules. These experiments showed (when measuring dipolar coupling between ^{13}C alanine co-crystallized with ^{15}N alanine) that intermolecular C-N distances of 4 – 6 angstroms could be determined using the method(8). A binding site – ligand interaction was characterized using a combination of ^{13}C - ^{15}N REDOR between ^{13}C labeled glutamine and ^{15}N labeled His156 of the *E. coli* Glutamine-Binding Protein and molecular dynamics simulations(10). ^{13}C , ^{15}N labeled Acetyl-L-carnitine was investigated using ^{13}C - ^{15}N REDOR to calculate the internuclear distance between the two nuclei to determine whether Acetyl-L-carnitine was in an extended or folded structure in the solid state. The dipolar couplings measured from these experiments indicated that the ^{13}C – ^{15}N internuclear distance was much longer (5.05×10^{-10} m between C(1) and N for the REDOR method compared to previous X-ray crystallography results of 4.24×10^{-10} m between C(1) and N) than reported

previously. These results were confirmed by subsequent crystallization and X-ray crystallography of Acetyl-L-carnitine.

^{13}C - ^{15}N REDOR was used to investigate oligomeric assemblies in the HIV fusion peptide (HFP) in membranes through a combination of dipolar coupling measurements and modeling of experimental results using SIMPSON. The results of this body of work suggest that HFP has multiple populations of antiparallel β -sheet registries, in contrast to earlier work that suggested that HFP assembled as in-register parallel β -sheets(3).

By using a short, fixed dephasing period, ^{13}C - ^{15}N REDOR can effectively be used to filter spectra to obtain structural information about one or multiple residues of large proteins. Earlier work in the Weliky group utilized this method to examine the secondary structure of the Influenza fusion protein HA2 in the context of lipid membranes. Portions of the protein had been studied by crystallography previously and FP structural studies had been performed in micelles and lipid bilayers(11-13). The first atomic resolution structural data for the HA2 fusion peptide in lipid bilayers (in the context of the full ectodomain of the protein) was obtained by ^{13}C - ^{15}N REDOR. By selective isotopic labeling, spectra were obtained that yielded structural information for individual amino acids in the protein(14).

REDOR can be utilized on nuclei other than ^{13}C - ^{15}N as well. Work in the Weliky group has included ^{13}C - ^{31}P REDOR to simultaneously probe the secondary structure of regions of the HIV fusion peptide (HFP) and proximity to the headgroups of phospholipid bilayers(15, 16). Results of these studies indicated that in membranes containing cholesterol, HFP retains

primarily β -sheet structure and that Ala5 (in the mid-FP region) does not interact with lipid headgroups while Ala15 (near the end of the FP) does interact with the lipid headgroups. In SSNMR samples prepared without cholesterol, the Ala5 residue of a monomer of HFP has closer contact with the lipid headgroups than Ala5 in cross-linked dimer or trimer molecules of HFP. In the samples without cholesterol, Ala16 showed similar dephasing curves between monomer, dimer, and trimer molecules. ^{13}C - ^{19}F REDOR was used on a similar system to investigate the correlation between insertion depth of HFP into ^{19}F labeled lipid bilayers and fusogenicity, finding that HFP constructs that inserted more deeply into the lipid bilayers (shown by experimental dephasing of ^{13}C on the HFP by ^{19}F incorporated at the end of the acyl chain of the lipid molecules) with higher fusogenic activity (17).

My dissertation work involved using ^{13}C - ^{15}N REDOR with a set dephasing time (either 1 or 2 ms, depending on the sample) to investigate the secondary structure of proteins in either lipid bilayer samples, whole *E. coli* cells, or inclusion body samples. The first project presented includes atomic resolution structural studies of Fgp41 (an ectodomain construct of the HIV fusion protein gp41) embedded in lipid bilayers using REDOR, and development of protein production and purification protocols. Part of the work presented in this dissertation involves the application of REDOR SSNMR to quantitatively determine the level of recombinant protein expression measured in mg of protein produced / L of bacterial expression in either whole *E. coli* cell samples or insoluble cell pellets (primarily composed of bacterial inclusion bodies). A short, final project involved structural studies of human proinsulin within bacterial inclusion bodies.

Human Immunodeficiency Virus (HIV) Fusion Protein gp41

The human immunodeficiency virus (HIV) is enveloped by a membrane obtained during budding from an infected host cell. An early step in HIV infection of a new cell is joining or “fusion” of the HIV and host cell membranes. This process is catalyzed by the ~350-residue HIV gp41 protein which is an integral membrane protein of the viral envelope(18). The ~175 N-terminal residues form the ectodomain which lies outside HIV; the ectodomain of gp41 is pictured below in Figure 1-5.

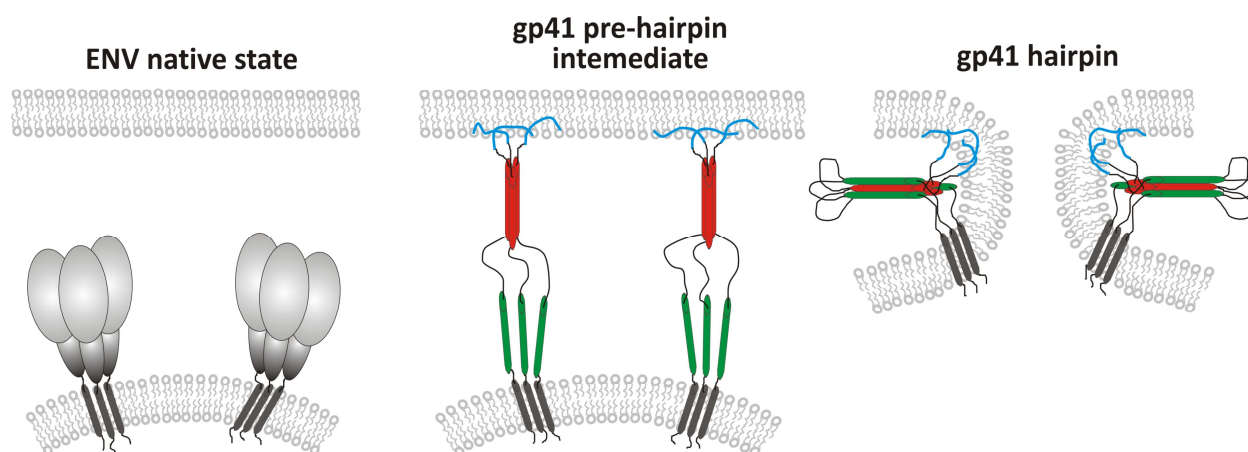


Figure 1-5: Conceptual representation of gp41 structural states with time increasing from left to right. In the middle and right panels, the region shown in blue represents the fusion peptide, red represents the C-terminal helix, and green represents the N-terminal helix. In the right panel of the figure, the red and green helices are antiparallel to one another.

Prior to fusion, gp41 is non-covalently associated with the gp120 protein. Productive infection begins with binding of gp120 to receptor proteins in a target cell membrane and is followed by gp120 dissociation from gp41(19). There are ensuing structural changes of gp41 and likely binding of the ~20-residue N-terminal “fusion peptide” (FP) region to target cell membranes with concurrent changes in the two membranes including mixing of lipids,

formation of a single hemifusion diaphragm bilayer that separates the HIV and cell contents; and opening of the diaphragm to form a single membrane that encloses HIV and the cell(20). Although there are no high-resolution structures of full-length gp41, other structural and functional data support: (1) trimeric gp41; (2) an early-stage “pre-Hairpin intermediate” (PHI) state with a parallel trimer of fully extended ectodomains between the HIV membrane and the FP in the cell membrane; and (3) a final “six-helix bundle” (SHB) state with a gp41 trimer with each gp41 molecule having a N-helix-turn-C-helix Hairpin structure and parallel N-helices in the trimer interior and parallel C-helices on the trimer exterior, as depicted in Figure 1-6 (21-24). Studies of cell-cell fusion induced by gp120/gp41 complexes indicate that most membrane fusion steps with the exception of diaphragm opening occur prior to formation of the final SHB state (25).

gp41 fusion peptide (FP)

The importance of the FP in fusion and infection has been highlighted by reduction in both functions with point mutations in the FP (26). Current understanding of gp41 is also based on smaller fragments of gp41 where fusogenic function has typically been assayed by fragment-induced perturbation/fusion of membrane vesicles. One such fragment is the HIV fusion peptide (HFP) which corresponds to the 20-30 N-terminal residues of gp41 and which has moderate fusogenicity (27). The functional significance of the PHI trimeric topology has been supported by high fusogenicity of: (1) a cross-linked HFP trimer (HFPtr); and (2) “N70”, the 70 N-terminal residues of gp41 (4, 27-29). The higher fusogenicity of N70 relative to HFP may also have a contribution from the N-helix residues that are C-terminal of the FP. Much larger ectodomain constructs have also been produced with N-helix and C-helix regions and form the

thermostable SHB structure, Figure 1-6, which is the final gp41 state. Different approaches were used to obtain these FP-containing Hairpin constructs. In one approach, the FP and Hairpin regions were produced separately by chemical synthesis and bacterial expression, respectively, and “FP-Hairpin” was then made by native chemical ligation (4, 28). In another approach, a chimera was expressed in *E. coli* bacteria and contained a N-terminal molecular carrier protein (e.g. glutathione S-transferase) followed by the gp41 ectodomain (30-32). The carrier was cleaved during purification. There are conflicting results from different studies of the fusogenicity of such Hairpin constructs with reports of both very high and no fusion. There were some differences among the studies including: (1) deletion of loop residues between the N- and C-helices in some constructs; (2) lipid compositions of the vesicles including different fractions of negatively charged lipid; (3) use of smaller and less stable sonicated vesicles vs larger and more stable extruded vesicles; and (4) pHs that ranged between 3.0 and 7.5 (33).

Structural studies have also been carried out for some of the aforementioned fragments. A helical monomer HFP has been observed in detergent with one report of a continuous helix between residues 4 and 22 (34-38). However, to our knowledge, HFP does not induce fusion between detergent micelles. The structure of membrane-associated HFP has been probed mostly by solid-state nuclear magnetic resonance (SSNMR) spectroscopy with supporting data from other techniques such as infrared spectroscopy (39-41). SSNMR spectra of HFP associated with membranes lacking cholesterol show distinct populations of predominant β sheet and predominant α helical molecules while HFP associated with membranes with ~30 mol% cholesterol show only the β sheet conformation with antiparallel alignment of adjacent hydrogen bonded HFPs (3, 16, 27, 42-44). The biological relevance of membrane cholesterol is

supported by the ~25 mol% cholesterol in host cell membranes and the ~45 mol% in HIV membranes (45). The FP structure appears to be similar in the highly fusogenic HFP trimer and in N70 whereas the FP-Hairpin construct with SHB structure showed approximately equal populations of molecules with either β sheet or helical FP structures (4, 16, 27). For membrane-associated N70, the N-helix residues appear to be predominantly helical and N70 is recognized by an antibody specific for trimeric coiled-coil N-helices (46). There are several high-resolution structures of Hairpin constructs without FP which show: (1) Hairpin structure of individual molecules; and (2) molecular trimers with SHB structure (21-24).

Fgp41 – an ectodomain construct of gp41

SSNMR requires production of multi-mg quantities of isotopically labeled protein and protein yields may be reduced by ligation and/or cleavage steps. This motivated one of the goals of the study presented in Chapter 2 – expression of the FP-containing gp41 ectodomain (“Fgp41”) in bacteria without a chimera or ligation. This goal seemed reasonable because recently developed protocols yielded 20 mg protein/L culture for the full-length “FHA2” ectodomain of the influenza virus fusion protein (14, 47, 48). There is considerable diversity among HIV protein sequences in patient sera and in cell cultures. This motivated a second goal of the present study – functional and structural experiments on a gp41 ectodomain sequence that differed from the sequence of the earlier studies to address the generality of the functional and structural findings across strains of HIV. In these earlier studies, the sequence was from the HXB2 strain of HIV-1 which was first created in cell culture in 1984 and which is grouped with “clade B” HIV-1 prevalent in patients in North America and Europe (49). In some contrast, the gp41 sequence of the present study is from the primary HIV-1 isolate Q45D5 from the sera of a

newly infected Kenyan woman (50). The Q45D5 isolate is grouped with clade A HIV-1 that is prevalent in Central and East Africa. The HXB2 and Q45D5 gp41 ectodomain sequences are compared in Figure 2-9. A third motivation for the present study was to provide comparative functional and SSNMR structural studies to the FP-Hairpin construct in which 46 contiguous residues including the native loop were replaced by the non-native SGGRGG sequence (4, 28, 33). FP-Hairpin did not induce vesicle fusion and inhibited fusion by constructs such as N70. The deletion of these residues in FP-Hairpin may be important as a 35-residue peptide which included the native loop region induced vesicle fusion under some conditions (51). Comparisons between the Fgp41 construct which includes the native loop and FP-Hairpin are discussed in detail in Chapter 2.

Bacterial Inclusion Bodies

Recombinant protein production within bacteria such as *E. coli* has become a standard way to produce proteins for further study. *E. coli* are attractive hosts to utilize for protein production for many reasons, including their relatively simple genome, ease in maintaining cultures, and high heterologous protein expression levels – often producing multi-mg quantities of recombinant protein per liter of fermentation culture.

One aspect of protein production in *E. coli* that has been considered a drawback is that overexpression of recombinant proteins often leads to the production of inclusion bodies, or insoluble aggregates of protein. Inclusion bodies have been described as amorphous aggregates which are spherical in shape as observed by transmission electron microscopy(52). It is unknown exactly what factors cause proteins to form inclusion bodies when expressed in *E. coli*,

but factors such as hydrophobicity of the protein, size, growth medium conditions, and promoter systems have all been implicated as possible causes. A review of older literature describes many different eukaryotic proteins expressed in *E. coli* K-12 strain under different conditions, where no clear pattern was observed to trigger the formation of inclusion bodies (53). It has been suggested more recently that the use of minimal media for *E. coli* growth (as is needed for isotopic labeling often used in NMR sample preparations) can cause proteins to be more likely to form inclusion bodies, perhaps due to the difference in cell environment in the different growth media (54).

Utilization of inclusion bodies

There are positive aspects to inclusion body formation during recombinant protein expression. When human insulin was expressed as A and B chains separately in *E. coli*, both chains were found in the insoluble portion of the cell lysate(55). In the case of human insulin A and B chains, the sequestration of the polypeptides into inclusion bodies was utilized as part of the purification procedure, allowing the researchers to discard the soluble proteins present in the *E. coli* cell lysate. Previous work in the Weliky group utilized inclusion bodies to increase the yield of isotopically labeled viral membrane protein FHA2 from ~5 mg per liter of bacterial cell culture to ~20 mg per liter of culture by solubilizing and refolding the protein within inclusion bodies (47). Additionally, FT-IR spectroscopy has been proposed as a method to quantify recombinant protein in inclusion bodies within intact cells by observing a shift in the amide I band toward the β sheet region of the spectrum(56). The previous work assumes knowledge that proteins within inclusion bodies have primarily β sheet structure, though work in our group

has shown that some proteins retain native α -helical structure within bacterial inclusion bodies(57, 58).

Quantitative detection of protein in inclusion bodies

We have developed a SSNMR method to detect recombinant protein expression levels within *E. coli* by utilizing inclusion bodies. The method requires small sample volumes, 20 – 40 mg of isotopically labeled amino acids per sample, moderate NMR fields (9.4 Tesla), and is quick and straightforward. In addition, we have applied the method to a variety of proteins in different plasmid types and *E. coli* strains, including proteins with native α -helical structure. This work is discussed in Chapter 3.

Diabetes and the prehoromone human proinsulin

Diabetes is a disease caused by either a lack of insulin production by the pancreas (Type I Diabetes) or ineffective processing of insulin (Type II Diabetes). The Centers for Disease Control and Prevention reports that Diabetes affects 8.3% of the American population, or 25.8 million people (59). Of these 25.8 million people, 26% are treated with insulin therapy, where a suspension of insulin is administered to the patient via injection.

Synthetic production of insulin

Due to the large demand for human insulin and the cost-effectiveness of bacterial expression of eukaryotic proteins, insulin has been produced via *E. coli* in several different ways. One method included separate expression of the A and B chains of insulin (55). Another method of production of human insulin includes expression of the prehormone proinsulin in *E.*

coli. After purification and refolding of proinsulin, insulin can be obtained after enzymatic cleavage of the C-chain (60).

An analogue of human proinsulin that contained three mutations which increased its biological activity was expressed in *E. coli* and purified to study the activity of the PC1 and PC2 enzymes, which are responsible for cleavage of the C chain from proinsulin (61). The structure of this proinsulin analogue was solved by solution NMR and showed a native-like insulin moiety in the A and B chains, while the structure of the C chain was less ordered (62).

Structural studies of human proinsulin

Proinsulin has been previously determined to be sequestered into inclusion bodies during production in *E. coli* (60, 61). Since there are several α -helical regions within the structure of proinsulin in solution, it would be interesting to investigate whether these helical regions are retained within bacterial inclusion bodies. In Chapter 4, a short project is discussed in which SSNMR was used to probe the secondary structure of human proinsulin within inclusion bodies.

REFERENCES

REFERENCES

1. Pochapsky, T. C. (2007) *NMR for Physical and Biological Scientists*, Garland Science, New York.
2. Sun, Y., and Weliky, D. P. (2009) ^{13}C - ^{13}C Correlation spectroscopy of membrane-associated Influenza virus fusion peptide strongly supports a helix-turn-helix motif and two turn conformations, *J. Am. Chem. Soc.* **131**, 13228-13229, PMID: 2772195.
3. Schmick, S. D., and Weliky, D. P. (2010) Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide, *Biochemistry* **49**, 10623-10635.
4. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) 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.* **397**, 301-315.
5. Tong, K. I., Yamamoto, M., and Tanaka, T. (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in E-coli, *J. Biomol. NMR* **42**, 59-67.
6. Ross, A., Kessler, W., Krumme, D., Menge, U., Wissing, J., van den Heuvel, J., and Flohe, L. (2004) Optimised fermentation strategy for C-13/N-15 recombinant protein labelling in Escherichia coli for NMR-structure analysis, *J. Biotechnol.* **108**, 31-39.
7. Weliky, D. P. (1999) Chemistry 988 Lecture Notes.
8. Gullion, T., and Schaefer, J. (1989) Rotational-echo double-resonance NMR, *J. Magn. Reson.* **81**, 196-200.
9. Gullion, T. (1998) Introduction to rotational-echo, double-resonance NMR, *Concepts Magn. Reson.* **10**, 277-289.
10. Hing, A. W., Tjandra, N., Cottam, P. F., Schaefer, J., and Ho, C. (1994) An investigation of the ligand-binding site of the glutamine-binding protein of Escherichia coli using rotational-echo double-resonance NMR, *Biochemistry* **33**, 8651-8661.
11. Gray, C., and Tamm, L. K. (1997) Structural studies on membrane-embedded influenza hemagglutinin and its fragments, *Protein Science* **6**, 1993-2006.
12. Chen, J., Skehel, J. J., and Wiley, D. C. (1999) 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.* **96**, 8967-8972.

13. Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, *Nature* **289**, 366-373.
14. Curtis-Fisk, J., Preston, C., Zheng, Z. X., Worden, R. M., and Weliky, D. P. (2007) Solid-state NMR structural measurements on the membrane-associated influenza fusion protein ectodomain, *J. Am. Chem. Soc.* **129**, 11320-11321.
15. Qiang, W., Yang, J., and Weliky, D. P. (2007) Solid-state nuclear magnetic resonance measurements of HIV fusion peptide to lipid distances reveal the intimate contact of beta strand peptide with membranes and the proximity of the Ala-14-Gly-16 region with lipid headgroups *Biochemistry* **46**, 4997-5008, PMCID: 2631438.
16. Qiang, W., and Weliky, D. P. (2009) 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* **48**, 289-301.
17. Qiang, W., Sun, Y., and Weliky, D. P. (2009) A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15314-15319.
18. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme, *Crit. Rev. Biochem. Mol. Biol.* **43**, 189-219.
19. Melikyan, G. B. (2008) Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm, *Retrovirology* **5**, 111.
20. Chernomordik, L. V., Zimmerberg, J., and Kozlov, M. M. (2006) Membranes of the world unite!, *J. Cell Biol.* **175**, 201-207.
21. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* **17**, 4572-4584.
22. Yang, Z. N., Mueser, T. C., Kaufman, J., Stahl, S. J., Wingfield, P. T., and Hyde, C. C. (1999) The crystal structure of the SIV gp41 ectodomain at 1.47 Å resolution, *J. Struct. Biol.* **126**, 131-144.
23. Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem.* **70**, 777-810.
24. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, *Plos Pathogens* **6**, e1000880.

25. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation, *Mol. Biol. Cell* **14**, 926-938.
26. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., and Panganiban, A. T. (1992) 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.* **89**, 70-74.
27. Yang, R., Prorok, M., Castellino, F. J., and Weliky, D. P. (2004) 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.* **126**, 14722-14723
28. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, *Biochemistry* **48**, 2714-2722.
29. Pan, J. H., Lai, C. B., Scott, W. R. P., and Straus, S. K. (2010) Synthetic fusion peptides of tick-borne Encephalitis virus as models for membrane fusion, *Biochemistry* **49**, 287-296.
30. Lev, N., Fridmann-Sirkis, Y., Blank, L., Bitler, A., Epand, R. F., Epand, R. M., and Shai, Y. (2009) Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action, *Biochemistry* **48**, 3166-3175.
31. Cheng, S. F., Chien, M. P., Lin, C. H., Chang, C. C., Lin, C. H., Liu, Y. T., and Chang, D. K. (2010) The fusion peptide domain is the primary membrane-inserted region and enhances membrane interaction of the ectodomain of HIV-1 gp41, *Mol. Membr. Biol.* **27**, 31-44.
32. Lin, C. H., Lin, C. H., Chang, C. C., Wei, T. S., Cheng, S. F., Chen, S. S. L., and Chang, D. K. (2011) An efficient production and characterization of HIV-1 gp41 ectodomain with fusion peptide in Escherichia coli system, *J. Biotech.* **153**, 48-55.
33. Sackett, K., TerBush, A., and Weliky, D. P. (2011) HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion, *Eur. Biophys. J.* **40**, 489-502.
34. Chang, D. K., Cheng, S. F., and Chien, W. J. (1997) The amino-terminal fusion domain peptide of human immunodeficiency virus type 1 gp41 inserts into the sodium dodecyl sulfate micelle primarily as a helix with a conserved glycine at the micelle-water interface, *J. Virol.* **71**, 6593-6602.
35. Morris, K. F., Gao, X. F., and Wong, T. C. (2004) The interactions of the HIV gp41 fusion peptides with zwitterionic membrane mimics determined by NMR spectroscopy, *Biochim. Biophys. Acta* **1667**, 67-81.

36. Jaroniec, C. P., Kaufman, J. D., Stahl, S. J., Viard, M., Blumenthal, R., Wingfield, P. T., and Bax, A. (2005) Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain, *Biochemistry* 44, 16167-16180.
37. Li, Y. L., and Tamm, L. K. (2007) Structure and plasticity of the human immunodeficiency virus gp41 fusion domain in lipid micelles and bilayers, *Biophys. J.* 93, 876-885.
38. Gabrys, C. M., and Weliky, D. P. (2007) Chemical shift assignment and structural plasticity of a HIV fusion peptide derivative in dodecylphosphocholine micelles, *Biochim. Biophys. Acta* 1768, 3225-3234.
39. Pereira, F. B., Goni, F. M., Muga, A., and Nieva, J. L. (1997) Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: dose and sequence effects, *Biophys. J.* 73, 1977-1986.
40. Grasnick, D., Sternberg, U., Strandberg, E., Wadhwani, P., and Ulrich, A. S. (2011) Irregular structure of the HIV fusion peptide in membranes demonstrated by solid-state NMR and MD simulations, *Eur. Biophys. J.* 40, 529-543.
41. Tristram-Nagle, S., Chan, R., Kooijman, E., Uppamoochikkal, P., Qiang, W., Weliky, D. P., and Nagle, J. F. (2010) HIV fusion peptide penetrates, disorders, and softens T-cell membrane mimics, *J. Mol. Biol.* 402, 139-153.
42. Yang, J., Gabrys, C. M., and Weliky, D. P. (2001) Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide, *Biochemistry* 40, 8126-8137.
43. Zheng, Z., Yang, R., Bodner, M.L., and Weliky, D.P. (2006) Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy, *Biochemistry* 45, 12960-12975.
44. Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) 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.* 130, 5459-5471.
45. Brugger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F. T., and Krasslich, H. G. (2006) The HIV lipidome: A raft with an unusual composition, *Proc. Natl. Acad. Sci. U.S.A.* 103, 2641-2646.
46. Sackett, K., Wexler-Cohen, Y., and Shai, Y. (2006) Characterization of the HIV N-terminal fusion peptide-containing region in context of key gp41 fusion conformations, *J. Biol. Chem.* 281, 21755-21762.

47. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in *E. coli*, purification, and refolding of the full ectodomain of the Influenza virus membrane fusion protein, *Prot. Expr. Purif.* 61, 212-219.
48. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* 286, 13226-13234.
49. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., and Wongstaal, F. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III, *Nature* 313, 277-284.
50. Painter, S. L., Biek, R., Holley, D. C., and Poss, M. (2003) Envelope variants from women recently infected with clade A human immunodeficiency virus type 1 confer distinct phenotypes that are discerned by competition and neutralization experiments, *J. Virol.* 77, 8448-8461.
51. Pascual, R., Moreno, M. R., and Villalain, J. (2005) A peptide pertaining to the loop segment of human immunodeficiency virus gp41 binds and interacts with model biomembranes: Implications for the fusion mechanism, *J. Virol.* 79, 5142-5152.
52. Marston, F. A. O. (1986) The Purification Of Eukaryotic Polypeptides Synthesized In *Escherichia-Coli*, *Biochemical Journal* 240, 1-12.
53. Kane, J. F., and Hartley, D. L. (1988) Formation Of Recombinant Protein Inclusion-Bodies In *Escherichia-Coli*, *Trends In Biotechnology* 6, 95-101.
54. Tao, H., Liu, W., Simmons, B. N., Harris, H. K., Cox, T. C., and Massiah, M. A. (2010) Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS, *Biotechniques* 48, 61-64.
55. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. D. (1979) Expression In *Escherichia-Coli* Of Chemically Synthesized Genes For Human Insulin, *Proceedings Of The National Academy Of Sciences Of The United States Of America* 76, 106-110.
56. Gross-Selbeck, S., Margreiter, G., Obinger, C., and Bayer, K. (2007) Fast quantification of recombinant protein inclusion bodies within intact cells by FT-IR spectroscopy, *Biotechnology Progress* 23, 762-766.
57. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy, *J. Am. Chem. Soc.* 130, 12568-12569.

58. Curtis-Fisk, J. (2009) Structural studies of the Influenza and HIV viral fusion proteins and bacterial inclusion bodies, Ph. D. Thesis, Michigan State University.
59. CDC. (2011) National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011, Atlanta, GA U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.
60. Cowley, D. J., and Mackin, R. B. (1997) Expression, purification and characterization of recombinant human proinsulin, *Febs Letters* 402, 124-130.
61. Mackin, R. B., and Choquette, M. H. (2003) Expression, purification, and PC1-mediated processing of (H10D, P28K, and K29P)-human proinsulin, *Protein Expression And Purification* 27, 210-219.
62. Yang, Y., Hua, Q.-x., Liu, J., Shimizu, E. H., Choquette, M. H., Mackin, R. B., and Weiss, M. A. (2010) Solution Structure of Proinsulin CONNECTING DOMAIN FLEXIBILITY AND PROHORMONE PROCESSING, *Journal Of Biological Chemistry* 285, 7847-7851.

Chapter 2 – Studies of Fgp41, an ectodomain construct of HIV fusion protein gp41

Introduction

This chapter will discuss structural and functional studies of recombinantly produced constructs of gp41, the fusion protein of the Human Immunodeficiency Virus (HIV) as well as advances in biochemistry techniques such as protein expression and purification that I have made while working with the Fgp41 protein. Chapter 1 provides a brief introduction to the gp41 protein and its significance in HIV infection. The majority of the work discussed in this chapter was published in Biochemistry in 2011(1). By working with a construct that represents the majority of the ectodomain of gp41 including the native loop between the N and C helices (as defined in Figure 1-5), I was able to examine whether past studies in our group utilizing smaller constructs accurately modeled the fusion peptide in the context of the protein. In addition to looking for structural and functional similarities between the engineered constructs and Fgp41, I was able to examine structural differences that might arise from the difference of the protein sequence in the fusion peptide region between different strains of HIV-1. Figure 2-9 highlights the sequence variation between the strain of HIV-1 utilized in these studies and the lab isolated HXB2 strain which is used in most other structural and functional studies of the gp41 protein. The sequence used in these studies comes from a strain of HIV-1 which uses the CCR5 co-receptor (in addition to CD4 receptor) for entry (an “M-tropic” strain), as opposed to the HXB2 strain which uses the CXCR4 co-receptor for entry (a “T-tropic” strain)(2, 3). M-tropic strains initiate infection, and individuals deficient in CCR5 receptors are resistant to HIV-1(4).

Fgp41 Construct Information

Source of Fgp41

The Fgp41 plasmid was constructed in the lab of Dr. Jun Sun (Department of Engineering, Michigan State University, East Lansing, MI). The plasmid was engineered by inserting cDNA into the commercially available pET24a(+) vector. The cDNA was obtained from the lab of Dr. William Wedemeyer (Department of Biochemistry, Michigan State University, East Lansing, MI). The source of this strain of HIV was patient sera of a recently infected Kenyan woman, and belongs grouped within the Clade A strains of HIV (Los Alamos HIV Database accession id: AY288087)(5).

DNA Sequence of Fgp41

Shown below is the DNA sequence of Fgp41, with the DNA corresponding to Fgp41 shown in bold, and the rest corresponding to the surrounding vector DNA.

TTTTGTTAACTTTAGAAGGAGATATACATAT**GGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGG**
GAGCAGCAGGGAGCACTATGGGCGCGGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTC
TGGCATAGTGCAACAGCAAAGCAATTTGCTGAAGGCTATAGAGGCTCAACAGCATCTGTTGAAACTCA
CGGTCTGGGGTATTAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGATACCTACAGGATCAACA
GCTCCTGGGAATTTGGGGCTGCTCTGGAAAACTCATCTGCACCTCTTTTGTGCCCTGGAACAATAGTTG
GAGTAACAAGACTTATAATGAGATTTGGGACAACATGACCTGGTTGCAATGGGATAAAGAAATTAGC
AATTACACAGACACAATATACAGGCTACTTGAAGACTCGCAGAACCAGCAGGAAAAGAATGAACAAG
ACTTATTGGCATTAGATAAACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCC

Protein Sequence of Fgp41

The protein sequence of Fgp41 is shown below. Underlined in the sequence are two non-native residues (which act as a linker) as well as a polyhistidine tag for purification purposes.

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIEA
QQHLLKLTWVGIKQLQARVLAVERYLQDQQLLGIWGC SGKLICTSFVPWN

NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKLEHHHHHH

Fgp41 Expression Optimization

Experiments were designed to investigate the effects on Fgp41 expression of: (1) [glycerol] in the expression medium; (2) [IPTG]; and (3) induction time. The protocol included: (1) overnight 37 °C cell growth from glycerol stock in 2 L of LB; (2) cell pelleting by centrifugation followed by resuspension in 1 L of LB; (3) growth at 37 °C for one hour; (4) transferring 100 mL aliquots of medium into separate flasks; (5) addition of glycerol and then IPTG with concomitant induction of expression at 23 °C; (6) cell pelleting by centrifugation followed by lysis in buffer with 1% SDS; and (7) SDS-PAGE of the soluble cell lysates with visual comparison of their Fgp41 band intensities. In general, only one parameter, e.g. [IPTG], was varied among a group of aliquots. Results included: (1) comparison between [IPTG] = 0.2 mM, 1.0, or 2.0 mM showed the darkest band at 2.0 mM; (2) comparison between [glycerol] = 0.1, 0.25, or 0.5% (v/v) showed the darkest bands for 0.1 and 0.25%; and (3) comparison between induction time = 2, 4, or 6 hours showed the darkest band for 6 hours. Subsequent experiments were done using [IPTG] = 2 mM, 0.25% glycerol, and 6 hour induction.

The protocol to produce isotopically labeled Fgp41 for NMR experiments was based on a previous protocol for the influenza virus fusion protein ectodomain FHA2(6). One key feature was initial bacterial growth in rich medium (LB) to high cell densities. Relative to initial growth in minimal medium, protein production was augmented by the cell densities and by the larger number of ribosomes per cell. Bacterial cell cultures were grown in media containing 15 mg/L kanamycin because the pET24a(+) vector contains a gene for kanamycin resistance. Bacterial

cells in 1 mL of 80/20 (v/v) H₂O/glycerol were added to two 2.8 L baffled fernbach flasks which each contained 1 L of LB and were capped with a foam plug. Bacterial growth to OD₆₀₀ ~4 occurred during overnight incubation at 37 °C with shaking at 140 rpm. The cell suspensions were centrifuged (10000g, 10 min) and the cell pellets were harvested and then resuspended in a single flask containing 1 L of fresh medium with M9 minimal salts, 2.0 mL of 1.0 M MgSO₄, and 5.0 mL of 50% glycerol solution. Growth resumed after approximately one hour of incubation at 37 °C. At this time, 100 mg/L of 1-¹³C amino acid and 100 mg/L of ¹⁵N amino acid (or 100 mg/L of 1-¹³C, ¹⁵N amino acid) were added to the medium. IPTG was then added to a final concentration of 2 mM which induced expression of Fgp41 (6 hours, 23 °C). The cell pellet was harvested after centrifugation and stored at -80 °C. The wet cell mass was ~8 g.

Fgp41 Purification Protocol Development

The basis for the development of the Fgp41 purification protocol was an earlier protocol developed in our lab for FHA2(6). Initial cell lysis buffers contained either 8 M urea, 0.5% N-lauroylsarcosine, 0.5% Triton X-100, or 10% SDS, and all buffers contained 50 mM sodium phosphate, 300 mM NaCl, and were at pH = 8. The solubilization efficiency of the buffer was assessed using detection of a band at ~19 kDa (assigned to Fgp41) in the SDS-PAGE of the soluble lysate and then consideration of the absolute intensity of this band as well as its intensity relative to other bands in the gel lane. A dark Fgp41 band that was intense relative to other proteins was observed with lysis in buffer containing SDS, shown in Figure 2-1.

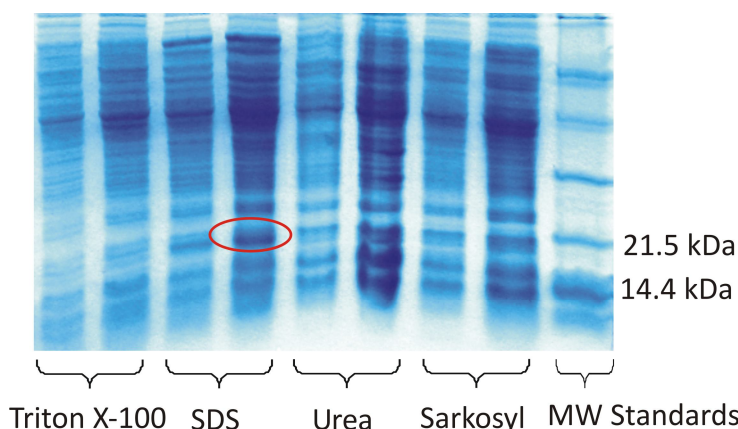


Figure 2-1: Representative SDS-PAGE gel of soluble cell lysates produced using buffers with different detergents or urea. For each buffer, the left and right lanes respectively correspond to 2 and 5 μ L aliquots of lysate. The ~19 kDa band apparent in some lanes is assigned to Fgp41. One example is circled in red in lane 4 for lysis in SDS.

Bands that may be Fgp41 were also apparent for lyses in either urea or N-lauroylsarcosine but purifications of these lysates consistently yielded <1 mg Fgp41/L culture whereas purifications of SDS lysates yielded >1 mg Fgp41/L culture. Subsequent lyses were therefore done with SDS. The effect of SDS concentration on Fgp41 solubilization was further investigated by comparison of lysis in buffer containing either 0.5%, 1%, 3%, or 5% SDS. For 1%, a dark band that was intense relative to other proteins was observed as shown in Figure 2-2.

Subsequent lyses were done using 1% SDS. The effect of different sonication conditions during lysis on Fgp41 solubilization was also investigated. The darkest Fgp41 band was observed using four 1-minute cycles at 80% amplitude with 0.8 seconds on/0.2 seconds off. Increasing the number of cycles did not result in a darker band.

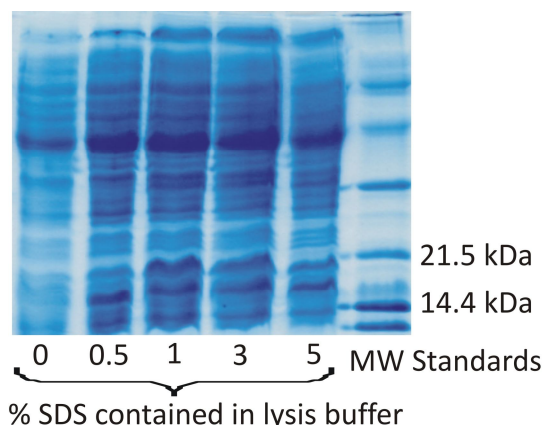


Figure 2-2: Representative SDS-PAGE gel of lysates of soluble cell lysates produced using buffers containing different concentrations of SDS. The ~19 kDa band was assigned to Fgp41 and is most apparent in the lane corresponding to 1% SDS lysis buffer.

The Fgp41 band was observed in elution fractions with a modified protocol using buffers that contained 50 mM sodium phosphate at pH 8.0, 0.5% SDS, 300 mM NaCl, and imidazole with different concentrations. Relative to only washing with buffer containing [imidazole] = 20 mM, SDS-PAGE showed that sequential washes with buffers containing [imidazole] = 1, 20, and then 50 mM was more effective at washing non-Fgp41 proteins from the resin while leaving most Fgp41 bound to the resin. After the washes, the Fgp41 was eluted from the resin using buffer containing [imidazole] = 250 mM. The eluent was incubated overnight at 4 °C with consequent precipitation of excess SDS. Negligible Fgp41 precipitated as evidenced by very similar A_{280} measurements for the eluent before and after incubation. SDS-PAGE showed that the eluent contained Fgp41 at high purity, Figure 2-3a, and that Fgp41 could be membrane-reconstituted, Figure 2-3b.

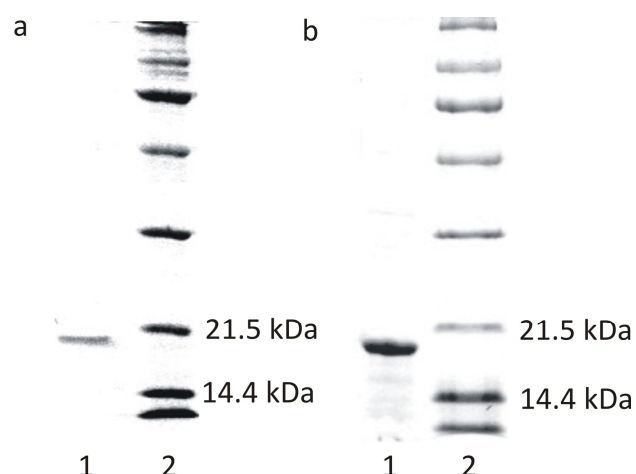


Figure 2-3: (a) SDS-PAGE gels of (lane 1) an elution aliquot of Fgp41 in buffer containing 250 mM imidazole and (lane 2) molecular weight standards. (b) SDS-PAGE gel of (lane 1) an aliquot of the proteoliposome complexes formed during membrane reconstitution of Fgp41 and (lane 2) molecular weight standards. The samples were boiled prior to loading on the gel.

The final purified yield of Fgp41 (as determined by A_{280}) was ~5 mg/L culture. This yield was obtained using one hour initial mixing of the lysate and resin with similar yield obtained for two hour mixing and reduced 3 mg/L yield for four hour mixing. Increased proteolysis is one explanation for reduced yield with longer mixing time.

Circular Dichroism Spectroscopy of Fgp41

Spectra were obtained using a CD instrument (Chirascan, Applied Photophysics, Surrey, United Kingdom), 1 mm pathlength, a 260-200 nm spectral window, wavelength points separated by 0.5 nm, and 1 s signal averaging per point. Fgp41 samples were prepared by precipitation of excess SDS followed by overnight dialysis into HEPES/MES buffer at pH 7.4 with DTT added at two times the molar concentration of Fgp41 to prevent disulfide bond formation. Most spectra were obtained with $[Fgp41] = 20 \mu M$. For each sample, a reference spectrum was also taken of buffer without Fgp41 and the relevant Fgp41 spectrum was the difference between the Fgp41 + buffer and buffer only spectra.

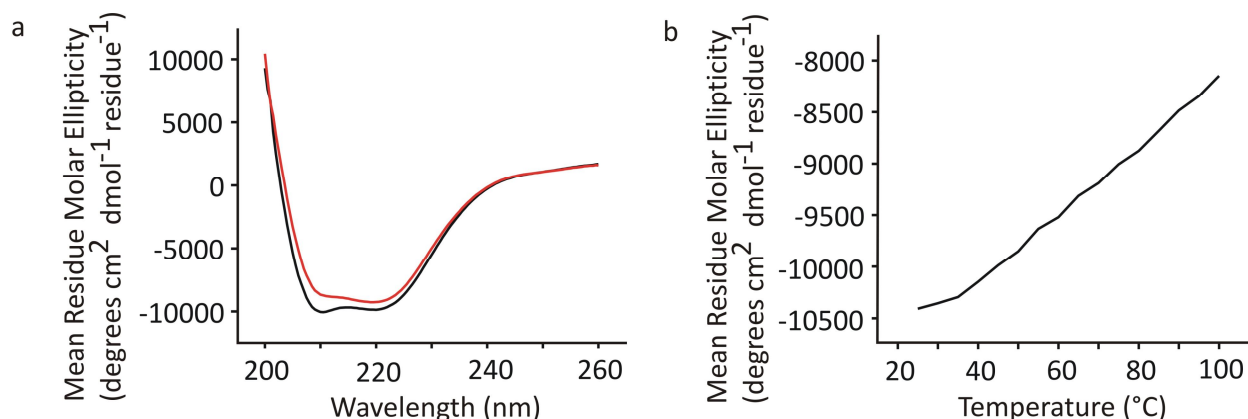


Figure 2-4: (a) CD spectra of Fgp41 at 25 °C. The black trace is for a sample that has not been heated, and the red trace was obtained after the sample had been heated to 100 °C with subsequent cooling to 25 °C. Each trace is the difference between the CD spectrum of Fgp41 with buffer and the spectrum of buffer alone. Fgp41 samples were prepared by precipitation of excess SDS, subsequent dialysis in HEPES/MES buffer (pH 7.4), and addition of DTT at two times the molar concentration of Fgp41 to inhibit disulfide bond formation. For these spectra, the concentration of Fgp41 was 20 μM . Spectra for other Fgp41 samples were similar with minima near 208 and 222 nm that were diagnostic of α -helical structure. In some spectra, θ_{222} could be as low as $-15000 \text{ deg cm}^2 \text{ dmol}^{-1}$. (b) Plot of CD θ_{222} vs temperature for Fgp41. No unfolding transition is apparent for temperatures up to 100 °C. Sample conditions were the same as those described in (a).

Figure 2-4(a) (black trace) displays the CD spectrum of the purified Fgp41 after dialysis into HEPES/MES buffer at pH 7.4. Minima near 208 and 222 nm were diagnostic of α -helical

conformation as might be expected from the Hairpin structure, Figure 1-1. The magnitude of θ_{222} showed a small linear decrease over the 25 – 100 °C range which can be seen in Figure 2-4b, where $(\theta_{222}^{100^{\circ}\text{C}}) \approx 0.8 \times (\theta_{222}^{25^{\circ}\text{C}})$. The CD spectra at 25 °C were very similar before and after heating, as shown in Figure 2-4(a), and showed that the temperature-dependent changes were reversible. This behavior was very similar to the temperature dependences of the CD spectra of the shorter Hairpin and FP-Hairpin constructs whose sequence was from the laboratory HXB2 strain of HIV-1(7, 8). For these constructs, 46 contiguous residues including the native loop were replaced by 6 non-native residues. Subsequent differential scanning calorimetry experiments showed an unfolding transition centered at 110 °C for both constructs. Consideration of other CD measurements on this unfolded state indicate that for Fgp41, $(\theta_{222}^{\text{unfolded}}) \approx 0.2 \times (\theta_{222}^{25^{\circ}\text{C}})$ so even at 100 °C, Fgp41 appears to retain hyperthermostable hairpin structure.

Fluorescence Based Lipid-Mixing Assays for Activity of Fgp41

One early step in fusion between the HIV and target cell membranes is mixing of lipids between the two membranes. This aspect of Fgp41 fusogenicity was probed by a fluorescence based assay that detected Fgp41-induced mixing of lipids between membrane vesicles. Initially, there are two populations of vesicles, some containing unlabeled lipids and some containing unlabeled lipids and a small percentage of a fluorescence donor (FD) and acceptor (FA) pairs. Fluorescence resonance energy transfer efficiency is proportional to r^{-3} . If the distance between the FD and FA is small, most of the fluorescence emitted by the FD will be absorbed by FA and thus minimal fluorescence will be observed experimentally. However, if the FD and FA

are further apart, there will be increasingly less transfer, and more fluorescence will be observed experimentally.

Experimental Details

The initial fluorescence is monitored and recorded as a “zero percent lipid mixing”. Protein is added to the solution of lipid vesicles, and if the protein perturbs the vesicles, it will cause lipid mixing, and ultimately the formation of larger lipid vesicles. As this occurs, the fluorescent donor and quencher molecules end up with larger intermolecular distances; this leads to an observed increase in fluorescence. To determine the extent of lipid mixing caused by the addition of protein, Triton X-100 is added to the system. Triton X-100 is thought to completely solubilize lipid vesicles, thereby resulting in the largest possible fluorophore-quencher intermolecular distance and maximal fluorescence. The observed level of fluorescence after the addition of Triton X-100 is considered “100 percent lipid mixing”.

A set of vesicles was prepared that contained POPC:POPG lipids in 4:1 mol ratio and another set of “labeled” vesicles was prepared that contained an additional 2 mol % of the fluorescent lipid N-NBD-PE and 2 mol % of the quenching lipid N-Rh-PE. Large unilamellar vesicles (LUVs) were prepared by: (1) dissolving lipids in chloroform and then removing chloroform by nitrogen gas and overnight vacuum; (2) formation of pH 7.5 aqueous lipid dispersions with [total lipid] $\approx$ 5 mM and [HEPES] = 25 mM including five freeze-thaw cycles; and (3) ~20-fold extrusion through a polycarbonate filter with 0.1 μ M diameter pores. The assay was done at 37 °C with continuous stirring in the HEPES buffer using a mixture of unlabeled vesicles ([total lipid] = 135 μ M) and labeled vesicles ([total lipid] = 15 μ M). After measuring the initial fluorescence F_0 , an aliquot of 30 μ M Fgp41 in HEPES/MES buffer was

added to the vesicle solution so that final $[Fgp41] = 3 \mu M$ and $Fgp41:total\ lipid = 0.02$. Fgp41-induced fusion between labeled and unlabeled vesicles resulted in larger fluorophore-quencher distance and increased fluorescence. The fluorescence increase ΔF_{Fgp41} was compared to the maximum fluorescence increase (ΔF_{max}) obtained after subsequent addition of Triton X-100 detergent which solubilized the vesicles. Assay parameters included: (1) fluorimeter (Photon Technology International); (2) excitation and emission wavelengths of 465 and 530 nm with 4 nm bandwidths; and (3) 1.8 mL of initial vesicle solution, 0.2 mL aliquot of Fgp41, and $\sim 20 \mu L$ aliquot of 10% Triton X-100.

Fgp41 induced negligible intervesicle fusion at pH 7.5 as assayed by lipid mixing, Figure 2-5. The fluorescence increase was $\sim 2\%$ of that observed for Triton X-100 detergent where Triton is commonly considered to induce 100% lipid mixing.

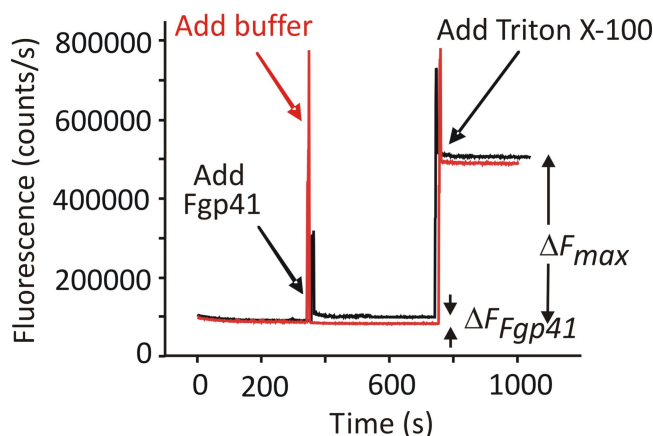


Figure 2-5: Vesicle fusion assayed by fluorescence. An aliquot of either Fgp41 with buffer (black trace) or buffer alone (red trace) was added to a vesicle solution at 350 s. Fgp41-induced vesicle fusion was evidenced by the fluorescence increase (ΔF_{Fgp41}) of the black trace. In either trace, Triton X-100 was added at 750 s and solubilized the vesicles, resulting in maximal fluorescence and fluorescence increase (ΔF_{max}). The spikes at 350 and 750 s were artifacts caused by transient exposure to stray light. Assay parameters included vesicles with 4:1 POPC:POPG composition, Fgp41:total lipid molar ratio of 1:50, pH 7.5, 37 °C.

Solid-State NMR Analysis of Membrane Associated Fgp41

Membrane Reconstitution

For studies of Fgp41 using Solid-State NMR, purified Fgp41 was reconstituted into lipid vesicles so that the protein could be studied in a biologically relevant environment. The composition of the lipid vesicles utilized in these studies was designed to include a 4:1 ratio of choline : negatively charged lipid headgroups as is seen in HIV membranes (9).

A homogeneous mixture of the POPC (27 mg) and POPG (7 mg) lipids and the bTOG (136 mg) detergent was made by: (1) dissolution in chloroform; (2) removal of chloroform by nitrogen gas and overnight vacuum; and (3) dissolution in HEPES/MES buffer. Fgp41 (~10 mg) was added to the solution and had been in affinity column eluents for which excess SDS had been removed by overnight incubation at 4 °C. Dialysis of the bTOG/lipid/Fgp41 solution against HEPES/MES buffer removed bTOG with consequent liposome formation with bound Fgp41. Dialysis parameters included: (1) bTOG/lipid/Fgp41 solution in 10 KDa MWCO tubing (~15 mL initial volume); (2) 3L buffer volume; and (3) 3 day duration at 4 °C while stirring with one buffer change. The proteoliposome pellet was harvested after centrifugation (50000g, 3 hours) and unbound Fgp41 did not pellet under these conditions. The pellet was packed into a 4 mm diameter magic angle spinning (MAS) rotor with ~5 mg Fgp41 and ~20 mg total lipid in the 40 µL active sample volume.

SSNMR Experimental Parameters

Data were obtained with a 9.4 T instrument (Agilent Infinity Plus) and a triple-resonance MAS probe whose rotor was cooled with nitrogen gas at -10 °C. Because of heating from MAS and RF radiation, we expect that water in the sample was liquid rather than solid. Experimental

parameters included: (1) 8.0 kHz MAS frequency; (2) 5 μ s ^1H $\pi/2$ pulse and 2 ms cross-polarization time with 50 kHz ^1H field and 70-80 kHz ramped ^{13}C field; (3) 1 or 2 ms rotational-echo double-resonance (REDOR) dephasing time with a 9 μ s ^{13}C π pulse at the end of each rotor period except the last period and for some data, a 12 μ s ^{15}N π pulse at the center of each rotor period; and (4) ^{13}C detection with 90 kHz two-pulse phase modulation ^1H decoupling (which was also on during the dephasing time); and (5) 0.8 sec pulse delay(10). Data were acquired without (S_0) and with (S_1) ^{15}N π pulses during the dephasing time and respectively represented the full ^{13}C signal and the signal of ^{13}C s not directly bonded to ^{15}N nuclei. The $S_0 - S_1$ (ΔS) difference signal was therefore dominated by the labeled ^{13}CO s in the sequential pairs targeted by the labeling. Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm so that the ^{13}CO shifts could be directly compared to those of soluble proteins(11).

SSNMR Experimental Results

Figure 2-6 displays S_0 , S_1 , and ΔS REDOR SSNMR spectra of membrane-reconstituted Fgp41 labeled with different amino acids. Many of these spectra were deconvolved into a few Gaussian line shapes, see Figure 2-6, 2-7, and 2-8. Table 2-1 presents the best-fit peak chemical shifts, line widths, and integrated intensities of the individual line shapes of the S_0 spectra and Table 2-2 presents a numerical breakdown of the S_0 line shape into contributions from natural

abundance signals, labeled signal, and labeled signal within the helices of Fgp41. Table 2-2 also presents a calculated $\Delta S/S_0$ value to compare with the experimental $\Delta S/S_0$ for each labeling. Table 2-3 presents the line shape parameters of the ΔS spectra. All fits were excellent as judged by the close agreement between the line shape sum and the experimental intensity, see Figures 2-7 and 2-8. These fittings were used to understand whether or not the N-helix and C-helix structures of the six-helix bundle were retained in the membrane-associated Fgp41 and to assess the distribution of conformations in the FP region.

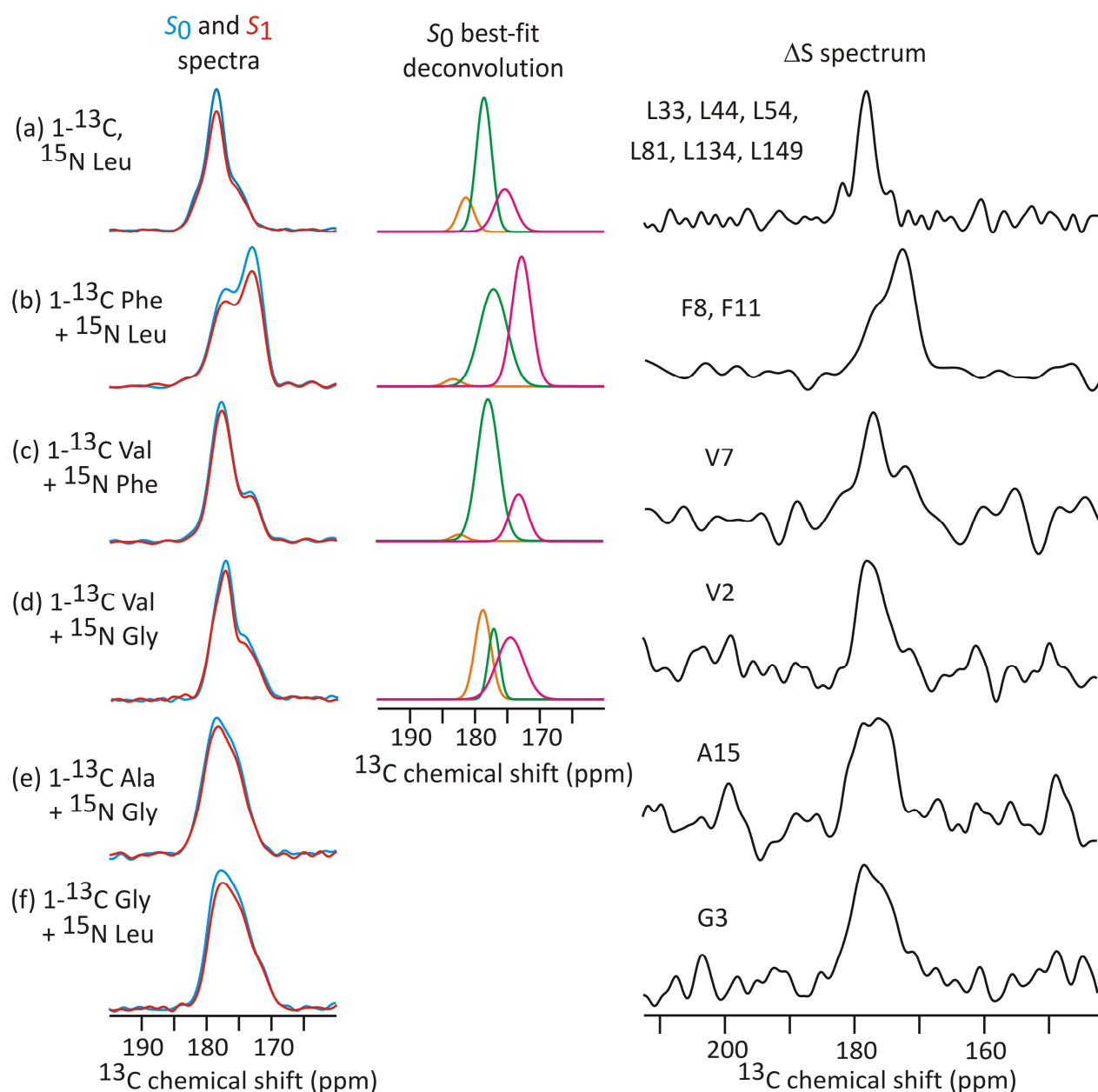


Figure 2-6: REDOR ^{13}C CO NMR spectra of Fgp41 reconstituted in membranes. The labeled amino acids in the expression medium are shown. The left panels display S_0 (blue) and S_1 (red) spectra; the middle panels display the best-fit Gaussian deconvolutions of the S_0 spectra, and the right panels display $\Delta S \equiv S_0 - S_1$ spectra. The REDOR dephasing time was either (a) 1 or (b-f) 2 ms, and the dominant contribution to each ΔS spectrum was from residues labeled with ^{13}C that were directly bonded to labeled ^{15}N atoms. The major contribution to each ΔS spectrum is indicated. Each S_0 or S_1 spectrum was processed with 100 Hz Gaussian line broadening, and each ΔS spectrum was processed with (a and b) 100 or (c-f) 200 Hz line broadening. Polynomial

baseline correction (typically fifth order) was applied to each spectrum. Each S_0 or S_1 spectrum was the sum of (a) 93424, (b) 115610, (c) 109504, (d) 110736, (e) 165216, or (f) 103717 scans.

Table 2-1: Analysis and deconvolution of S_0 SSNMR spectra of membrane reconstituted Fgp41.

^aSpectral deconvolution was conducted with three Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For all cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 2-7. Deconvolution was not meaningful for the $1\text{-}^{13}\text{C}$ Ala and $1\text{-}^{13}\text{C}$ Gly samples because the S_0 spectra were broad and relatively featureless, resulting in deconvolutions that were dominated by a line shape with ~ 7 ppm line width. ^bThe conformations designated are assigned based on RefDB(12). ^cFull width at half-maximal line width.

Fgp41 labeling	S_0 spectral deconvolution ^a			
	Peak shift (ppm) ^b		Peak width (ppm) ^c	Intensity (fraction of total)
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu	181.3	<i>helix</i>	2.8	0.15
	178.5	<i>helix</i>	2.8	0.60
	175.3	β	3.7	0.25
$1\text{-}^{13}\text{C}$ Phe + ^{15}N Leu	183.2	<i>helix</i>	3.0	0.02
	177.1	<i>helix</i>	5.0	0.51
	172.7	β	3.5	0.47
$1\text{-}^{13}\text{C}$ Val + ^{15}N Phe	182.1	<i>helix</i>	2.7	0.02
	177.7	<i>helix</i>	3.9	0.76
	173.1	β	3.3	0.22
$1\text{-}^{13}\text{C}$ Val + ^{15}N Gly	178.6	<i>helix</i>	2.9	0.37
	177.0	<i>helix</i>	2.0	0.20
	174.4	β	4.9	0.43

Table 2-2: Comparison between experimental and calculated REDOR dephasing for membrane reconstituted Fgp41.

Fgp41 labeling	Fraction of calculated S_0 intensity ^a				$(\Delta S/S_0)^{calc, b}$	$(\Delta S/S_0)^{exp}$ (integrated) ^c
	Labeled	Nat. abund. Fgp41	Nat. abund. lipid	Labeled in N- and C-helices		
$1-^{13}\text{C}, ^{15}\text{N}$ Leu	0.86	0.07	0.07	0.68	0.15	0.12
$1-^{13}\text{C}$ Phe + ^{15}N Leu	0.42	0.31	0.27	0	0.24	0.15
$1-^{13}\text{C}$ Val + ^{15}N Phe	0.69	0.19	0.12	0.43	0.07	0.07
$1-^{13}\text{C}$ Val + ^{15}N Gly	0.64	0.17	0.19	0.40	0.07	0.08
$1-^{13}\text{C}$ Ala + ^{15}N Gly	0.79	0.12	0.09	0.34	0.05	0.08
$1-^{13}\text{C}$ Gly + ^{15}N Leu	0.77	0.15	0.08	0.14	0.06	0.11

^aContribution to spectral intensities were calculated with the following considerations: (1) 100% labeling of the Fgp41 residues corresponding to the labeled amino acid(s) with no scrambling to other amino acid types, (2) 1.0 relative intensity for each labeled ^{13}CO , (3) 0.011 relative intensity for each natural abundance ^{13}CO , (4) the Fgp41 natural abundance contribution as the sum from backbone ^{13}CO groups and Asn, Asp, Gln, and Glu side chain ^{13}CO groups, and (5) the lipid natural abundance signal calculated using the experimental Fgp41:total lipid molar ratios. The specific ratio in each sample was as follows: $1-^{13}\text{C}, ^{15}\text{N}$ Leu: 0.011; $1-^{13}\text{C}$ Phe + ^{15}N Leu: 0.012; $1-^{13}\text{C}$ Val + ^{15}N Phe: 0.016; $1-^{13}\text{C}$ + ^{15}N Gly: 0.009; $1-^{13}\text{C}$ Ala + ^{15}N Gly: 0.013; and $1-^{13}\text{C}$ Gly + ^{15}N Leu: 0.019. The labeled ^{13}CO fraction in N- and C-helices was based

on the red and green regions in Figure 1-1a. $^b(\Delta S/S_0)^{calc}$ values were based on (1) the fraction of the S_0 signal from labeled ^{13}CO directly bonded to labeled ^{15}N atoms and (2) an S_1/S_0 intensity ratio for these ^{13}CO of 0.70 (1 ms dephasing time) or 0.85 (2 ms dephasing time). These ratios were based on experimental REDOR data of crystalline glycine as well as simulations (Jun Yang Ph.D. Dissertation 2003). The 1 ms dephasing time was used for the ^{13}C , ^{15}N Leu Fgp41 sample and 2 ms dephasing time was used for all other Fgp41 samples. c The typical uncertainty of $(\Delta S/S_0)^{exp}$ was ± 0.02 as determined from the standard deviation of integrals of regions of the S_0 and S_1 spectra that contained noise rather than signal.

Figure 2-6a displays the ^{13}CO spectra of the ^{13}C , ^{15}N Leu-labeled sample. The S_0 spectrum targeted the 24 Leus in the Fgp41 sequence and the ΔS spectrum targeted the L33, L44, L54, L81, L134, and L149 ^{13}CO s which are the N-terminal Leus in LL repeats. The ^{13}CO signal was the only discernible feature in the ΔS spectrum. Both the S_0 and ΔS spectra had high signal-to-noise and were fitted well to the sum of three components. In both cases, the two higher shift components comprised >75% of the integrated intensity and were assigned to helical conformation because their peak shifts were much closer to the characteristic shifts of helical Leus (Gaussian distribution of 178.5 ± 1.3 ppm) than to β strand Leus (175.7 ± 1.5 ppm)(12). The ^{13}CO S_0 spectrum had contributions from the labeled Fgp41 Leus, as well as natural abundance sites in Fgp41 and lipids. Calculated relative fractional contributions are listed in Table 1 and show that the Fgp41 Leus dominate the spectrum. Using a S_1/S_0 intensity ratio of 0.3 for the N-terminal Leus of the LL pairs and a ratio of 1.0 for other ^{13}CO s (based on model compound studies and simulations), the $(\Delta S/S_0)^{calc}$ for the sample was 0.15 and correlated reasonably well with the $(\Delta S/S_0)^{exp}$ of 0.12 ± 0.02 (13).

If the SHB structure were retained in membrane-associated Fgp41, then the fractional contribution to the S_0 ^{13}C O intensity of Leus in the N- and C-helices would be 0.68. This correlated well with the experimental fractional S_0 intensity of 0.75 in helical conformation and supports retention of SHB structure upon membrane binding. Further support for this structure was the correlation between the experimental helical fractional intensity of 0.92 in the ΔS spectrum and the location of the six LL repeats in the N- and C-helices.

Spectra of the remaining labeled samples provided information about structure in the putative SHB region as well as in the FP. Figure 2-6b displays spectra from a sample with 1- ^{13}C Phe and ^{15}N Leu labeling. There are three Phes in the sequence: F8 and F11 in the FP, and F96 which would be in the loop region of a SHB structure. There was ~ 0.4 fractional contribution of the labeled Phe ^{13}C O to the S_0 spectrum and ~ 0.3 contributions each from natural abundance ^{13}C O in Fgp41 and lipid. The S_0 spectrum was well-fitted to the sum of three line shapes. The two line shapes with higher peak shifts comprised ~ 0.5 fractional contribution of the total intensity and the shifts were generally consistent with helical protein conformation. The peak shift of the other line shape was consistent with β strand protein conformation and with lipid shifts. The labeled F8 and F11 ^{13}C O in the FP were directly bonded to labeled Leu ^{15}N s with S_1/S_0 of ~ 0.15 for 2 ms dephasing time(13). The other ^{13}C O had S_1/S_0 of ~ 1 . The $(\Delta S/S_0)^{\text{calc}}$ was close to $(\Delta S/S_0)^{\text{exp}}$ and the ΔS spectrum was dominated by the F8 and F11 ^{13}C O signals. The ΔS spectrum was well-fitted to two line shapes with the higher (lower) peak shifts

consistent with helical (β strand) Phe ^{13}C CO shift distributions of 177.1 ± 1.4 (174.3 ± 1.6) ppm. The lower ~ 173 ppm experimental peak shift matched well with the 173 ppm peak shifts measured for F8 and F11 of the membrane-associated HFP fragment(14-16). This peptide has been shown to form small oligomers with anti-parallel β sheet structure(17). For membrane-bound Fgp41, the ratio for F8 + F11 of helical to β strand/sheet intensities was $\sim 1:2$ and was consistent with two Fgp41 populations with different FP conformations.

Figure 2-6c displays the spectra and analysis for a sample labeled with $1\text{-}^{13}\text{C}$ Val and ^{15}N Phe. The analysis approach was the same as in the previous paragraph. The eight labeled Val ^{13}C COs made a fractional contribution of ~ 0.7 to the S_0 signal. The S_0 spectrum was well-fitted to three line shapes and the two higher shift line shapes comprised ~ 0.8 fraction of the total intensity and had shifts that correlated with the helical rather than the β strand Val ^{13}C CO distribution (177.7 ± 1.4 vs 174.8 ± 1.4 ppm)(12). The line shape with lowest peak shift correlated with β strand/sheet conformation. The high helical content was consistent with SHB structure for membrane-bound Fgp41. The $(\Delta S/S_0)^{\text{calc}}$ matched $(\Delta S/S_0)^{\text{exp}}$. The ΔS spectrum was dominated by V7 and was well-fitted to three line shapes which indicated a ratio of helical to β strand/sheet populations of $\sim 2:1$. This ΔS spectrum confirmed two Fgp41 populations with different FP conformations while the difference in population ratio relative to the Figure 2-6b ΔS spectrum may reflect lower signal-to-noise of the Figure 2-6c spectrum, sample-to-sample variation, and/or conformational differences between V7 and F8 + F11.

Figure 2-6d displays the spectra and analysis for a sample labeled with $1\text{-}^{13}\text{C}$ Val and ^{15}N Gly. As with Figure 2-6c, analysis of the S_0 spectrum of Figure 2-6d supported a dominant helical conformation consistent with six-helix bundle structure. Comparison of the two spectra provided insight into sample-to-sample variation and the robustness of the S_0 deconvolution. The $(\Delta S/S_0)^{\text{calc}}$ matched $(\Delta S/S_0)^{\text{exp}}$. The ΔS spectrum was dominated by V2 and extended broadly over 170-180 ppm region so that deconvolution was not meaningful. As noted in the previous paragraph, this shift range includes the helical and β strand/sheet shift distributions and the ΔS spectrum was therefore consistent with a mixture of Fgp41 populations with helical and β strand/sheet conformations at V2 in the FP. We note that the V2 ^{13}CO signal of the membrane-associated HFP was also broader than signals from residues 6-12 in the interior hydrophobic region(14).

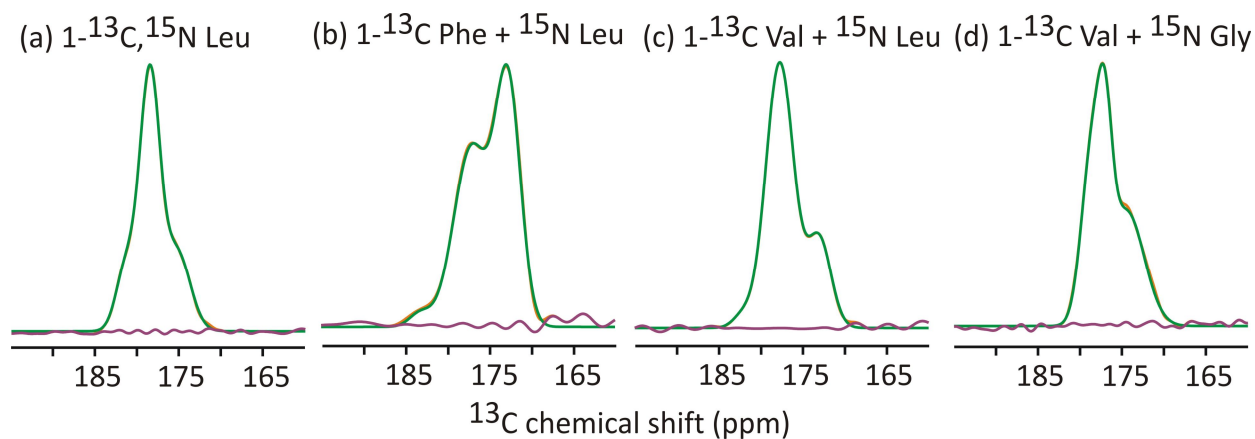


Figure 2-7: The fittings of S_0 deconvolutions for membrane associated Fgp41 samples are displayed. The labeling present in each sample is indicated. The experiment is shown in orange, the best-fit deconvolution sum is shown in green, and the difference is shown in purple. The best-fit deconvolution sum is the sum of the Gaussian curves shown previously in Figure 2-6.

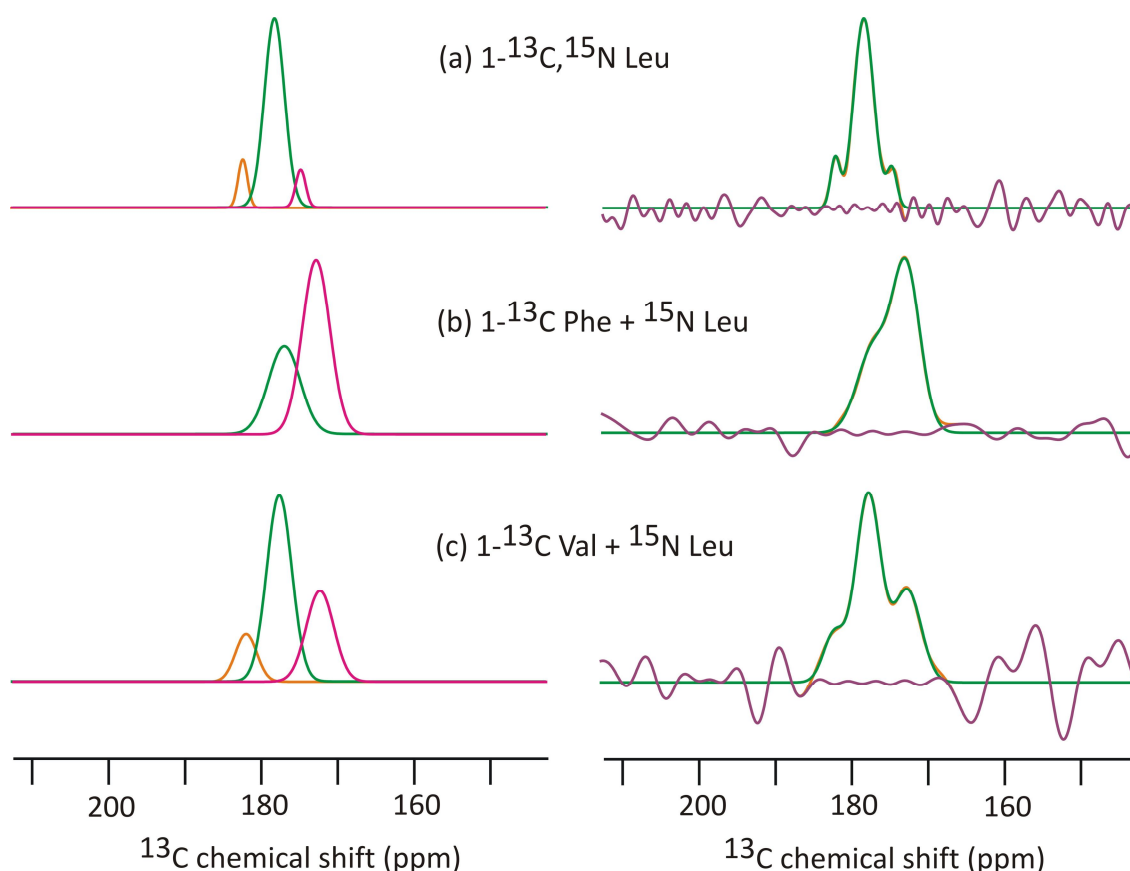


Figure 2-8: Deconvolutions of ΔS spectra are displayed. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.

Figure 2-6 e and f display spectra from samples that were labeled with $1\text{-}^{13}\text{C}$ Ala + ^{15}N Gly or $1\text{-}^{13}\text{C}$ Gly + ^{15}N Leu. The analyses are presented together because of the similar results.

The S_0 spectra were broad and featureless over the 170-185 ppm range so that deconvolution was not meaningful. This spectral breadth was understood by considering that although the fractional contribution of the labeled ^{13}C COs to the total S_0 intensity was ~ 0.8 , the labeled contribution from N- and C-helices in a SHB structure would be ~ 0.25 . About half of the S_0 intensity would be from labeled ^{13}C COs in the FP and loop regions. The earlier Figure 2-6a-d

analyses supported a mixture of helical and β strand/sheet shifts for FP ^{13}C O and broad signals are also expected from ^{13}C O in the less-ordered loop region. For the Figure 2-6e,f spectra, there were relatively good agreements between $(\Delta S/S_0)^{\text{calc}}$ and $(\Delta S/S_0)^{\text{exp}}$ and the ΔS spectra were respectively dominated by the A15 and G3 ^{13}C O. These ΔS spectra extended over 170-180 ppm and as with the V2 ΔS spectrum, the breadth correlated with being near one end of the FP region and with the spectral breadth observed for the corresponding residues in the membrane-associated HFP(14).

Discussion of Results of Fgp41 studies

The CD spectra and melting curves of purified Fgp41 support thermostable SHB structure and this structure was retained upon membrane binding as evidenced by a predominant sharp (3 ppm) helical ^{13}C O feature in the ΔS spectrum of Fgp41 produced with ^{13}C , ^{15}N Leu. This feature was assigned to the sum of ^{13}C O signals from six Leus which are in N- and C-helices in SHB structure. The SHB was also observed for the membrane-associated FP-Hairpin construct whose sequence was from a different HIV clade than Fgp41 and for which 46 contiguous residues including the native loop were replaced by a six non-native residues. By contrast, Fgp41 had the full native sequence of its clade. The similar results for Fgp41 and FP-Hairpin support the SHB as the final stable structure for membrane-associated gp41, Figure 1-1.

Fgp41 induced negligible inter-vesicle lipid mixing at pH 7.5 which correlated with the same result for FP-Hairpin. gp41 in the final SHB state may therefore be fusion-inactive at least with respect to lipid mixing which occurs early in either fusion of membranes of HIV and host

cells or in gp41-mediated cell-cell fusion. This view is supported by other fusion data showing that most membrane changes occur prior to formation of the final gp41 SHB state(18). For vesicles with negative charge, FP-Hairpin and related SHB gp41 constructs induce lipid mixing at pHs much lower than 7 (e.g. 4) and the pH-dependent functional difference has been correlated to changes in protein-membrane electrostatics (19). It is therefore likely that Fgp41 will also induce lipid mixing at these lower pHs. Over the past 25 years, there have been a series of experimental studies by different groups to determine whether HIV infects cells through direct fusion at the plasma membrane or through an endocytic mechanism(20, 21). In our view, the preponderance of data for either route support HIV-cell fusion at $\text{pH} \approx 7$ where SHB gp41 is fusion-inactive. There may be some differences among enveloped viruses as there is significant evidence for fusion activity of the folded influenza virus fusion protein ectodomain FHA2 (22, 23).

Relative to the sharp 3 ppm ΔS ^{13}C O signal from six Leu residues in the SHB, broader (4-10 ppm) ΔS ^{13}C O signals were observed from (typically) one residue in the FP. These breadths indicate conformational heterogeneity in the FP (24, 25). This point was further supported by the ΔS spectra of V7, F8, and F11 which were reasonably deconvolved into helical and β -sheet signals and indicated two populations of Fgp41 with distinct FP conformations. Helical and β sheet FP signals were also observed for membrane-associated FP-Hairpin samples even though there were differences between the Fgp41 and FP-Hairpin samples including: (1) two of the sixteen FP residues were different; (2) lipids were ester-linked (Fgp41) vs ether-linked (FP-Hairpin); (3) membrane reconstitution was based on detergent dialysis (Fgp41) vs simple mixing

of protein and vesicle solutions (FP-Hairpin); and (4) unfrozen Fgp41 vs frozen FP-Hairpin samples(8). Detection of helical and β sheet FP populations in both sample types strongly supports existence of these populations in membrane-associated gp41 in its final SHB state. In the future, it would be very interesting to study a larger gp41 construct that contains the transmembrane domain and for which there may be close contact between the FP and transmembrane domains.

MAS SSNMR structural studies of proteins are generally done by one of two approaches: (1) uniform ^{13}C and ^{15}N labeling, unambiguous assignment of most crosspeaks in multi-dimensional NMR spectra, and structural interpretation of the peak shifts and the crosspeak intensities of nuclei far apart in the sequence; or (2) specific (often residue or at least amino-acid type) labeling, and quantitative SSNMR measurements (e.g. shifts or dipolar couplings) to test specific structural models(26-28). The choice of approach for a particular protein depends on protein size and quantity as well as NMR linewidths. Approach (1) is more feasible for smaller proteins, high protein concentrations, and narrow (<1 ppm) linewidths. The present study is an example of approach (2) which was appropriate given the 162-residues, Fgp41:lipid ≈ 0.01 (with additional dilution of Fgp41 in the sample from water); the 3-10 ppm ^{13}CO linewidths; and the possibility of FP conformational heterogeneity (shown to be true in this study). The approach considered a model based on the existing high-resolution SHB structures of gp41 fragments and the extensive residue-specific SSNMR data for membrane-associated HFP.

Expanded Studies of Fgp41

Mutations to Fgp41 to Enhance Solubility

There are two Cys residues in the Fgp41 sequence that are separated by five residues. These Cys residues are likely on either side of the tip of the loop in the hairpin structure and therefore positioned to form an intramolecular disulfide bond(29). For the laboratory strain HXB2 sequence, the Cys residues have been mutated to Ala residues, as shown below in Figure 2-9. The unfolding temperature of the HXB2 Hairpin structure is 105 °C which should be within a few degrees of that of Fgp41, Figure 2-4b (8, 30). It is therefore unlikely that the disulfide bond of Fgp41 contributes appreciably to the thermostability of the hairpin structure of Fgp41. In addition, Fgp41 was initially quite difficult to purify given its low solubility in a variety of buffers. Sarkosyl was successfully used to solubilize the FHA2 protein which is largely similar to Fgp41, and one possible reason that Fgp41 was not able to be solubilized with sarkosyl is that the native Cys residues caused excessive aggregation of Fgp41 within inclusion bodies. In the FHA2 sequence, the native Cys residues had been mutated to Ala residues to avoid disulfide bond formation that could interfere with attempts to solubilize and purify the protein.

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGAASGKLICTSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKLEHHHHHH Stop

Second C to A mutation:

Forward primer: CTCATCGCCACCTCTTTTGTGC

Reverse primer: GCACAAAAGAGGTGGCGATGAG

DNA sequence of Fgp41 after second C to A mutation (Fgp41_{noCys}):

ATGGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGC
GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTGCAACAGCAAAGCAATTTGCT
GAAGGCTATAGAGGCTCAACAGCATCTGTTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAG
AGTCCTGGCTGTGGAAAGATACCTACAGGATCAACAGCTCCTGGGAATTTGGGGCGCCTCTGGAAAAC
CATCGCCACCTCTTTTGTGCCCTGGAACAATAGTTGGAGTAACAAGACTTATAATGAGATTTGGGACAAC
ATGACCTGGTTGCAATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTACTTGAAGAC
TCGCAGAACCAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAGATAAACTCGAGCACCACCACCAC
CACCCTGA

Protein sequence of Fgp41 after second C to A mutation (Fgp41_{noCys}):

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGAASGKLIATSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKLEHHHHHH Stop

After the above described mutations were performed to the Fgp41 plasmid, the plasmid for Fgp41_{noCys} was transformed into Rosetta2 *E. coli* competent cells (commercially available from EMD Chemicals). The Rosetta2 strain was chosen for its ability to process codons that correspond to rare tRNA in *E. coli*. Rosetta2 cells contain an extra plasmid to produce these rare tRNA molecules, which makes this particular strain of *E. coli* an excellent choice for recombinant expression of eukaryotic proteins such as Fgp41 and mutants. Unfortunately, a

rare codon analysis for the sequence of Fgp41 was not performed earlier, so all studies discussed previously only utilized the BL21(DE3) strain of *E. coli*.

Rare codon analysis of Fgp41 DNA sequence shows rare codons in *E. coli* underlined:

ATGGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGC
GTCAATGACGCTGACGGTACAGGCCCAGACAATTATTGTCTGGCATAGTGCAACAGCAAAGCAATTTGCT
GAAGGCTATAGAGGCTCAACAGCATCTGTTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAG
AGTCCTGGCTGTGGAAAGATACTACAGGATCAACAGCTCCTGGGAATTTGGGGCTGCTCTGGAAAACT
CATCTGCACCTCTTTTGTGCCCTGGAACAATAGTTGGAGTAACAAGACTTATAATGAGATTTGGGACAAC
ATGACCTGGTTGCAATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTACTTGAAGAC
TCGCAGAACCCAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAGATAAACTCGAGCACCACCACCAC
CACCCTGA

Expression and Purification of Fgp41_{noCys}

Expression of the mutated construct was performed as previously described. Several different purification attempts of Fgp41_{noCys} were performed, most notably an initial attempt to purify Fgp41_{noCys} in buffers lacking detergent.

Purification #1

5.0 grams of cells induced to express Fgp41_{noCys} were sonicated in 40 mL of buffer containing 50 mM sodium phosphate at pH 8.0, 300 mM NaCl, and 20 mM imidazole. The lysate was centrifuged at 50000g for 20 minutes at 4°C, and the supernatant was combined with 0.25 mL of prepared His-Select cobalt resin. After one hour of mixing at room temperature, the resin was loaded onto a column and washed with 3 mL of fresh lysis buffer. Protein was eluted from the resin with a buffer differing only in [imidazole] = 250 mM. An intense band (estimated ~40%

of the total protein) corresponding to Fgp41_{noCys} was observed by SDS-PAGE as shown below in Figure 2-10. The purity of this band could be increased by washing with buffers including increasing [imidazole] as in the purification protocol for Fgp41.

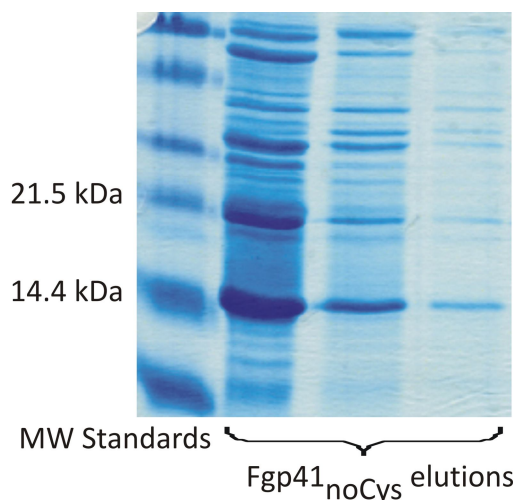


Figure 2-10: An initial attempt at purification of Fgp41_{noCys} involved solubilizing the protein in a buffer containing no detergent. There was no detectable band in earlier attempts to solubilize Fgp41 under the same conditions (data not shown).

Purification #2

The insoluble material from Purification #1 (obtained as a pellet after centrifugation of the lysate) was sonicated in 40 mL of urea lysis buffer, which contained 50 mM sodium phosphate at pH 8.0, 300 mM NaCl, 20 mM imidazole, and 8 M urea. The lysate was centrifuged at 50000g for 20 minutes at 4°C, and the supernatant was combined with 0.50 mL of prepared His-Select cobalt resin. More resin was used for this purification because it was likely that more recombinant protein was solubilized by sonication in urea. After one hour of mixing at room temperature, the resin was loaded onto a column and washed with 6 mL of fresh lysis buffer. The washes were done until the A₂₈₀ reading of the eluent was small and constant (about 0.2

mg/mL). Protein was eluted from the resin with urea elution buffer (50 mM sodium phosphate at pH 8.0, 300 mM NaCl, 250 mM imidazole, and 8 M urea). Approximately 1.5 mg of pure Fgp41_{noCys} was obtained from this purification. One advantage of a purification utilizing urea to solubilize the protein rather than SDS is that urea is easily removed through dialysis, where SDS is difficult to remove. An SDS-PAGE gel of the Fgp41_{noCys} obtained in elutions is shown below in Figure 2-11.

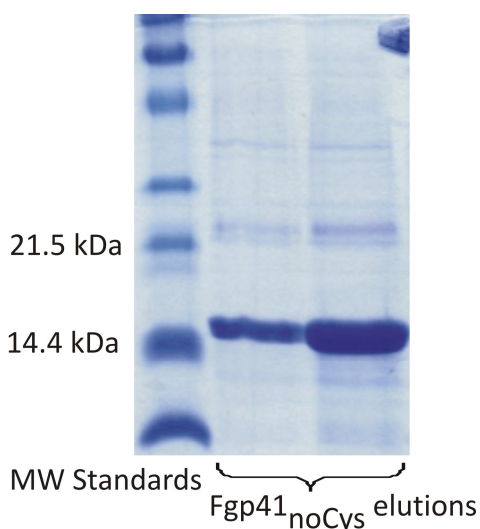


Figure 2-11: Purification of the insoluble fraction of protein using urea resulted in ~95% pure Fgp41_{noCys} in elution fractions. The yield of this particular purification was estimated as ~1.5 mg pure protein per 5 grams of cells.

Future Work

1. To gain a better understanding of the relationship between protein conformation, fusion activity, and pH effects in the context of gp41, a shorter version of Fgp41 could be engineered. I would propose to engineer a construct that models N70, containing the fusion peptide through the end of the N-helix of gp41. N70 exhibits high lipid mixing activity at physiological pH, while Fgp41 showed ~2% lipid mixing activity under the same assay conditions(1, 31). By creating a

construct that consists of the same regions as N70, we would be able to better understand how the presence of the six-helix bundle affects lipid mixing ability of the protein.

An N70-like construct could be made by simply introducing a stop codon into the plasmid DNA for Fgp41 at the desired position (likely at the end of the N-helix). This would allow the beginning portion of the protein to be expressed, though the protein would be truncated to only include those residues before the stop codon. At this point, the produced protein could be purified utilizing HPLC, or further mutations to the plasmid DNA could be performed to introduce a poly-histidine tag or MAT tag before the stop codon to utilize IMAC based methods of protein purification(32).

2. Fgp41 could be engineered to model the proposed “pre-hairpin intermediate” (PHI) conformation of gp41. By disrupting hydrophobic contacts between the N- and C-helices, it would be possible to assay activity of the proposed PHI. Crystallographic studies have suggested hydrophobic contacts involving residues Ile559, Val570, and Ile573 of the N helix and residues Trp631, Ile635, and Ile646 of the C helix(33). The hydrophobic contacts between the N and C helices could be disrupted by mutating the implicated residues in the C helix to Ala. Disrupting these contacts should lead to a much less thermostable structure than the SHB, and this could be investigated by performing a melt in the CD instrument on successively mutated constructs. If the hydrophobic contacts were disrupted, one would expect to observe a thermal transition much below the ~110°C observed for the hairpin structure.

REFERENCES

REFERENCES

1. Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) Solid-State Nuclear Magnetic Resonance (NMR) Spectroscopy of Human Immunodeficiency Virus gp41 Protein That Includes the Fusion Peptide: NMR Detection of Recombinant Fgp41 in Inclusion Bodies in Whole Bacterial Cells and Structural Characterization of Purified and Membrane-Associated Fgp41, *Biochemistry* 50, 10013-10026.
2. Davis, C. B., Dikic, I., Unutmaz, D., Hill, C. M., Arthos, J., Siani, M. A., Thompson, D. A., Schlessinger, J., and Littman, D. R. (1997) Signal transduction due to HIV-1 envelope interactions with chemokine receptors CXCR4 or CCR5, *Journal of Experimental Medicine* 186, 1793-1798.
3. Painter, S. L., Biek, R., Holley, D. C., and Poss, M. (2003) Envelope variants from women recently infected with clade A human immunodeficiency virus type 1 confer distinct phenotypes that are discerned by competition and neutralization experiments, *J. Virol.* 77, 8448-8461.
4. Hill, C. M., and Littman, D. R. (1996) AIDS - Natural resistance to HIV?, *Nature* 382, 668-669.
5. Painter, S. L., Biek, R., Holley, D. C., and Poss, M. (2003) Envelope variants from women recently infected with clade A human immunodeficiency virus type 1 confer distinct phenotypes that are discerned by competition and neutralization experiments, *Journal Of Virology* 77, 8448-8461.
6. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in *E. coli*, purification, and refolding of the full ectodomain of the Influenza virus membrane fusion protein, *Prot. Expr. Purif.* 61, 212-219.
7. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, *Biochemistry* 48, 2714-2722.
8. Sackett, K., Nethercott, M. J., Epand, R. F., Epand, R. M., Kindra, D. R., Shai, Y., and Weliky, D. P. (2010) 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.* 397, 301-315.
9. Brugger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F. T., and Krasslich, H. G. (2006) The HIV lipidome: A raft with an unusual composition, *Proc. Natl. Acad. Sci. U.S.A.* 103, 2641-2646.
10. Gullion, T., and Schaefer, J. (1989) Rotational-echo double-resonance NMR, *J. Magn. Reson.* 81, 196-200.

11. Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR, *J. Magn. Reson.* 162, 479-486.
12. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts, *J. Biomol. NMR* 25, 173-195.
13. Yang, J. (2003) Ph. D. Dissertation, Michigan State University, East Lansing, MI.
14. Yang, J., Gabrys, C. M., and Weliky, D. P. (2001) Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide, *Biochemistry* 40, 8126-8137.
15. Zheng, Z., Yang, R., Bodner, M.L., and Weliky, D.P. (2006) Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy, *Biochemistry* 45, 12960-12975.
16. Qiang, W., Bodner, M. L., and Weliky, D. P. (2008) 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.* 130, 5459-5471.
17. Schmick, S. D., and Weliky, D. P. (2010) Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide, *Biochemistry* 49, 10623-10635.
18. Markosyan, R. M., Cohen, F. S., and Melikyan, G. B. (2003) HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation, *Mol. Biol. Cell* 14, 926-938.
19. Sackett, K., TerBush, A., and Weliky, D. P. HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion, *European Biophysics Journal With Biophysics Letters* 40, 489-502.
20. Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions, *J. AIDS* 3, 965-974.
21. Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G. B. (2009) HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes, *Cell* 137, 433-444.
22. Curtis-Fisk, J., Preston, C., Zheng, Z. X., Worden, R. M., and Weliky, D. P. (2007) Solid-state NMR structural measurements on the membrane-associated influenza fusion protein ectodomain, *J. Am. Chem. Soc.* 129, 11320-11321.

23. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* **286**, 13226-13234.
24. Grasnick, D., Sternberg, U., Strandberg, E., Wadhwani, P., and Ulrich, A. S. (2011) Irregular structure of the HIV fusion peptide in membranes demonstrated by solid-state NMR and MD simulations, *Eur. Biophys. J.* **40**, 529-543.
25. Tristram-Nagle, S., Chan, R., Kooijman, E., Uppamoochikkal, P., Qiang, W., Weliky, D. P., and Nagle, J. F. (2010) HIV fusion peptide penetrates, disorders, and softens T-cell membrane mimics, *J. Mol. Biol.* **402**, 139-153.
26. Tycko, R. (2006) Molecular structure of amyloid fibrils: insights from solid-state NMR *Quarterly Reviews of Biophysics* **39**, 1-55.
27. McDermott, A. (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR, *Ann. Rev. Biophys.* **38**, 385-403.
28. Fowler, D. J., Weis, R. M., and Thompson, L. K. (2010) Kinase-active signaling complexes of bacterial chemoreceptors do not contain proposed receptor-receptor contacts observed in crystal structures, *Biochemistry* **49**, 1425-1434.
29. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* **17**, 4572-4584.
30. Lev, N., Fridmann-Sirkis, Y., Blank, L., Bitler, A., Epand, R. F., Epand, R. M., and Shai, Y. (2009) Conformational stability and membrane interaction of the full-length ectodomain of HIV-1 gp41: Implication for mode of action, *Biochemistry* **48**, 3166-3175.
31. Sackett, K., TerBush, A., and Weliky, D. P. (2011) HIV gp41 six-helix bundle constructs induce rapid vesicle fusion at pH 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for HIV-host cell fusion, *Eur. Biophys. J.* **40**, 489-502.
32. Watson, N., Davis, R. L., Zobrist, J. M., Stephan, J., Scott, M., Davis, G., Mehig, R. J., and Kappel, W. K. (2007) The MAT-Tag system: Versatile for recombinant protein purification and expression, *Biotechniques* **42**, 768-768.
33. Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12303-12308.

Chapter 3 – Development of a quantitative method of recombinant protein expression in whole E. coli cells and bacterial inclusion bodies

Introduction

Recombinant protein expression in bacteria is a method heavily utilized to produce large amounts of proteins for structural and functional studies. For those working with membrane proteins or other insoluble proteins, solubilization of the proteins can be difficult as these proteins are often sequestered within bacterial inclusion bodies. Inclusion bodies are large insoluble aggregates of protein, where little is known about the structure of the protein within. Inclusion bodies are often difficult to solubilize, and as a result the proteins are difficult to purify to high yields. With this situation, it is difficult to tell whether the target protein is not being produced at a high level within the cells, or is just not well solubilized.

Previous methods of quantifying recombinant protein have been suggested. One method utilizes SDS-PAGE and scanning laser densitometry, though it requires multiple samples of pure protein with known concentrations to quantify recombinant protein from a fermentation culture(1). FT-IR was also proposed as a high-throughput method to quantify recombinant protein expression in whole cells, but this method relied heavily on the shift of the amide I band into the β -strand region to indicate the presence of protein in inclusion bodies; the method also required advanced data analysis consisting of multivariate calibration utilizing 23 different samples and multiple principle component plots (2). Additionally, previous work from our group has shown that proteins can retain native α -helical structure within inclusion bodies, which suggests the FT-IR method may not be applicable to all recombinant proteins in inclusion bodies(3).

We have developed a solid-state nuclear magnetic resonance (SSNMR) method to detect recombinant protein expression levels within *E. coli* by taking spectra of either whole bacterial cells or insoluble cell pellet (ICP). For a 40 μ L rotor volume, (estimate sample density of an ICP as ~ 1.2 g/mL) there will be ~ 50 mg of sample. The ICP is primarily comprised of insoluble proteins and lipids, as well as cell organelles. From our NMR data and calculations (shown later in this chapter and based on Fgp41 data) we can estimate that ~ 3 mg of recombinant protein is present within the sample ($\sim 6\%$ of the mass of the ICP sample then is comprised of recombinant protein within inclusion bodies). The method utilizes small sample volumes (25 – 50 mL of bacterial cell culture), 20 – 40 mg of isotopically labeled amino acids per sample, moderate NMR fields (9.4 Tesla), is quick (less than 2 days total for sample preparation and analysis by NMR spectroscopy) and straightforward.

The REDOR (rotational echo double resonance) pulse sequence was utilized in this work because of its utility as a filter(4). In REDOR experiments, two different spectra are acquired: S_0 represents the full ^{13}C spectrum containing signals from all ^{13}C nuclei in the sample, while S_1 represents the ^{13}C spectrum of all nuclei not directly bonded to ^{15}N nuclei. By subtracting S_1 from S_0 , we can obtain a spectrum representative of signals from all ^{13}C nuclei that are directly bonded to ^{15}N nuclei. To determine whether or not a recombinant protein is being produced within the bacterial system, a ΔS spectrum ($S_0 - S_1$) should be obtained. We utilize a labeling scheme that should detect a unique sequential pair of amino acids (XY) within the protein sequence. By labeling all ^{13}C of the first amino acid type in the pair (X), and ^{15}N labeling all of

the second amino acid type in the pair (Y), we can obtain one position where the two nuclei (^{13}C and ^{15}N of X and Y, respectively) are chemically bonded. If the recombinant protein is produced, the ΔS spectrum will show one spectral feature within the carbonyl region of the spectrum (corresponding to the ^{13}C from the X residue). An example of a ΔS spectrum is shown in Figure 3-1a. This feature corresponds to the one position where the ^{13}C labeled residue is followed by the ^{15}N labeled residue. If the linewidth of the peak is narrow, conformational homogeneity of the protein's secondary structure at that position is inferred. With samples that give narrow difference spectrum peaks, we can compare the chemical shift to a reference database to predict the likely secondary structure of the protein at that residue(5). Previous work in our group utilized a similar method to determine the structure of different proteins within bacterial inclusion bodies, however for these experiments we are using the ΔS spectrum primarily to decide whether a protein is being produced or not(3).

Protein Construct Information

To test the generality of the application of REDOR SSNMR to detect recombinant protein expression in whole *E. coli* cells as well as bacterial inclusion bodies, a variety of protein constructs, plasmid types, and strains of *E. coli* were utilized in these studies. The plasmid, target protein, and *E. coli* strains used are outlined in Table 3-1.

Table 3-1: Protein construct information. The name of the protein construct, plasmid type, and *E. coli* cell type used are listed for each protein.

Protein Construct	Plasmid Type	Cell Type
human proinsulin	pQE-31	BL21(DE3)
Hairpin	pGEMT	BL21(DE3)
Fgp41	pET24a+	Rosetta2
Fgp41+	pET24a+	Rosetta2
FHA2	pET24a+	Rosetta2
Control (no protein insert)	pET24a+	Rosetta2

Listed below are the amino acid sequences for the recombinant protein inserts within the plasmids. Some of the proteins include polyhistidine tags to enable affinity purification.

Human Proinsulin (HPI)

GSSHHHHHHSSGLDPVLMFVNQHLCGSHLVEALYLVCGERGFFYTPKTRRE
AEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSIKSLYQLENY
CN

Hairpin

CTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILSGGR
GGWMEWDREINNYTSLIHSLIEESQNQQEKNEQEELLELDKW

Fgp41

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQSNLLKAIEA
QQHLLKLTWVGKQLQARVLAVERYLQDQQLLGIWGASGKLIATSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKLEHHHHHH

Fgp41+

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVHQQSNLLKAIEA
QQHLLKLTWVGKQLQARVLAVERYLQDQQLLGIWGASGKLIATSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKWANLWNWFSITNWLWYIKLEHHHHHH

FHA2

GLFGAIAAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
KLN RVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALE
NQHTIDLTDSMNKLF EKTRRLRENAEEMGNGSFKIYHKADNAAIESIRN
GTYDHDVYRDEALNNRFQIKGV ELKSGYKDWVEHHHHHH

Sample Preparation

Protein Expression

One 250 mL flask containing 100 mL of LB and the proper antibiotic was inoculated with 0.5 mL of a glycerol stock of *E. coli* cells (containing a plasmid for recombinant protein expression). The flask was placed in an incubator shaker with shaking at 180 rpm and a temperature of 37°C. After ~16 hours, the cells were harvested by centrifugation at 10,000 *g* / 4°C / 10 minutes. The cells were then resuspended into a baffled flask containing 50 mL of M9 minimal media, antibiotic, 100 µL of 1.0 M MgSO₄, and 250 µL of 50% v/v glycerol. After approximately one hour of shaking at 180 rpm and 37°C (once log phase growth was reached) the *E. coli* were induced to express recombinant protein by addition IPTG to a concentration of 2.0 mM. Then an amino acid mixture containing both unlabeled amino acids and isotopically labeled amino acids was added to the media. 10 mg of each amino acid (either labeled or unlabeled) was contained within the mixture. One hour later, another dose of the same amino acid mixture was added to the media. An expression period of 3 hours with shaking at 37°C was utilized. At the end of the expression period, the cells were once again harvested by centrifugation at 10,000 *g* / 4°C / 10 minutes. The cell pellets were stored at -20°C until they were prepared for the NMR experiments. For whole cell NMR experiments, the cell pellets were lyophilized.

Suppression of Isotopic Label Scrambling

NMR experiments were set up as described in the NMR Experiment section to study dehydrated whole bacterial cell samples. Initially, 10 mg of each labeled amino acid was added to the minimal medium at the time of induction with glycerol present as the only other carbon source. When a protein sample was “mis-labeled” (i.e. for FHA2, the 1-¹³C Ala and ¹⁵N Val were

added to the medium, but there is no AV sequential pair of amino acids in the sequence of FHA2) we actually observed unexpected dephasing in the S_1 spectrum, indicating that ^{15}N nuclei were bonded to the $1\text{-}^{13}\text{C}$ Ala. Since the labeling scheme should have prevented this, our conclusion is that the ^{15}N label from Val was shuffled into other amino acid types, which were then incorporated into the protein. The difference spectrum for this experiment is shown below in Figure 3-1a. When a dose of 10 mg each of every amino acid (labeled $1\text{-}^{13}\text{C}$ Ala and ^{15}N Val and all other amino acids unlabeled) was added to the culture at the time of induction, and another dose one hour later, the dephasing was suppressed(6, 7). The difference spectrum for this result is shown below in Figure 3-1b. In conclusion, we were able to utilize product feedback inhibitory loops of the E. coli amino acid metabolic pathways and suppress isotopic label scrambling by supplementing growth medium with all amino acids.

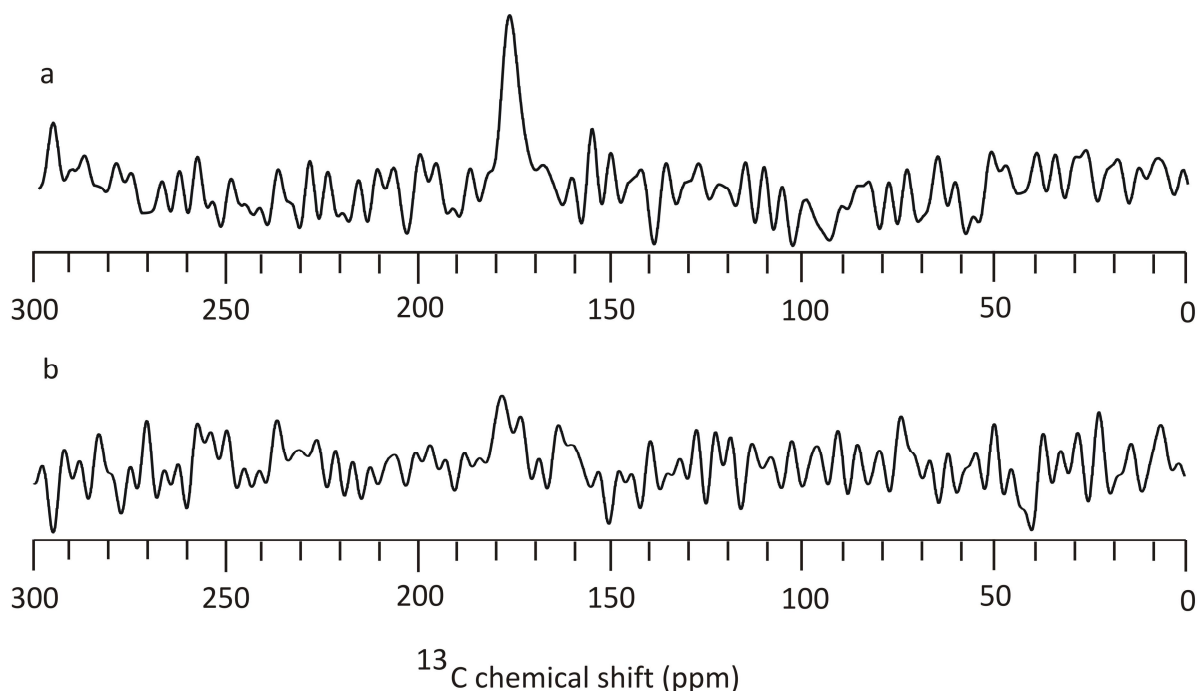


Figure 3-1: ΔS spectra for a) $1\text{-}^{13}\text{C}$ Ala, ^{15}N Val labeled dry whole *E. coli* cells induced to produce FHA2 with glycerol as the only other carbon source, and b) $1\text{-}^{13}\text{C}$ Ala, ^{15}N Val labeled dry whole *E. coli* cells induced to produce FHA2 where the growth medium was supplemented with all unlabeled amino acids as well as glycerol. Each ΔS spectrum was the result of a) 46652 ($S_0 - S_1$) scans and b) 43647 ($S_0 - S_1$) scans. The spectra were processed with no line broadening and a) 5^{th} order and b) 7^{th} order polynomial baseline corrections.

In order to investigate the precision of the integrated signal intensities, the signal intensity was integrated in regions of each spectrum that do not contain spectral features. This allows for determination of how much variation in signal intensity can be attributed to spectral noise. 13 regions of noise (in 15 ppm sections) were integrated for each spectrum. The values of integrated signal intensity are reported in Table 3-2. In addition, the integrated signal intensity in the carbonyl region is reported for each spectrum.

Table 3-2: Integrated signal intensities in 15 ppm regions from spectra corresponding to either whole bacterial cells induced to express FHA2 that had been 1-¹³C Ala, ¹⁵N Val labeled with glycerol present as the only additional carbon source in the growth medium, or whole bacterial cells induced to express FHA2 that had been 1-¹³C Ala, ¹⁵N Val labeled with glycerol and all other unlabeled amino acids present in the growth medium. The Ala-Val sequential pair of amino acids does not appear within the FHA2 protein sequence.

range of spectrum integrated	integrated signal intensities over 15 ppm ranges	
	all amino acids added	only labeled amino acids added
carbonyl (170 → 185 ppm)	21.341	58.1279
400 → 385 ppm	5.7787	2.6401
380 → 365 ppm	5.7584	-14.3034
360 → 345 ppm	-22.852	-0.0376
340 → 325 ppm	-3.7395	-4.2526
320 → 305 ppm	2.6227	-6.1709
300 → 285 ppm	-17.4505	17.543
280 → 265 ppm	-12.6827	-8.8076
240 → 225 ppm	9.506	-12.7001
220 → 205 ppm	-10.7618	-8.9171
0 → -15 ppm	-7.8094	-2.5523
-20 → -35 ppm	2.4773	-2.0299
-40 → -55 ppm	2.4773	4.2796
-60 → -75 ppm	-1.3408	-0.3178

The standard deviations in integrated signal intensity were calculated to be 9.9 for cells supplemented with all amino acids and 8.3 for cells supplemented with glycerol. By comparing the standard deviations of the noise integrals using the statistical F test, the difference between the calculated standard deviations are not found to be statistically significant at the 95% confidence level (with 12 degrees of freedom for each data set, the critical value of F is 2.69, where 1.43 was calculated from the data sets)(8). This ensures that the variation in the noise should not affect analysis of the spectra.

In conclusion, for cells that were supplemented with the additional, unlabeled amino acids, the integrated signal intensity in the carbonyl region can be expressed as 20 ± 10 while it is calculated as 58 ± 8 for cells that were not supplemented with unlabeled amino acids. This data, along with the difference spectra shown in Figure 3-1, supports that we have substantially limited the amount of difference signal that will be observed for a sample where there is not a sequential pair of amino acids labeled by simply supplementing the growth medium with all amino acids to prevent conversion between amino acid types.

NMR Sample Preparation for Insoluble Cell Pellet Experiments

Each cell pellet was combined with ~40 mL PBS (pH 7.3) and placed on ice. Lysis of the *E. coli* cells was achieved by sonication with a tip sonifier (using 4 one minute cycles, 80% amplitude, 0.8 seconds on, 0.2 seconds off). After sonication, the samples were centrifuged at $50,000\ g / 4^{\circ}\text{C} / 20$ minutes. The supernatant of each sample (containing soluble proteins) was discarded, and the pellet was packed into a 4 mm solid state NMR magic angle spinning rotor. The active sample volume of the rotor was approximately 40 μL .

NMR Sample Preparation for Whole Cell Experiments

The cell pellets obtained after the expression period were lyophilized overnight to remove all water. The pellets were ground into a fine powder with a mortar and pestle and packed into a 4 mm solid state NMR magic angle spinning rotor.

NMR Experimental Parameters

The following parameters were used for all samples. Data were obtained with a 9.4 T instrument (Agilent Infinity Plus) and a triple-resonance MAS probe whose rotor was cooled with nitrogen gas at -20°C . Experimental parameters included: (1) 8.0 kHz MAS frequency; (2)

5 μ s ^1H $\pi/2$ pulse and 2 ms cross-polarization time with 50 kHz ^1H field and 70-80 kHz ramped ^{13}C field; (3) 1 ms rotational-echo double-resonance (REDOR) dephasing time with a 9 μ s ^{13}C π pulse at the end of each rotor period except the last period and for some data, a 12 μ s ^{15}N π pulse at the center of each rotor period; and (4) ^{13}C detection with 90 kHz two-pulse phase modulation ^1H decoupling (which was also on during the dephasing time); and (5) 0.8 sec pulse delay. Data were acquired without (S_0) and with (S_1) ^{15}N π pulses during the dephasing time and respectively represented the full ^{13}C signal and the signal of ^{13}C s not directly bonded to ^{15}N nuclei. The $S_0 - S_1$ (ΔS) difference signal was therefore dominated by the labeled ^{13}CO s in the sequential pairs targeted by the labeling. Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm so that the ^{13}CO shifts could be directly compared to those of soluble proteins(9).

Whole Cell SSNMR Spectroscopy

The bacterial growth and Fgp41 expression conditions were very similar to those used for FHA2, a construct corresponding to the full-length ectodomain (including fusion peptide) of the influenza virus HA2 fusion protein(10). Like Fgp41, FHA2 had a C-terminal hexahistidine tag and bacterial cell lysis and protein solubilization in buffer containing N-lauroylsarcosine detergent followed by affinity chromatography resulted in 10 mg purified FHA2/L culture. In some contrast, application of this protocol to cells induced to synthesize Fgp41 gave only 0.1

mg Fgp41/L culture. It was unclear whether the poor yield was due to low Fgp41 expression or to poor Fgp41 solubilization by the detergent.

The FHA2 solubilized by detergent was likely initially associated with the cell membrane. It was also shown that a much larger fraction of FHA2 was not solubilized by detergent and was likely constituted in inclusion bodies. The poor yield of Fgp41 might therefore be due to dominant incorporation in inclusion bodies. The molecular structure of FHA2 in inclusion bodies had been probed by: (1) adding specific ^{13}C and ^{15}N labeled amino acids immediately prior to induction; and (2) recording REDOR SSNMR spectra of the whole cells after induction so that the filtered ΔS signal corresponded to the ^{13}C of a targeted residue in FHA2; and (3) correlation of the experimental peak ^{13}C shift to local conformation at this residue (3). A modified approach was applied to cells induced to express Fgp41 with the goal of assessing Fgp41 production. Addition of 10 mg of 1- ^{13}C , ^{15}N Leu to 50 mL culture just prior to induction of expression targeted the 24 Leus and 6 LL repeats in the Fgp41 sequence, Figure 3-2.

1 AVGLGAVFLG FLGAAGSTMG AASMTLTVQA RQLLSGIVQQ QSNLLKAIEA QQHLLKLTWW
61 GIKQLQARVL AVERYLQDQQ LLGIWGCSGK LICTSFVPWN NSWSNKTYNE IWDNMTWLQW
121 DKEISNYTDT IYRLLED SQN QQEKNEQDLL ALDKLEHHHH HH

Figure 3-2: Amino acid sequence of the Fgp41 protein construct. The LL pairs targeted with the 1- ^{13}C , ^{15}N Leu labeling are bolded in the sequence. The fusion peptide region is shown in blue, the N helix and C helix in red and green, respectively. All LL pairs are located either within or right at the end of the helical regions of the protein.

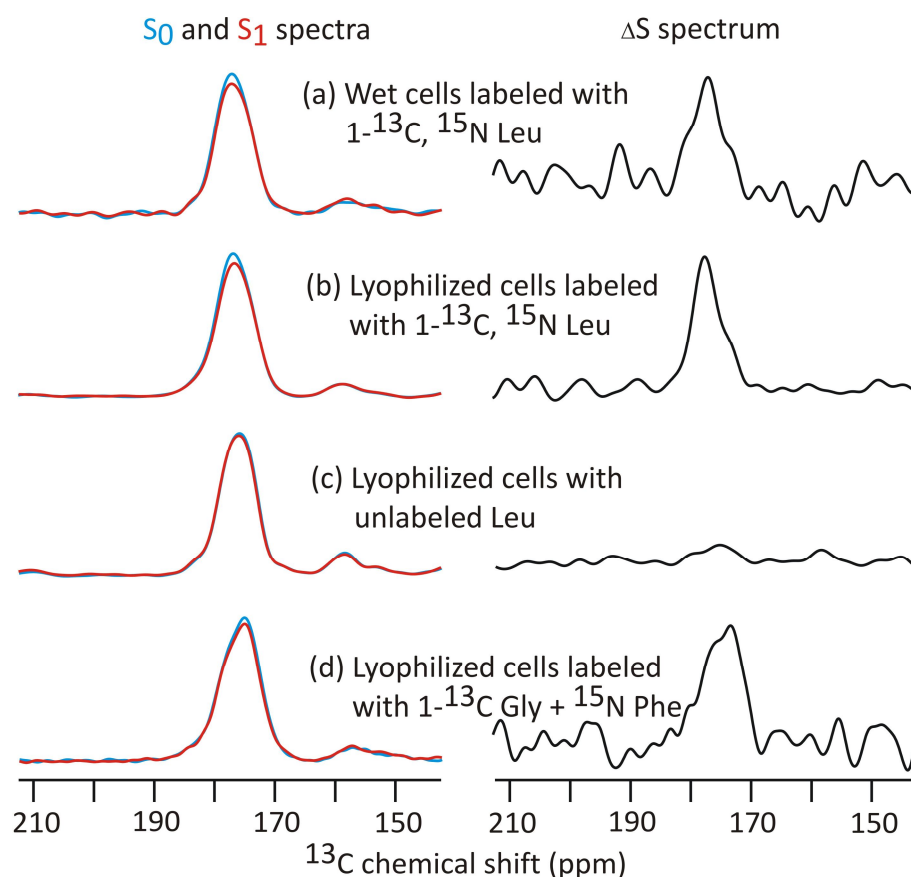


Figure 3-3: REDOR ^{13}C CO NMR spectra of whole bacterial cells induced to produce Fgp41 by sequential steps: (1) growth in rich medium, (2) growth in minimal medium, (3) addition of labeled or unlabeled amino acids, (4) induction of Fgp41 expression, (5) centrifugation. The induction temperature and duration were either (a-c) 23 °C and ~2 hr or (d) 37 °C and ~5hr. The left panels display S_0 (blue) and S_1 (red) spectra and the right panels display ΔS spectra. The REDOR dephasing time was either (a-c) 1 ms or (d) 2 ms. For panels a, b, and d, the dominant contribution to each ΔS spectrum was from residues with labeled ^{13}C CO groups that were directly bonded to ^{15}N atoms. These residues were (a and b) L33, L44, L54, L81, L134, and L149 of the LL sequential pairs of Fgp41 and (d) G10 of the G10-F11 unique sequential pair. Each S_0 or S_1 spectrum was processed with 100 Hz Gaussian line broadening, and each ΔS spectrum was processed with either (a and d) 200 or (b and c) 100 Hz line broadening. Polynomial baseline correction (typically fifth order) was applied to each spectrum. Each S_0 or S_1 was the sum of (a) 100000, (b) 100000, (c) 127222, or (d) 48448 scans.

The isotropic ^{13}C CO regions of the REDOR S_0 , S_1 , and ΔS spectra of the cells are displayed in Figure 3-3. The Figure 3-3a sample was an aliquot of the wet cell pellet obtained after the induction period and subsequent centrifugation and the Figure 3-3b sample was an aliquot of

this whole cell pellet that had been lyophilized. The spectra were similar for both wet and lyophilized cells with ~4 times greater signal-per-scan in the lyophilized cell sample because this sample had a higher fraction of non-aqueous cell mass. For either sample type, the intensity of the S_1 spectrum was reduced relative to S_0 . This supported the presence of LL repeats in the protein produced during the induction period and correlated with the 6 LLs in the Fgp41 sequence. The ΔS spectra had prominent signals in the ^{13}CO region and these were the only signals detectable above the noise. Control cells were produced using unlabeled rather than labeled Leu. The resultant NMR spectra are displayed in Figure 3-3c and had comparable S_0 and S_1 intensities with little ^{13}CO ΔS signal. This provided further support that the ^{13}CO ΔS signal from the labeled cells could be ascribed to LL repeats in protein produced during expression. Cells were also labeled with 1- ^{13}C Gly and ^{15}N Phe which targeted the 11 Glys in the Fgp41 sequence and the single GF pair at G10-F11. The resulting NMR spectra are displayed in Figure 3-3d and included a prominent ^{13}CO ΔS signal that was consistent with Fgp41 production.

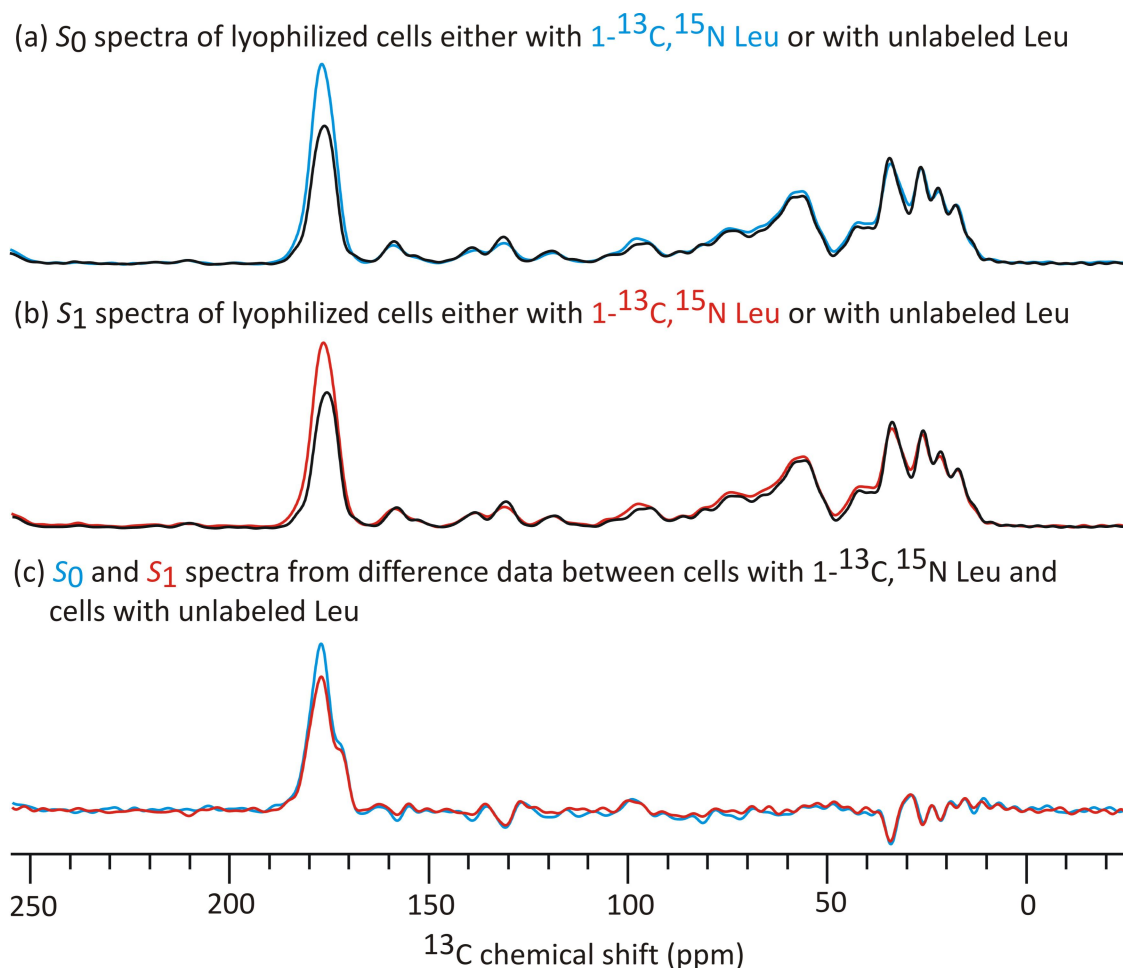


Figure 3-4: REDOR ^{13}C NMR spectra of lyophilized whole bacterial cells induced to produce Fgp41 with either $1\text{-}^{13}\text{C}, ^{15}\text{N}$ labeled Leu or unlabeled Leu. The cell production and NMR parameters are described in the legend of Figure 3-3. Panel a displays the S_0 spectra of the labeled (blue) and unlabeled (black) cells with the relative intensities adjusted to yield the best agreement in the 0 to 90 ppm region, as this region should be unaffected by labeling. The incorporation of the labeled Leu synthesized during the induction period is evidenced by the larger ^{13}C CO intensity for the labeled cell spectrum. Panel b displays the S_1 spectra of the labeled (red) and unlabeled (black) cells. Panel c displays the S_0 (blue) and S_1 (red) spectra processed from the difference NMR data: labeled cells – $0.75 \times$ unlabeled cells. The 0.75 factor reflects the ratio of the number of scans summed for the labeled cells relative to number for the unlabeled cells and resulted in a minimal signal in the 0 to 90 ppm region. The spectra in panel c are representative of the $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu incorporated into the cellular protein. Spectra were processed with no line broadening and a 5^{th} order polynomial baseline correction.

Both labeled and natural abundance ^{13}C Os contribute to the S_0 and S_1 NMR signals of the labeled whole cells. Figure 3-4a(b) provides quantitative assessment of these two contributions and shows the full $S_0(S_1)$ spectra of the Leu-labeled and unlabeled cells. In each panel, the two spectra were scaled to have equal intensity in the 0-90 ppm region because this region should be unaffected by labeling. The ratio of the unlabeled to labeled scaling factors was ~ 0.75 and matched the ratio of numbers of scans summed for the labeled vs unlabeled samples. This matching was expected because the signal intensities of individual scans were approximately equal to each other so the sum signal intensity increased linearly with number of scans. For panel a(b), the difference between the intensities in the ^{13}C O region was the labeled Leu contribution to the $S_0(S_1)$ signal. For these labeled Leu ^{13}C Os, there was smaller S_1 intensity relative to S_0 . This is shown more clearly in Figure 3-4c which displays the S_0 and S_1 spectra processed from labeled cell data – $(0.75 \times \text{unlabeled cell data})$. For labeled Leu in the cells, the normalized experimental dephasing $(\Delta S/S_0)^{\text{exp}} = 0.13 \pm 0.01$ and was determined from the ^{13}C O S_0 and S_1 intensities in panel c.

The following model and analysis support that most of the labeled Leu was in Fgp41, i.e. Fgp41 was the dominant protein produced during expression. Consider the model: (1) The 24 Leus of Fgp41 are ^{13}C O, ^{15}N labeled; (2) the ^{13}C Os of the N-terminal Leus of the 6 LL repeats (directly bonded to ^{15}N) have S_1/S_0 intensity ratio = 0.3; and (3) the other 18 Leu ^{13}C Os have $S_1/S_0 = 1.0$. Points (2) and (3) are based on earlier experiments and simulations (11). For the

Fgp41 Leu ^{13}C O, the $(\Delta S/S_0)^{\text{calc}} = [24 - 18 - (6)(0.3)]/24 = 0.17$ which is close to $(\Delta S/S_0)^{\text{exp}}$ and supports dominant production of Fgp41 during the induction period.

Analysis of the SSNMR spectra of lyophilized whole cells

Comparison of the ΔS spectrum of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled cells, Figure 3-3b, to the ΔS spectrum of the unlabeled cells, Figure 3-3c, shows a clear effect from using labeled Leu. The “labeled cell difference” S_0 (S_1) spectrum, Figure 3-3c, is the difference between the S_0 (S_1) spectra of the labeled and unlabeled cells and shows only the contribution of the labeled Leu. Deconvolution was applied to the labeled cell difference S_0 spectrum and to the labeled cell ΔS spectrum. Both spectra were well-fitted to the sum of three Gaussian line shapes, see Table 3-3 and Figure 3-5, and were dominated by the $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu incorporated into cell protein produced during the expression period. In order to understand the fraction of Fgp41 in this protein, comparison was made between the deconvolutions of: (1) the ΔS spectrum of labeled cells and the ΔS spectrum of membrane-reconstituted Fgp41; and (2) the S_0 spectrum of labeled cell difference and the S_0 spectrum of membrane-reconstituted Fgp41. The ΔS spectrum of labeled cells and the ΔS spectrum of membrane-reconstituted Fgp41 are compared in Figure 3-6.

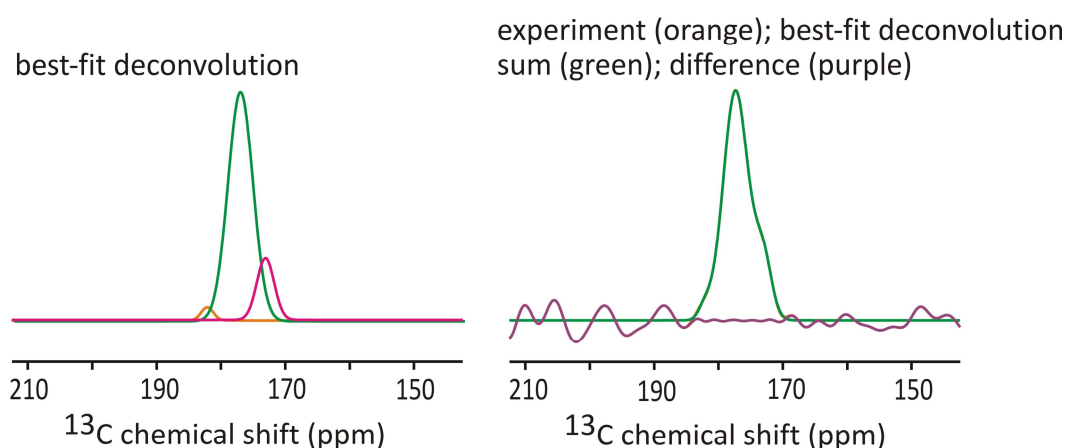
For either case, there were striking similarities in the deconvolutions including the peak chemical shifts and the large fraction of the total intensity in the two high shift peaks corresponding to helical conformation. These similarities as well as the detection of large ΔS signals provide additional strong evidence that Fgp41 is the predominant labeled protein in the

cells. This result was used to conservatively estimate that there was at least 3 mg of Fgp41 in the lyophilized labeled cell NMR sample. Other inputs for this estimate were: (1) the mass Fgp41 in the membrane-reconstituted sample was ~5 mg; (2) the membrane and whole cell data were acquired on the same spectrometer and were the sums of about the same numbers of scans; and (3) for the membrane-reconstituted and whole cell samples, the integrated ^{13}C intensities of the ΔS spectra were within 20% agreement and there was similar agreement for the S_0 spectra. There was ~50 mg total cell mass in the whole cell NMR sample so the ratio of mass Fgp41 to total dry cell mass was ~0.05. There was ~2 g dry cell mass/L culture so prior to solubilization and purification, there was ~100 mg Fgp41/L culture. The much smaller purified yield of ~5 mg Fgp41 /L culture points to solubilization and purification rather than expression as the limiting factors in Fgp41 production. Because relatively harsh conditions were needed to solubilize Fgp41 in the cells, it seems likely that most Fgp41 was in inclusion bodies. Detection of predominant helical conformation for the Leus in Fgp41 in the lyophilized cells including those in the N- and C-helices of a putative SHB structure suggests that this structure is retained in inclusion bodies.

Table 3-3: Deconvolution of spectra of lyophilized cells induced to produce Fgp41. ^aSpectral deconvolution was done with three Gaussian line shapes whose peak shifts, linewidths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and experimental line shape, see Figure 3-5. ^bThe reasons for assignment of peaks to specific conformations are provided in the main text. ^cFull-width at half-maximum linewidth.

Sample/ spectrum type	ΔS and S_0 spectral deconvolution ^a			
	Peak shift (ppm) ^b		Peak width (ppm) ^c	Intensity (fraction of total)
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu cells ΔS	182.1	<i>helix</i>	2.2	0.03
	177.4	<i>helix</i>	4.5	0.82
	173.1	β	3.2	0.15
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu cells - 0.75 $\times$ (unlabeled cells) S_0	180.8	<i>helix</i>	8.9	0.18
	177.6	<i>helix</i>	5.6	0.69
	172.1	β	3.1	0.13

Analysis of ΔS spectrum of lyophilized cells labeled with $1\text{-}^{13}\text{C}$, ^{15}N Leu



Analysis of S_0 spectrum from difference data between lyophilized cells with $1\text{-}^{13}\text{C}$, ^{15}N Leu and cells with unlabeled Leu

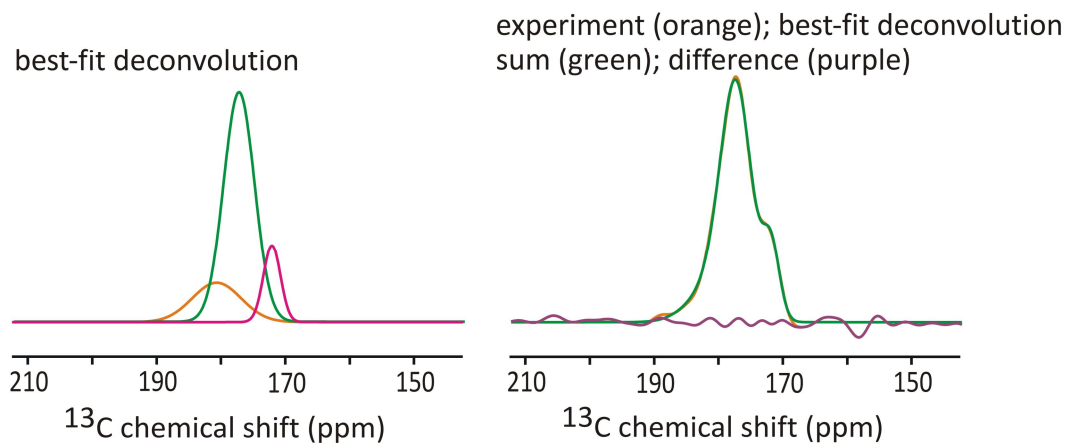


Figure 3-5: Deconvolutions are shown for (top) ΔS spectrum of lyophilized cells induced to produce Fgp41 and labeled with $1\text{-}^{13}\text{C}$, ^{15}N Leu, and (bottom) S_0 spectrum from [lyophilized cells induced to produce Fgp41 and labeled with $1\text{-}^{13}\text{C}$, ^{15}N Leu] – $0.75 \times$ [lyophilized cells induced to produce Fgp41 with no label].

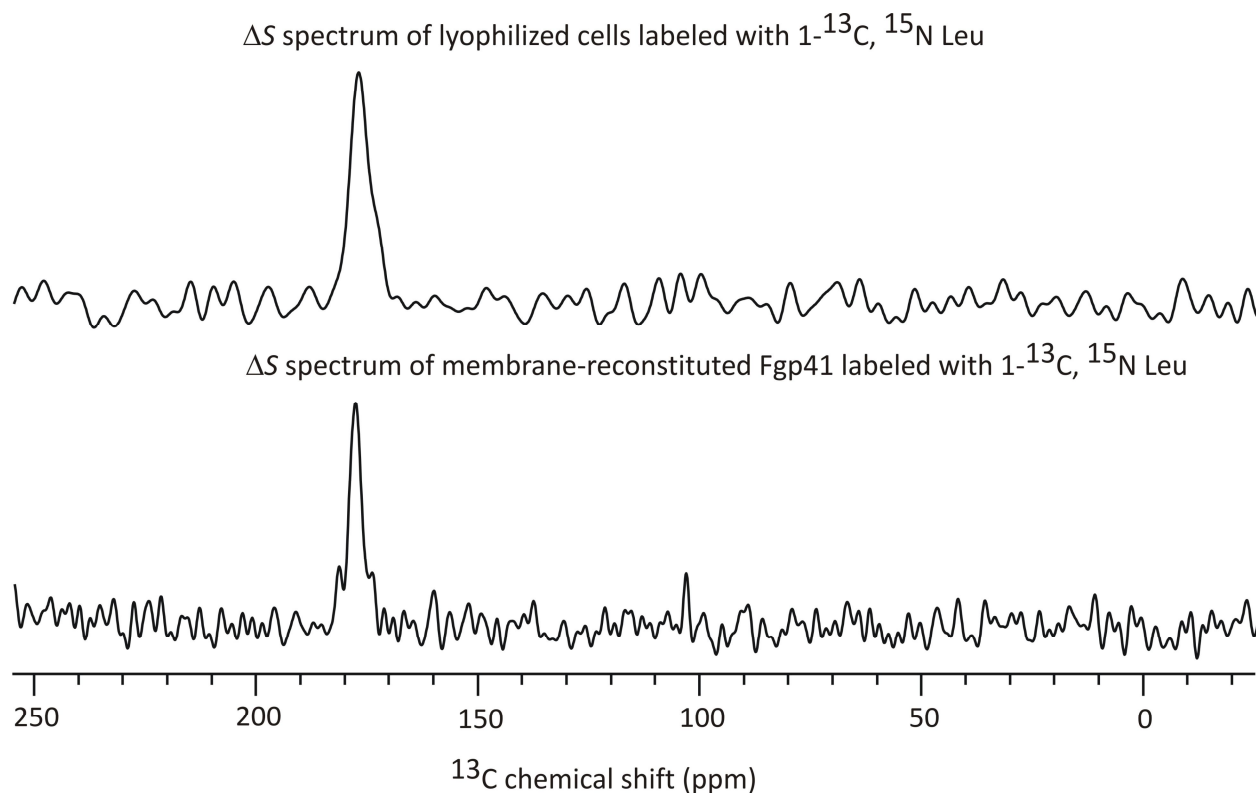


Figure 3-6: Difference spectra are displayed for (top) lyophilized whole cell samples that were induced to produce Fgp41 and labeled with 1-¹³C, ¹⁵N Leu and (bottom) membrane reconstituted purified Fgp41 labeled with 1-¹³C, ¹⁵N Leu. The similarity in line shape and chemical shift of the peak is indicative that Fgp41 is the primary labeled protein present in the lyophilized whole cell sample. The spectra were processed with 100 Hz Gaussian line broadening and a 3rd order polynomial baseline correction.

The successful approach to detecting Fgp41 within whole bacterial cells included identifying an abundant amino acid in Fgp41 (24 Leus) that was the first amino acid of an abundant sequential pair (6 LLs). The procedure included: (1) inducing cells in minimal medium with either 1-¹³C, ¹⁵N Leu or unlabeled Leu; (2) cell pellet lyophilization; and (3) taking ¹³C REDOR SSNMR spectra of the lyophilized whole cells with short dephasing time. As expected, the spectra of the labeled and unlabeled cells were very similar in the aliphatic ¹³C shift region but the labeled cells had greater intensity in the ¹³CO region. The labeled cell – unlabeled cell

difference spectra were therefore assigned to Leu ^{13}C O incorporated into protein produced during the expression period. This approach to detection of recombinant protein in whole cells by SSNMR has several strengths including: (1) small (~50 mL) culture volumes; (2) small (~10 mg) quantities of isotopically labeled amino acids; and (3) simple sample preparation protocol without protein solubilization or purification. The main drawback might be the few days of SSNMR spectrometer time. Interpretation of the SSNMR spectra using this approach will likely not be greatly affected by some “scrambling”, i.e. conversion of the labeled amino acids into other amino acids. For example, transfer of the ^{15}N from the labeled amino acid to other amino acids would likely result in a larger number of labeled ^{13}C - ^{15}N sequential pairs and therefore larger ΔS signal and more sensitive detection of the recombinant protein. Support for minimal scrambling of the Fgp41 sample labeled with $1\text{-}^{13}\text{C}$, ^{15}N Leu included: (1) expression done at lower temperature for short 2h duration; (2) $(\Delta S/S_0)^{\text{exp}}$ for both the whole cell and membrane-reconstituted samples that were close to the values calculated using models without scrambling; and (3) deconvolutions of the S_0 and ΔS ^{13}C O spectra of these samples which agreed nearly quantitatively with the expected secondary structure distributions of the 24 Leu's and the 6 N-terminal Leu's in LL pairs, respectively, see Tables 2-1, 2-2, and 3-3.

Conclusions from Whole Cell NMR Experiments

For most non-bacterial proteins produced in bacteria, a large fraction of the protein in the cells is found in “inclusion bodies” which are macroscopic non-crystalline solid aggregates (3, 12, 13). Inclusion body formation appears to be largely independent of protein sequence.

There are little data about the structure(s) of recombinant protein molecules in inclusion bodies. In the present study, deconvolutions of the S_0 and ΔS spectra of the $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu-labeled inclusion body Fgp41 in cells resulted in line shapes with similar peak shifts and relative intensities as those of membrane-associated Fgp41 with folded SHB structure. It therefore seems likely that at least the SHB fold exists for most Fgp41 molecules in inclusion bodies, as probed by SSNMR spectroscopy of the lyophilized whole cell samples.

Insoluble Cell Pellet SSNMR Spectroscopy

Many recombinantly expressed proteins are packed into inclusion bodies within the bacterial cells. This can be used to our advantage to get even more quantitative SSNMR data regarding recombinant protein. By isolating the insoluble protein within inclusion bodies from the soluble cellular proteins, lipids, and organelles, we can remove more background contributions to the signal. The ICP samples are enriched in inclusion bodies as compared to whole cell samples.

To investigate the contributions to the difference signal for the REDOR experiments, inclusion body samples were prepared as described earlier to study three different sample types: 1) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled Fgp41 inclusion bodies, 2) unlabeled Fgp41 inclusion bodies, and 3) $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled empty pET24a+ plasmid inclusion bodies (within BL21(DE3) Rosetta 2 *E. coli*).

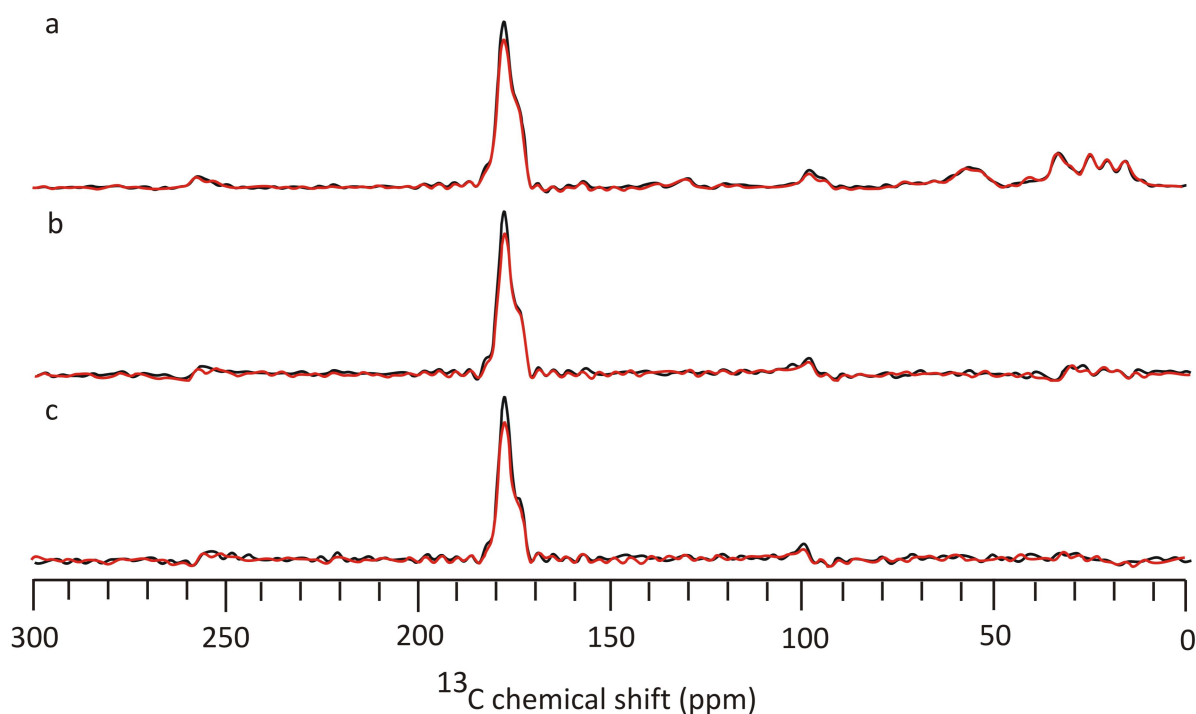


Figure 3-7: S_0 (black) and S_1 (red) spectra for a) $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41 ICP sample, b) $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41 ICP spectrum minus $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled pET24a+ ICP spectrum, c) $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41 ICP spectrum minus unlabeled Fgp41 ICP spectrum. Spectra were processed with no line broadening and a 5th order polynomial baseline correction.

When the spectrum corresponding to the $1\text{-}^{13}\text{C}$, ^{15}N labeled Leu pET24a+ ICP sample is subtracted from the spectrum corresponding to the $1\text{-}^{13}\text{C}$, ^{15}N labeled Leu Fgp41 ICP sample, the resulting spectrum should correspond only to the labeled Fgp41 present within the ICP sample. By running these control experiments, we are confident that the attenuation of the ^{13}C signal is indeed due to the Leu-Leu pairs within the Fgp41 protein sequence (as well as ~1% contribution from natural abundance dephasing). The natural abundance contribution to dephasing can be calculated based on the following model: 1) assume that 100% of the ^{13}C signal in the carbonyl region is due to labeled Leu, 2) assume 100% labeling of Leu residues with

1-¹³C, ¹⁵N Leu, 3) assume no scrambling of the labels, 4) 6 of 24 residues of Leu are immediately followed by another Leu residue. This leaves 18 residues that have a 0.37% chance of being followed by a natural abundance ¹⁵N. (.067 of 18 dephased signal due to natural abundance ¹⁵N). If the total dephased signal is 6 + 0.067, then natural abundance contribution to dephasing is 0.067/6.067 = ~1% contribution. Subtracting the pET24a+ spectrum ensures that contributions to the signal from both native *E. coli* proteins, as well as proteins produced as a result of the presence of the plasmid (i.e. the protein that confers kanamycin resistance) are eliminated. For the subtraction process, the spectra are scaled appropriately so that the signal intensity in the 0 to 90 ppm range is ~zero (not above the noise range) in the resulting spectrum. These subtracted S₀ and S₁ spectra are displayed in Figure 3-7. We presume that the majority of signal increase due to the production of the recombinant protein comes from incorporating the labeled amino acids into the protein. Since the label is within the carbonyl region, the increase in signal observed is primarily within the carbonyl region of the spectrum (approximately from 170 to 185 ppm).

Another set of controls investigated the effect of the labeled amino acid on the spectrum. This was done by comparing the spectrum corresponding to the 1-¹³C, ¹⁵N labeled Leu Fgp41 ICP sample to the spectrum corresponding to an unlabeled Fgp41 ICP sample. The sample were prepared in the exact same manner, except where one sample received labeled Leu, the other sample received unlabeled Leu in its dose of amino acids.

In comparing both of the subtracted spectra, the S_0 line shapes look remarkably similar, as do the ΔS spectra line shapes, found in Figure 3-8b,c. Upon deconvolution of the ΔS spectra, it becomes obvious that the dephasing of the ^{13}C signal is in fact due to the labeled recombinant protein that is present within the ICP samples. Table 3-4 contains the results of the deconvolution of the ΔS spectra for and the $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41 ICP sample, $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled Fgp41 ICP sample spectrum minus either $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled pET24a+ ICP spectrum, or unlabeled Fgp41 ICP spectrum. The chemical shift obtained from the deconvolution of each spectrum is ~ 178.4 ppm for all three samples, which is indicative of a helical secondary structure for the dephased Leu residues in each sample. This corresponds well with previous data on the folded, membrane reconstituted Fgp41 sample, as well as the crystal structures for gp41 constructs which depict two helices which should contain the residues in the sequential LL pairs (14-16).

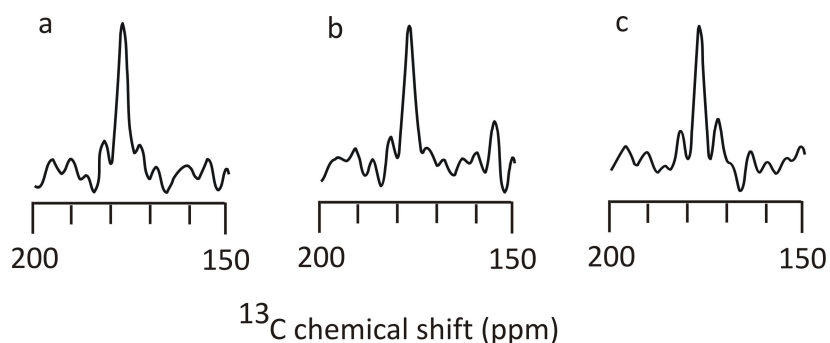


Figure 3-8: $\Delta S \equiv S_0 - S_1$ spectra derived from ICP samples. For panel a, both S_0 and S_1 are from the same $1\text{-}^{13}\text{C}$, ^{15}N Leu Fgp41 sample. For panels b and c, $S_0(S_1)$ is the difference between the individual $S_0(S_1)$ of two different samples: b) $1\text{-}^{13}\text{C}$, ^{15}N Leu Fgp41 sample minus $1\text{-}^{13}\text{C}$, ^{15}N Leu pET24a+ sample; c) $1\text{-}^{13}\text{C}$, ^{15}N Leu Fgp41 sample minus unlabeled Fgp41 sample. Both the S_0 and the S_1 spectrum of each ICP sample was the sum of 50,000 scans. Spectra were processed with 100 Hz Gaussian line broadening and a 5th order polynomial baseline correction.

Table 3-4: Best fit deconvolution of Figure 3-8 spectra. The parameters are for the best-fit Gaussian lineshape of the dominant spectral peak. The integrated signal intensity was obtained by integrating the peak in the difference spectrum that appears between 170 ppm to 185 ppm. The uncertainty in integrated signal intensity was calculated using the RMSD integrated intensity of 5 ppm regions without signal.

Fig. 3-7 panel	Peak ^{13}C shift (ppm)	FWHM linewidth (ppm)	Integrated intensity
a	178.4	3.0	61 ± 7
b	178.4	3.1	57 ± 6
c	178.4	2.8	53 ± 5

Quantitative Detection of Recombinant Protein Expression

Another set of experiments was designed to test whether the REDOR method could be used to quantitatively detect recombinant protein expression within ICP samples. In these experiments, several different protein constructs were utilized, as outlined in the beginning of the chapter. We chose constructs that were fairly well established to produce protein in inclusion body form and have been studied previously by other methods (14, 17-20). There were several different plasmid types and two different strains of *E. coli* used for the studies, which are outlined in Table 3-1.

To relate the integrated signal intensity to the amount of protein present in an ICP sample, a calibration curve was created. For the calibration experiments, samples consisting of $1\text{-}^{13}\text{C}$, ^{15}N Leu and talc (an inert substance that will not contribute to the ^{13}C NMR spectrum, as the chemical formula for talc is $\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$) were used to perform REDOR experiments in the same manner as those performed for the ICP samples. By measuring the integrated signal

intensity in the carbonyl region for samples containing a known amount of labeled leucine, it is straightforward to determine the amount of signal per mole of ^{13}C label present in the sample.

The S_0 spectra for these samples are presented in Figure 3-10. Data obtained from the S_0 spectra of these experiments is reported in Table 3-5 and the calibration curve is shown in Figure 3-9.

Table 3-5: Information obtained from REDOR S_0 spectra of 1- ^{13}C , ^{15}N Leu/talc samples. The error in integrated signal intensity was obtained by integrating regions of noise in the S_0 spectrum for 0.5 mg 1- ^{13}C , ^{15}N Leu containing sample. This sample was used because all spectra showed apodization of the signal, and this had the least amount. The noise should be the same in all spectra as the same conditions were used for the experiments.

Amount of 1- ^{13}C , ^{15}N Leu (mg)	Amount of 1- ^{13}C , ^{15}N Leu (moles)	Integrated carbonyl signal intensity
0.5	3.75×10^{-6}	1432 ± 12
5	3.75×10^{-5}	11666 ± 12
25	1.88×10^{-4}	40603 ± 12

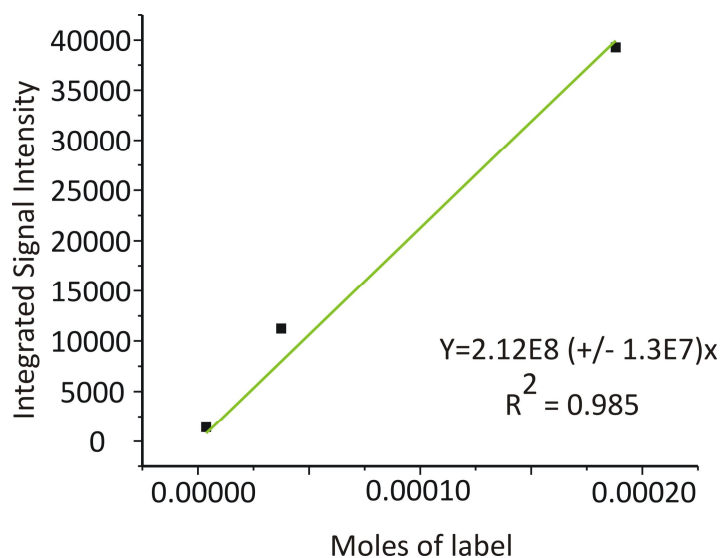


Figure 3-9: Plot of the integrated signal intensity in the carbonyl region of the ^{13}C spectrum (170 $\rightarrow$ 185 ppm) from 50,000 REDOR S_0 scans vs. the number of moles of label present. The

samples measured to create this calibration curve were made of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu manually mixed with talc to create a uniform distribution of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu to fill the 4 mm MAS rotor. The line shown is a linear regression fit with a forced (0,0) intercept. The equation of linear regression is $y=2.12\times 10^8 x$, and $R^2 = 0.985$. The standard error associated with the slope is 1.3×10^7 . Numerical data corresponding to this plot is presented in Table 3-5. S_0 spectra are shown in Figure 3-10.

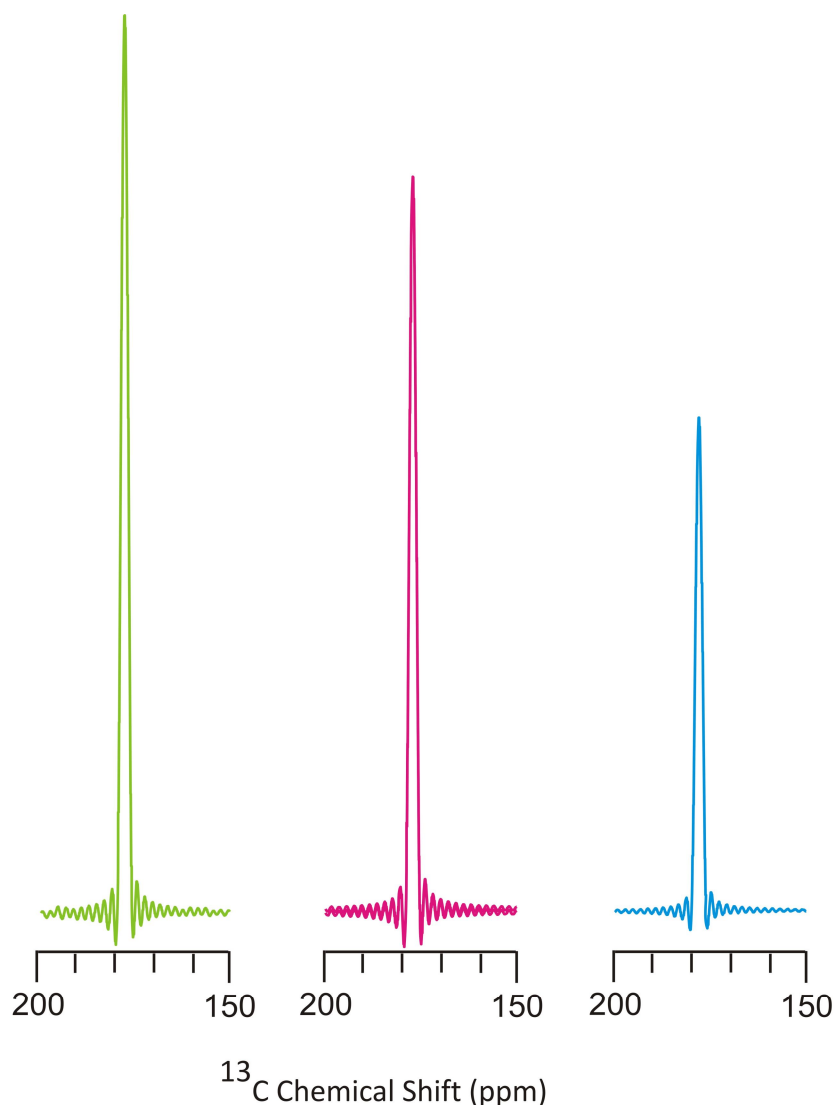


Figure 3-10: REDOR S_0 spectra of $1\text{-}^{13}\text{C}$, ^{15}N labeled Leu mixed with talc. Blue = 25 mg $1\text{-}^{13}\text{C}$, ^{15}N Leu, pink = 5 mg $1\text{-}^{13}\text{C}$, ^{15}N Leu, and green = 0.5 mg $1\text{-}^{13}\text{C}$, ^{15}N Leu. The spectra are scaled such that the y axis of the spectra containing 0.5 mg : 5 mg : 25 mg of $1\text{-}^{13}\text{C}$, ^{15}N Leu were multiplied by 50 : 10 : 1. This was done so that we may assess the linearity of the spectral intensities with respect to the amount of labeled material present. Each spectrum is the result of 50,000 S_0 scans. Spectra are processed with 200 Hz Gaussian line broadening and 5th order polynomial baseline correction.

Calculation of Expression Levels

The primary piece of data utilized to determine the level of recombinant protein expression is the S_0 integrated signal intensity in the 170 to 185 ppm region of the spectrum.

We feel this is appropriate since this region is where the signal intensity increases as more labeled recombinant protein is produced. In order to compare different spectra, we scaled the data so that the integrated signal intensity was the same in the 0 to 90 ppm region, as this should be unaffected by isotopic labeling. The following method was used to calculate the expression level for each ICP sample using its corresponding S_0 spectrum. Data used in the calculations and results of the calculations can be found in Tables 3-6 and 3-7, respectively.

$$\text{Expression Level (mg protein / L culture)} = [aA - bB]C$$

$a \equiv$ scaling factor for sample with recombinant protein

$$a = [1000 / \text{integrated signal intensity (0} \rightarrow 90 \text{ ppm)}]$$

$A =$ integrated S_0 signal intensity from 170 $\rightarrow$ 185 ppm for sample with recombinant protein

$b \equiv$ scaling factor for sample with empty pET24a+ plasmid

$$b = [1000 / \text{integrated signal intensity (0} \rightarrow 90 \text{ ppm)}]$$

$B =$ integrated S_0 signal intensity from 170 $\rightarrow$ 185 ppm for sample with empty pET24a+ plasmid

$C \equiv$ constant to convert $[aA - bB]$ to mg protein / L culture. C takes into consideration the molar mass of the recombinant protein and the number of Leu residues (and therefore the number of ^{13}C labels) present in the protein. C also contains a factor of 40 to compensate for the fact that only ~ 25 mL worth of culture is used for each ICP sample. (50 mL of *E. coli* cell culture is grown for each sample. The entire cell pellet after centrifugation is sonicated in PBS to remove soluble proteins, and centrifuged once again. After this step, approximately half of the total volume of ICP is able to fit into the rotor, corresponding to ICP from about 25 mL of culture.) Depending on the strain of *E. coli* utilized and the particular plasmid that the

recombinant DNA is inserted in, there may be other contributions to the NMR signals (not from the recombinant protein) within the ICP. For example, if the protein that confers antibiotic resistance for a particular protein is produced during the period when labels are present in the medium, that protein will also contribute to the observed ^{13}C spectrum as labeled amino acids will be incorporated into that protein as well. Additionally, if the recombinant protein is expressed in a strain where expression is not tightly controlled (such as BL21(DE3)) then there may be a population of recombinant protein present within the cell before labels are present within the medium. This protein will not contribute significantly to the NMR spectrum, and therefore will not be accounted for in this method of quantitation. 2.12×10^8 is a factor taken from the calibration curve created by measuring the NMR ^{13}C signal intensity with respect to the number of moles of labeled Leu present in a sample, and is used to convert from integrated signal intensity into moles of label present.

$$C = \frac{\text{molar mass of protein (mg/mol)}}{\text{\# Leu residues in protein}} \times \frac{40}{2.12 \times 10^8}$$

In addition, it can be noted from Table 3-6 that for all samples other than the $1\text{-}^{13}\text{C}$, ^{15}N Leu labeled pET24a+ sample, the integrated signal intensity in the carbonyl region is much greater than the integrated signal intensity in the 0 to 90 ppm region. The opposite is true for the pET24a+ sample. As the Leu with $1\text{-}^{13}\text{C}$ label is only in the medium during the expression period, the significant enhancement of the carbonyl signal in spectra of cells expressing recombinant protein supports that the recombinant protein is the major protein produced during the expression period.

Table 3-6: Integrated signal intensities from the S_0 spectrum for each ICP sample and the calculated scaling factors. The scaling factor was $[1000/(\text{integrated signal intensity in the 0 to 90 ppm region})]$.

Sample Description	170 to 185 ppm	0 to 90 ppm	Scaling Factor
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu pET24a+	353	802	1.25
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Hairpin	1269	576	1.74
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41	1041	926	1.08
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41+	918	857	1.17
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu FHA2	931	667	1.50
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Val HPI	3708	1338	0.75
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Ala HPI	2934	931	1.07
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Tyr HPI	3796	1150	0.87

Table 3-7: Calculated normalized carbonyl signal = $aA - bB$ and expression level for each ICP sample. The # of Leu $\equiv$ number of Leu residues in the recombinant protein sequence. The sample-to-sample variation in recombinant protein expression level is $\sim 10\%$ based on the analysis for the three $1-^{13}\text{C}$ labeled Leu HPI samples.

Sample Description	Normalized Carbonyl Signal	# of Leu	Expression Level (mg protein/L culture)	Expression Level ($\mu\text{mol protein/L culture}$)
$1-^{13}\text{C}, ^{15}\text{N}$ Leu Hairpin	1763	14	270 ± 2	25.2 ± 0.2
$1-^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41	684	24	105 ± 1	5.7 ± 0.1
$1-^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41+	632	26	101 ± 2	4.9 ± 0.1
$1-^{13}\text{C}, ^{15}\text{N}$ Leu FHA2	956	13	329 ± 4	14.7 ± 0.2
$1-^{13}\text{C}$ Leu, ^{15}N Val HPI	2331	14	378 ± 3	33.3 ± 0.2
$1-^{13}\text{C}$ Leu, ^{15}N Ala HPI	2710	14	439 ± 3	38.7 ± 0.2
$1-^{13}\text{C}$ Leu, ^{15}N Tyr HPI	2862	14	464 ± 3	40.9 ± 0.2

As we have three samples that all correspond to $1-^{13}\text{C}$ labeled Leu human proinsulin, these were useful in determining a threshold of precision. Ideally, these samples should give the same normalized level of expression value since the protein construct and the manner in which the samples were produced are the same. From the data corresponding to human proinsulin, as shown in Table 3-6, Table 3-7, and Figure3-11, we have determined that in this case, there is a standard deviation of $\sim 10\%$ in the calculated level of expression between the three samples. Due to the consistency between these values, and the large difference between the calculated level of expression for human proinsulin and the other constructs (in general, the human proinsulin samples yielded two to four times higher signal intensity), we believe the validity of using these calculated values to assess the level of recombinant protein expression.

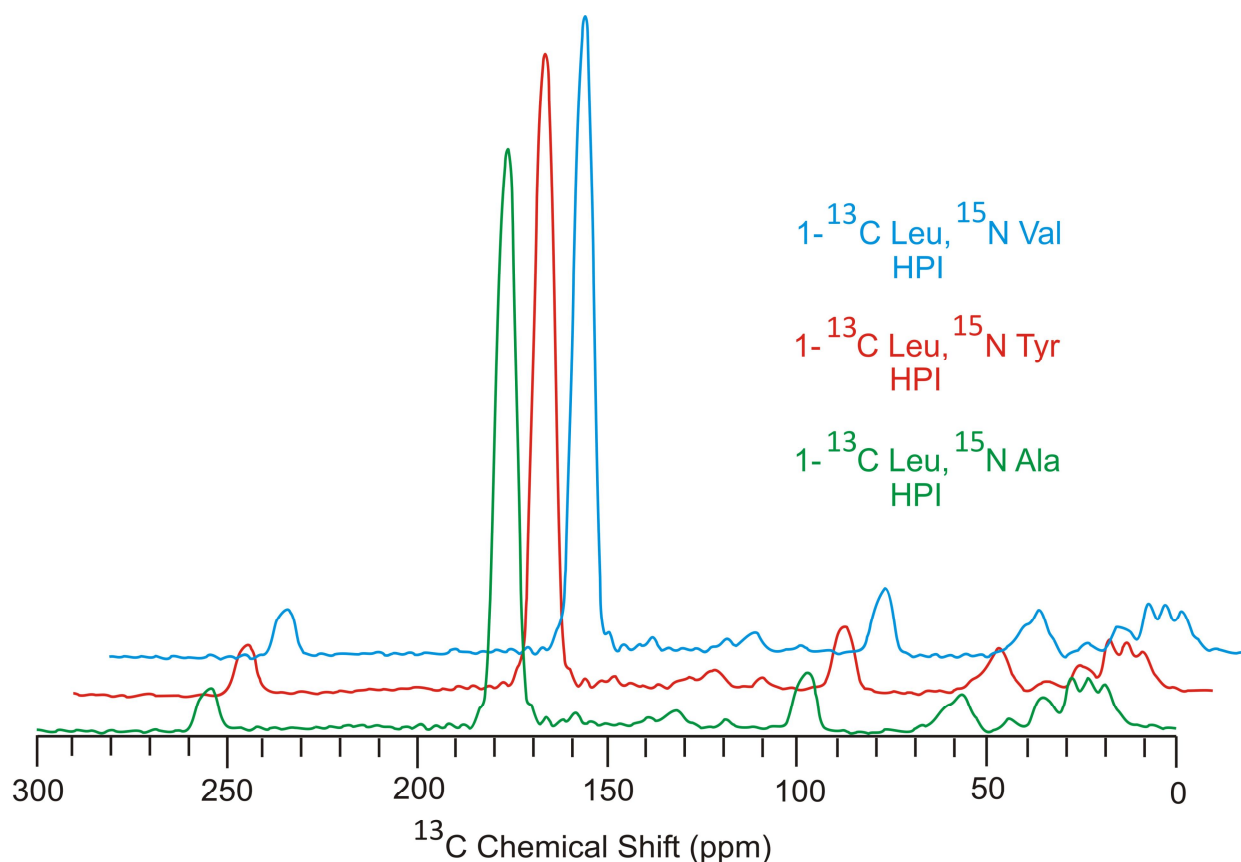


Figure 3-11: REDOR S_0 spectra for $1\text{-}^{13}\text{C}$ Leu labeled Human Proinsulin ICP samples. The labeling of each sample is indicated. Each spectrum is the sum of 50,000 S_0 scans. The spectra were processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra are scaled such that the signal in the 0 to 90 ppm region is the same, as this should be unaffected by isotopic labeling.

In addition to the NMR data, we have assessed the relative level of recombinant protein expression by boiling small amounts of the insoluble cell pellets in an SDS containing sample buffer and running SDS-PAGE of the samples. The resulting gel is depicted in Figure 3-12b. The recombinant protein seems to be ~the darkest band in the lanes for human proinsulin and hairpin samples, and fairly faint in the Fgp41 and FHA2 samples, which shows a correlation with the levels of expression calculated for each of these. While this may be a straightforward way to assess that protein is being produced in samples like Hairpin and Human proinsulin, Figure 3-12b illustrates how useful the NMR based approach is in instances where the recombinant

protein band is similar in intensity to native proteins, such as in the Fgp41 and FHA2 cases. The NMR data suggests that FHA2 expresses to a higher level (in mg/L) than Hairpin, though the band is more difficult to observe on the SDS-PAGE gel. This is likely due to FHA2 being more poorly solubilized than Hairpin and HPI, as the expression level of FHA2 is $14.7 \pm 0.2 \mu\text{mol/L}$ compared to the Hairpin expression level of $25.2 \pm 0.2 \mu\text{mol/L}$.

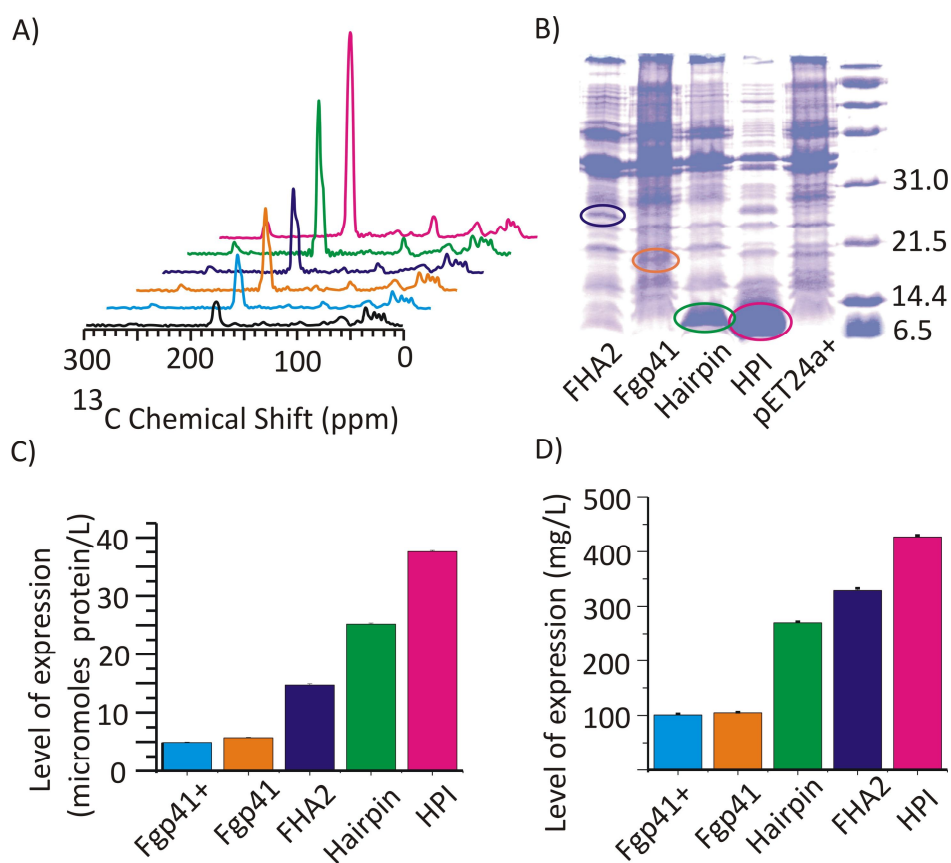


Figure 3-12: a) ^{13}C S₀ REDOR SSNMR spectra of ICP samples labeled with 1- ^{13}C Leu. Each spectrum is the sum of 50000 scans. The spectral intensities are scaled to approximate equal values in the 0 to 90 ppm range. The intensity in this region should be least affected by protein synthesized with 1- ^{13}C Leu in the medium. The spectra are all processed with 200 Hz of Gaussian line broadening and a 3rd order polynomial baseline correction. b) SDS-PAGE gel of insoluble cell pellets after boiling in SDS-containing sample buffer. The molecular weight standards are labeled in the right most lane in kDa and the band attributed to recombinant protein is circled in each sample lane. c,d) Recombinant protein (RP) expression levels calculated from the difference in ^{13}C CO signal intensity between the cells with RP and cells

without RP. These values were calculated based on analysis of the NMR data shown in panel A, and the colors correspond. Numerical values from the NMR data can be found in Tables 3-6 and 3-7.

Conclusions from ICP NMR Experiments

In conclusion, our method shows that the level of recombinant protein expression in bacterial cells can be quantified without purification using a straightforward application of solid-state NMR. As discussed previously, we had conservatively estimated that the Fgp41 construct expresses at ~100 mg/L of bacterial culture by comparing SSNMR signal intensities obtained by using whole *E. coli* cell samples expressing Fgp41 to signal intensities obtained from samples containing purified lipid reconstituted Fgp41. The new method agrees with this estimate, reporting Fgp41 expression at a level of 105 ± 1 mg / L of bacterial cell culture. The new method is quick and inexpensive, and only moderate NMR fields are required which should make this method widely applicable. This is the first instance of a way to quantify the amount of recombinant protein expressed within bacterial cells that does not depend on assumptions that the protein will be in a specific conformation inside of the cells nor does it depend on the ability to solubilize the protein.

Previous reports of recombinant protein yields from the constructs studied vary. Human proinsulin was reported to be isolated from *E. coli* cell culture within inclusion bodies estimated at approximately 200 mg inclusion bodies / 1 L bacterial cell culture, with a final yield of pure, active human proinsulin of 1-2 mg / 1 L culture(19). The Hairpin protein was reported to yield ~50 mg of pure protein / 1 L culture after harsh denaturation of the *E. coli* cells using ultrasonication in glacial acetic acid followed by RP-HPLC of the protein(20). The yield of isotopically labeled Fgp41 was ~5 mg / 1 L culture after sonication in SDS and subsequent

detergent removal(14). FHA2 was purified to a yield of up to 20 mg / L culture when initial solubilization with sarkosyl as the primary denaturant was followed by subsequent solubilization in urea(10). Measurements of total expression yields of these proteins are not available to my knowledge prior to this work. The method utilizing FT-IR had reported that inclusion bodies from the expression of a GFP-autoprotease fusion protein (which has been shown to primarily express in inclusion bodies) were observed at concentrations as high as 200 mg / g dry biomass after isolation from E. coli fermentation cultures, however the purified yield of this protein was not mentioned (2).

It appears that the difference between the amount of expressed recombinant protein and the purified yield of recombinant protein can vary greatly depending on different aspects of the recombinant protein itself. For example, the Hairpin protein has a fairly high yield of purified protein, and does not contain any transmembrane domains or fusion peptides that could cause aggregation problems. FHA2 has a much higher purified yield ($\sim 4\times$) than Fgp41, and also has a higher percentage of charged residues : hydrophobic residues ($2\times$ greater) than does Fgp41. HPI has a very low yield reported despite a high expression level detected, though this is likely due to the need for three disulfide bonds to be formed for the protein to be considered active.

A Possible Alternate Method of Calculating Expression Levels

Table 3-8 and Figure 3-13 contain information obtained from REDOR ΔS spectra of the ICP samples utilized for determining levels of recombinant protein expression. Table 3-9 contains information about each protein construct used to directly calculate the amount of recombinant protein per liter of E. coli culture from the ΔS spectra. The method of calculation is outlined

below. The number of dephased residues is defined as the # of ^{13}C labeled amino acids that are directly followed in the protein sequence by a ^{15}N labeled amino acid (according to the chosen labeling scheme). The factor of 0.7 accounts for the efficiency of 1 ms dephasing time in REDOR experiments to detect directly bonded labeled nuclei(11).

To calculate the milligrams of recombinant protein per L of culture:

$$\text{integrated signal} \times \frac{1 \text{ mol protein}}{\# \text{ dephased residues} \times 0.7} \times \frac{\text{molar mass protein (mg)}}{1 \text{ mol protein}} \times \frac{40 \text{ samples / L}}{2.12 \times 10^8}$$

To calculate the μmol protein per L of culture:

$$\text{integrated signal} \times \frac{1 \text{ mol protein}}{\# \text{ dephased residues} \times 0.7} \times \frac{10^6 \mu\text{mol}}{1 \text{ mol}} \times \frac{40 \text{ samples / L}}{2.12 \times 10^8}$$

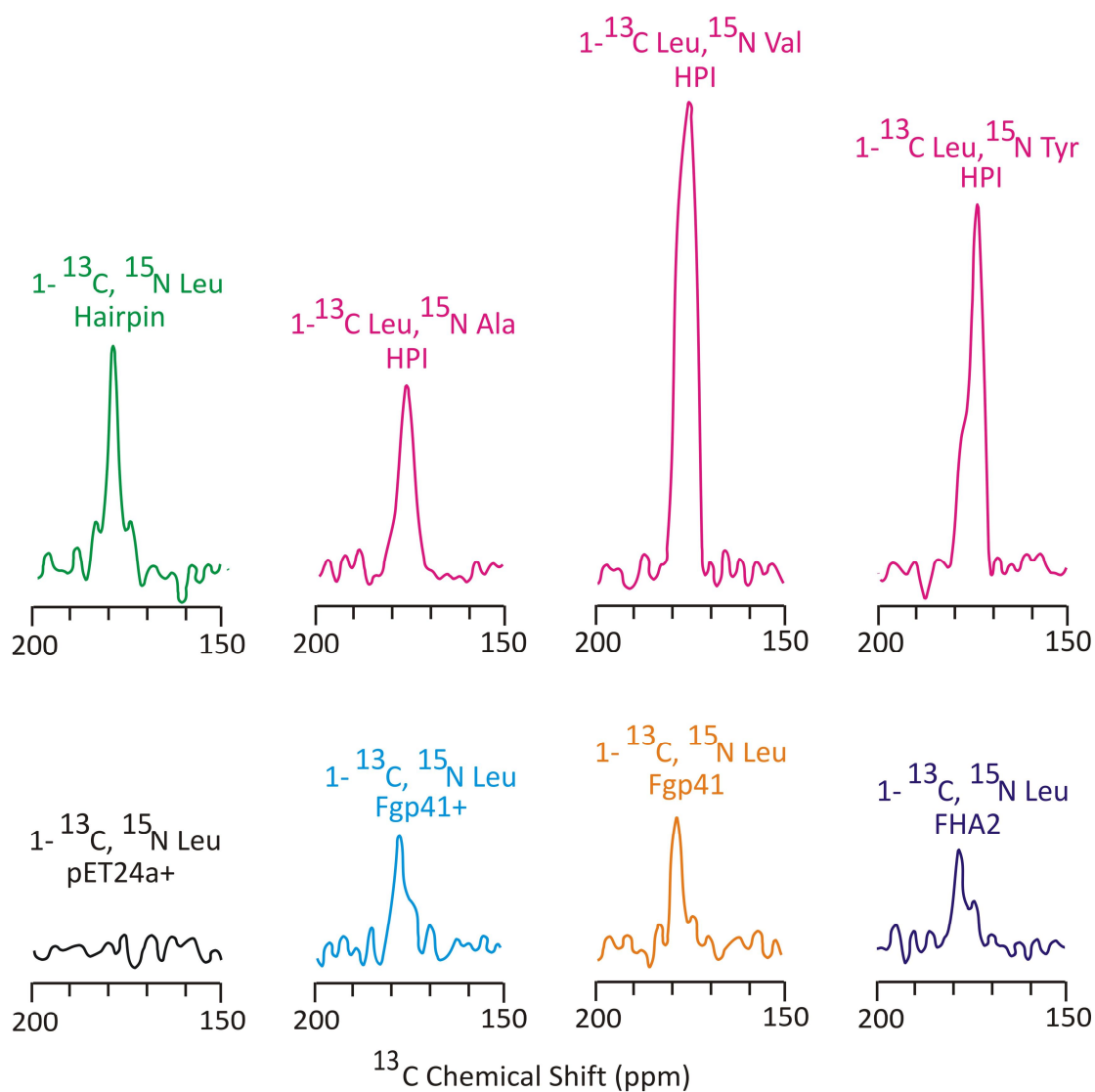


Figure 3-13: ^{13}C - ^{15}N REDOR ΔS spectra of ICP samples processed without line broadening and with a 5th order polynomial baseline correction. Each ΔS spectrum was the result of 50,000 S_0 scans – 50,000 S_1 scans. The labeling and protein construct is indicated above the spectrum for each sample.

Table 3-8: Data obtained from ^{13}C - ^{15}N REDOR ΔS spectra of ICP samples processed without line broadening and with a 5th order polynomial baseline correction. Each ΔS spectrum was the result of 50,000 S_0 scans – 50,000 S_1 scans. Line width reported is the Full Width at Half Maximal value, and was measured from the spectra.

Sample	^{13}C Chemical Shift (ppm)	Line Width (ppm)	Integrated Signal Intensity
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Hairpin	178.6	3.4	138.6
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41	178.3	3.5	78.6
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41+	178.1	4.5	78.9
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu FHA2	178.6	3.4	71.1
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Val HPI	175.2	7.0	394.6
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Ala HPI	176.6	5.3	146.3
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Tyr HPI	174.3	4.7	250.8

Table 3-9: Calculated recombinant protein expression levels using the ΔS spectra for the samples mentioned in Table 3-8.

Sample	Integrated Signal Intensity	# of dephased residues	molar mass (g/mol)	mg of protein / L	μmol of protein / L
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Hairpin	138.6	4	10723	100 ± 4	9.3 ± 0.4
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41	78.6	6	18376	65 ± 4	3.5 ± 0.2
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu Fgp41+	78.9	6	20809	74 ± 4	3.5 ± 0.2
$1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu FHA2	71.1	1	22363	429 ± 23	19.2 ± 1.0
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Val HPI	394.6	2	11348	603 ± 7	53.2 ± 0.6
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Ala HPI	146.3	1	11348	447 ± 17	39.4 ± 1.5
$1\text{-}^{13}\text{C}$ Leu, ^{15}N Tyr HPI	250.8	2	11348	384 ± 4	33.8 ± 0.4

Overall, the results of the calculated expression levels from ΔS spectra yield more conservative estimates of the amount of recombinant protein present in samples. In general, the results follow the same trend as the previous calculations (using S_0 data) showing a very high expression level of human proinsulin (average 340 mg/L), and lower expression levels for Fgp41 (46 mg/L) and Fgp41+ (52 mg/L). The hairpin and FHA2 expression levels calculated from the ΔS spectra do not follow the trend observed in the S_0 data.

An advantage of utilizing ΔS spectra to calculate the expression level of recombinant proteins in either whole *E. coli* cells or in insoluble cell pellets is that the ΔS spectrum filters out the majority of natural abundance contributions to the spectrum. This allows the researcher to skip the step of running control spectra of *E. coli* cells containing the empty plasmid (without a protein insert) and subtracting these signal intensities to obtain expression levels.

One disadvantage of using ΔS spectra to determine expression levels is evident when analyzing the HPI data. In the HPI amino acid sequence, there is not an adjacent pair of Leu residues, thus I was forced to use a different ^{15}N amino acid (not the doubly labeled Leu) to label the protein. In order to use the ΔS spectra to effectively quantify the recombinant protein, we must know the efficiency of labeling, i.e. quantitative dephasing is needed to accurately estimate the expression levels.

Future Work

Though the ^{13}C - ^{15}N REDOR experiment is technically only a double resonance NMR experiment, it does require a three channel probe in the HXY configuration. Aside from ^{13}C and ^{15}N , the

third channel, set up for ^1H frequency is utilized for cross-polarization of ^1H magnetization to ^{13}C . If this equipment is unavailable, we reason that the quantitative aspect of this work could still be performed in a double resonance experiment, utilizing a simple cross-polarization (^1H to ^{13}C) experiment and detecting on the ^{13}C channel. Then observing a change in signal resulting from different expression conditions could give the investigator a reasonable quantitative model for how changing different conditions (i.e. media components, concentration of inducer, etc.) changes the level of recombinant protein expression.

REFERENCES

REFERENCES

1. Miles, A. P., and Saul, A. (2005) Quantifying recombinant proteins and their degradation products using SDS-PAGE and scanning laser densitometry, *Methods in molecular biology (Clifton, N.J.)* 308, 349-356.
2. Gross-Selbeck, S., Margreiter, G., Obinger, C., and Bayer, K. (2007) Fast quantification of recombinant protein inclusion bodies within intact cells by FT-IR spectroscopy, *Biotechnology Progress* 23, 762-766.
3. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy, *J. Am. Chem. Soc.* 130, 12568-12569.
4. Gullion, T. (1998) Introduction to rotational-echo, double-resonance NMR, *Concepts Magn. Reson.* 10, 277-289.
5. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts, *J. Biomol. NMR* 25, 173-195.
6. Tong, K. I., Yamamoto, M., and Tanaka, T. (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in E-coli, *J. Biomol. NMR* 42, 59-67.
7. Waugh, D. S. (1996) Genetic tools for selective labeling of proteins with alpha-N-15-amino acids, *J. Biomol. NMR* 8, 184-192.
8. Harris, D. C. (2003) *Quantitative Chemical Analysis*, 6th ed., W.H. Freeman and Company, New York.
9. Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR, *J. Magn. Reson.* 162, 479-486.
10. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in E. coli, purification, and refolding of the full ectodomain of the Influenza virus membrane fusion protein, *Prot. Expr. Purif.* 61, 212-219.
11. Yang, J. (2003) Solid-state nuclear magnetic resonance structural studies of the HIV-1 fusion peptide in the membrane environment, Ph. D. Thesis, Michigan State University, East Lansing, MI.
12. Wang, L. (2009) Towards revealing the structure of bacterial inclusion bodies, *Prion* 3, 139-145.
13. Gatti-Lafranconi, P., Natalello, A., Ami, D., Doglia, S. M., and Lotti, M. (2011) Concepts and tools to exploit the potential of bacterial inclusion bodies in protein science and biotechnology, *Febs J.* 278, 2408-2418.

14. Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) Solid-State Nuclear Magnetic Resonance (NMR) Spectroscopy of Human Immunodeficiency Virus gp41 Protein That Includes the Fusion Peptide: NMR Detection of Recombinant Fgp41 in Inclusion Bodies in Whole Bacterial Cells and Structural Characterization of Purified and Membrane-Associated Fgp41, *Biochemistry* 50, 10013-10026.
15. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, *Plos Pathogens* 6, e1000880.
16. Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M., and Clore, G. M. (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, *EMBO J.* 17, 4572-4584.
17. Kim, C. S., Epand, R. F., Leikina, E., Epand, R. M., and Chernomordik, L. V. (2011) The final conformation of the complete ectodomain of the HA2 subunit of Influenza Hemagglutinin can by itself drive low pH-dependent fusion, *J. Biol. Chem.* 286, 13226-13234.
18. Curtis-Fisk, J., Preston, C., Zheng, Z. X., Worden, R. M., and Weliky, D. P. (2007) Solid-state NMR structural measurements on the membrane-associated influenza fusion protein ectodomain, *J. Am. Chem. Soc.* 129, 11320-11321.
19. Cowley, D. J., and Mackin, R. B. (1997) Expression, purification and characterization of recombinant human proinsulin, *Febs Letters* 402, 124-130.
20. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, *Biochemistry* 48, 2714-2722.

Chapter 4 – Structural analysis of human proinsulin within bacterial inclusion bodies by solid state NMR

Introduction

This chapter covers a short project which investigates the structure of human proinsulin within bacterial inclusion bodies. Proinsulin is the biological precursor to the hormone insulin and undergoes post-translational modifications in the islet beta cells to produce the active hormone insulin. Previous studies on this particular construct of human proinsulin have been performed and suggested that the protein will be found within inclusion bodies when it is expressed in *E. coli* (1). REDOR is a useful tool to study the secondary structure at particular residues throughout the protein sequence of proinsulin, and thus can give some insight into the structure of the protein within inclusion bodies.

A solution NMR structure was determined for a mutated human proinsulin construct (H10D, P28K, K29P) and showed a native, insulin-like moiety in the A and B chains, and a more disordered C-chain(2). This DKP-proinsulin structure will be the basis of comparison to my SSNMR structural study of proinsulin within bacterial inclusion bodies (PDB-ID for the structure is 2KQP). Below I have color coded the sequence of proinsulin to represent the structural findings from the solution NMR structure of DKP-proinsulin. The mutated residues are shown in pink, residues in coil conformation are shown in blue, helical conformation in green, and β -turn conformation is shown in gold.

MGSSSHHHHHSSGLDPVL¹ FVNQHLCGSH¹¹ LVEALYLVCG²¹ ERGFFYTPKT³¹ RREAEDLQVG
⁴¹ QVELGGGPGA⁵¹ GSLQPLALEG⁶¹ SLQKRGIVEQ⁷¹ CTSICSLYQ⁸¹ LENYCN

COIL TURN HELIX

Human Proinsulin Construct Information

Source of Human Proinsulin

The human proinsulin plasmid (contained within expression vector pQE-31) was provided by Dr. Robert B. Mackin (Department of Biomedical Sciences, Creighton University School of Medicine, Omaha, NE).

DNA Sequence of Human Proinsulin

Sequencing result was posted on the Finch data server on November 8, 2011, and can be accessed as file DPW277. The DNA corresponding to the human proinsulin construct is shown in bold, and the rest corresponds to vector DNA.

TTACTTTAGAAGGAGATATACCAT**GGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGATCC**
GGTGCTGATGTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTCTACCTAGTGTGCG
GGGAACGAGGCTTCTTCTACACACCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGTGGGGCAGG
TGGAGCTGGGCGGGGGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAA
GCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTACTGCAACT
AGAGTCGACCTGCAGCCAAG

Amino Acid Sequence of Human Proinsulin

The protein contains an N-terminal polyhistidine tag for purification purposes. Non-native residues are underlined.

MGSSSHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYLVCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGPGA ⁵¹GSLQPLALEG ⁶¹SLQKRGIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Human proinsulin expression

The pQE-31 plasmid contains a gene for ampicillin resistance, so all expression media contained ampicillin at 100 mg/L. The pQE-31/hpi plasmid was transformed into BL21(DE3) competent *E. coli* cells and plated on LB/agar plates containing ampicillin for selection. Colonies were picked and glycerol stocks were made and stored at -80°C.

A normal expression procedure for production of human proinsulin in *E. coli* cells was carried out as follows. 100 mL of LB (containing ampicillin) in a 250 mL flask was inoculated with 0.5 mL of glycerol stock containing *E. coli* cells with the pQE-31/hpi plasmid. The flask was incubated at 37°C while shaking at 180 rpm overnight (approximately 16 hours). The cells were reclaimed by centrifugation, and the pellet was resuspended into 50 mL of M9 minimal medium containing ampicillin, 1 mM MgSO₄, and 250 µL 50% glycerol. After one hour of shaking in the new medium at 37°C, 10 mg of each amino acid is added to the medium, and IPTG is added to a concentration of 0.2 mM. After one hour, another dose containing 10 mg of every amino acid is added to the medium. Expression continues for a total of 3 hours at 37°C. Cells are reclaimed by centrifugation (10,000 *g*, 4°C, 10 minutes). Cell pellets were stored at -20°C until preparation for NMR use.

NMR sample preparation

As the goal of the study presented in this chapter was to study the structure of human proinsulin within bacterial inclusion bodies, I made an attempt to rid the system of soluble protein. This step not only ensures that the REDOR difference signal obtained from the NMR experiments correlated with the insoluble protein, but it should also simplify the spectrum by removing any contribution to the signal from soluble proteins.

Each cell pellet was combined with ~40 mL PBS (pH 7.3) and placed on ice. Lysis of the *E. coli* cells was achieved by sonication with a tip sonifier (using 4 one minute cycles, 80% amplitude, 0.8 seconds on, 0.2 seconds off). After sonication, the samples were centrifuged at 50,000 *g* / 4°C / 20 minutes. The supernatant of each sample (containing soluble proteins) was

discarded, and the pellet was packed into a 4 mm solid state NMR magic angle spinning rotor. The active sample volume of the rotor was approximately 40 μ L.

Isotopic Labeling Considerations

Each human proinsulin sample was isotopically labeled with 1- 13 C and 15 N amino acids to observe unique sequential pairs of amino acids throughout the protein sequence. The positions selected for observation were chosen for several reasons. 1) Amino acids were selected that are known to label well (from previous work in our research group by Jaime Curtis-Fisk and me) were utilized for this project. 2) Since proinsulin contains the A and B chains from the insulin hormone as well as the signaling C chain, I attempted to observe positions in each of the three domains.

The three domains of human proinsulin are shown below. Residues 1-32 comprise the B chain of insulin, shown in blue. Residues 33-65 comprise the C-peptide signaling domain, shown in red. Residues 66-86 comprise the A chain of insulin, shown in green.

MGSSSHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYLVCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGPGA ⁵¹GSLQPLALEG ⁶¹SLQKRIVEEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Summary of NMR Labeling Schemes

Every residue observed in the structural studies of human proinsulin within bacterial inclusion bodies is underlined in the sequence below. Residues from A, B, and C chains of proinsulin were observed using the REDOR filtering method to determine the most likely secondary structure at the targeted residues.

MGSSSHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYLVCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGPGA ⁵¹GSLQPLALEG ⁶¹SLQKRIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Following is a list of the expected secondary structures for each labeling scheme based on the solution NMR structure of DKP-proinsulin.

Ala_{14,57} (double α helical) labeling: 1-¹³C Ala, ¹⁵N Leu

Leu_{15,78} (double α helical) labeling: 1-¹³C Leu, ¹⁵N Tyr

Leu_{11,17} (double α helical) labeling: 1-¹³C Leu, ¹⁵N Val

Leu₅₆ (single α helical) labeling: 1-¹³C Leu, ¹⁵N Ala

Gly₆₆ (single coil) labeling: 1-¹³C Gly, ¹⁵N Ile

Gly₂₃ (single coil) labeling: 1-¹³C Gly, ¹⁵N Phe

Gly₄₉ (single coil) labeling: 1-¹³C Gly, ¹⁵N Ala

Leu₄₄ (single coil) labeling: 1-¹³C Leu, ¹⁵N Gly

Ala₅₀ (single coil) labeling: 1-¹³C Ala, ¹⁵N Gly

NMR Experimental Parameters

The following parameters were used for all samples, and are identical to those discussed in Chapter 3. Data were obtained with a 9.4 T instrument (Agilent Infinity Plus) and a triple-resonance MAS probe whose rotor was cooled with nitrogen gas at $-20\text{ }^{\circ}\text{C}$. Experimental parameters included: (1) 8.0 kHz MAS frequency; (2) 5 μs ¹H $\pi/2$ pulse and 2 ms cross-polarization time with 50 kHz ¹H field and 70-80 kHz ramped ¹³C field; (3) 1 ms rotational-echo double-resonance (REDOR) dephasing time with a 9 μs ¹³C π pulse at the end of each rotor period except the last period and for some data, a 12 μs ¹⁵N π pulse at the center of each rotor

period; and (4) ^{13}C detection with 90 kHz two-pulse phase modulation ^1H decoupling (which was also on during the dephasing time); and (5) 0.8 sec pulse delay. Data were acquired without (S_0) and with (S_1) ^{15}N π pulses during the dephasing time and respectively represented the full ^{13}C signal and the signal of ^{13}C s not directly bonded to ^{15}N nuclei. The $S_0 - S_1$ (ΔS) difference signal was therefore dominated by the labeled ^{13}CO s in the sequential pairs targeted by the labeling. Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm so that the ^{13}CO shifts could be directly compared to those of soluble proteins(3).

Experimental Results

On the following pages, the labeling schemes are summarized and for each sample S_0 , S_1 , and ΔS spectra are shown. The samples are grouped according to the 1- ^{13}C label, as it is most informative to be able to compare the spectra with the same ^{13}C label. For each S_0/S_1 figure (Figure 4-1, 4-3, 4-5) it is expected that the S_0 signal, which represents the total ^{13}C spectrum, should look the same between each sample in the given figure as the samples were prepared in parallel. The S_1 spectra (shown in red in Figures 4-1, 4-3, 4-5) look different depending on the conformation and number of dephased residues. The ΔS spectra displayed in Figures 4-2, 4-4, and 4-6 provide information about the secondary structure of the protein at the targeted residues.

$1\text{-}^{13}\text{C}$ Leu Labeling Schemes

Leu_{11,17} (double α helical) labeling: $1\text{-}^{13}\text{C}$ Leu, ^{15}N Val

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYL²¹VCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGP⁵¹GA ⁵¹GSLQPLALEG ⁶¹SLQKRGIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Leu_{15,78} (double α helical) labeling: $1\text{-}^{13}\text{C}$ Leu, ^{15}N Tyr

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYL²¹VCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGP⁵¹GA ⁵¹GSLQPLALEG ⁶¹SLQKRGIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Leu₄₄ (single coil) labeling: $1\text{-}^{13}\text{C}$ Leu, ^{15}N Gly

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYL²¹VCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGP⁵¹GA ⁵¹GSLQPLALEG ⁶¹SLQKRGIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

Leu₅₆ (single α helical) labeling: $1\text{-}^{13}\text{C}$ Leu, ^{15}N Ala

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH ¹¹LVEALYL²¹VCG ²¹ERGFFYTPKT ³¹RREAEDLQVG
⁴¹QVELGGGP⁵¹GA ⁵¹GSLQPLALEG ⁶¹SLQKRGIVEQ ⁷¹CCTSICSLYQ ⁸¹LENYCN

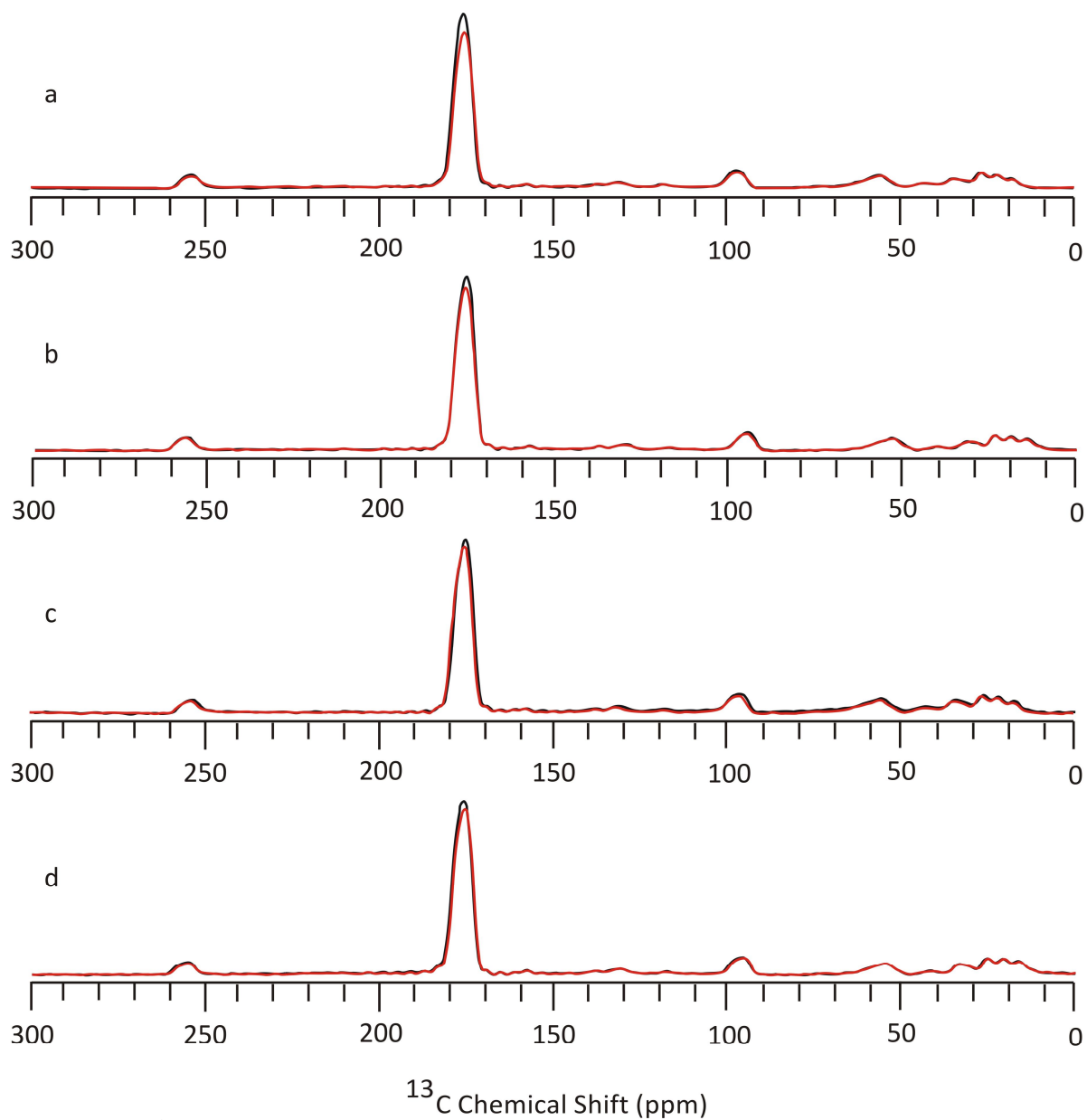


Figure 4-1: $1-^{13}\text{C}$ Leu S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1-^{13}\text{C}$ Leu and ^{15}N Val, b) $1-^{13}\text{C}$ Leu and ^{15}N Tyr, c) $1-^{13}\text{C}$ Leu and ^{15}N Gly, and d) $1-^{13}\text{C}$ Leu and ^{15}N Ala.

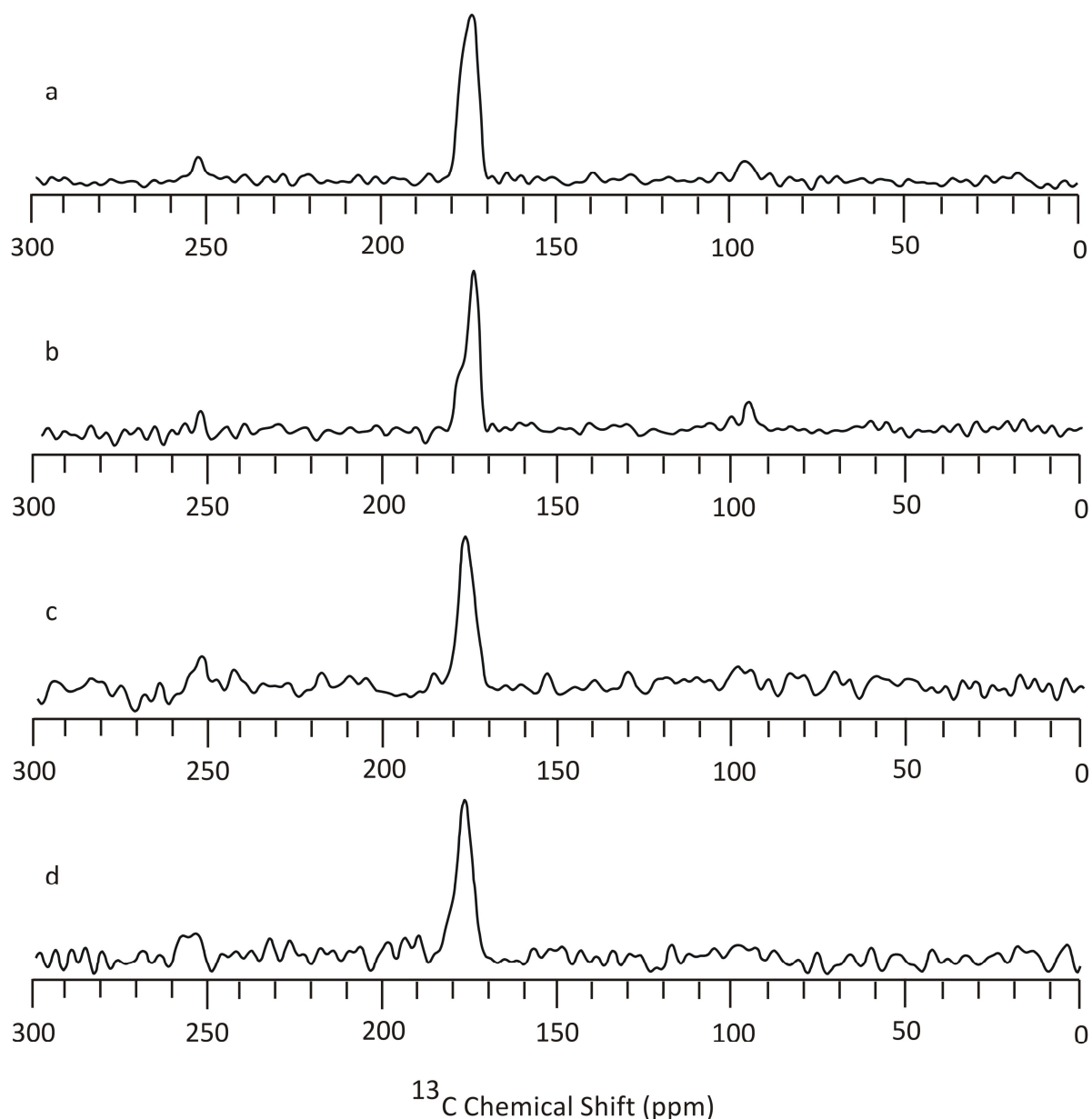


Figure 4-2: $1\text{-}^{13}\text{C}$ Leu REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Leu and ^{15}N Val, b) $1\text{-}^{13}\text{C}$ Leu and ^{15}N Tyr, c) $1\text{-}^{13}\text{C}$ Leu and ^{15}N Gly, and d) $1\text{-}^{13}\text{C}$ Leu and ^{15}N Ala.

$1\text{-}^{13}\text{C}$ Ala Labeling Schemes

Ala_{14,57} (double α helical) labeling: $1\text{-}^{13}\text{C}$ Ala, ^{15}N Leu

MGSSSHHHHHSSGLDPVL ¹FVNQHLCGSH¹¹LVEALYLVCG²¹ERGFFYTPKT³¹RREAEDLQVG
⁴¹QVELGGGPGA⁵¹GSLQPLALEG⁶¹SLQKRGIVEQ⁷¹CCTSICSLYQ⁸¹LENYCN

Ala₅₀ (single coil) labeling: $1\text{-}^{13}\text{C}$ Ala, ^{15}N Gly

MGSSSHHHHHSSGLDPVL ¹FVNQHLCGSH¹¹LVEALYLVCG²¹ERGFFYTPKT³¹RREAEDLQVG
⁴¹QVELGGGPGA⁵¹GSLQPLALEG⁶¹SLQKRGIVEQ⁷¹CCTSICSLYQ⁸¹LENYCN

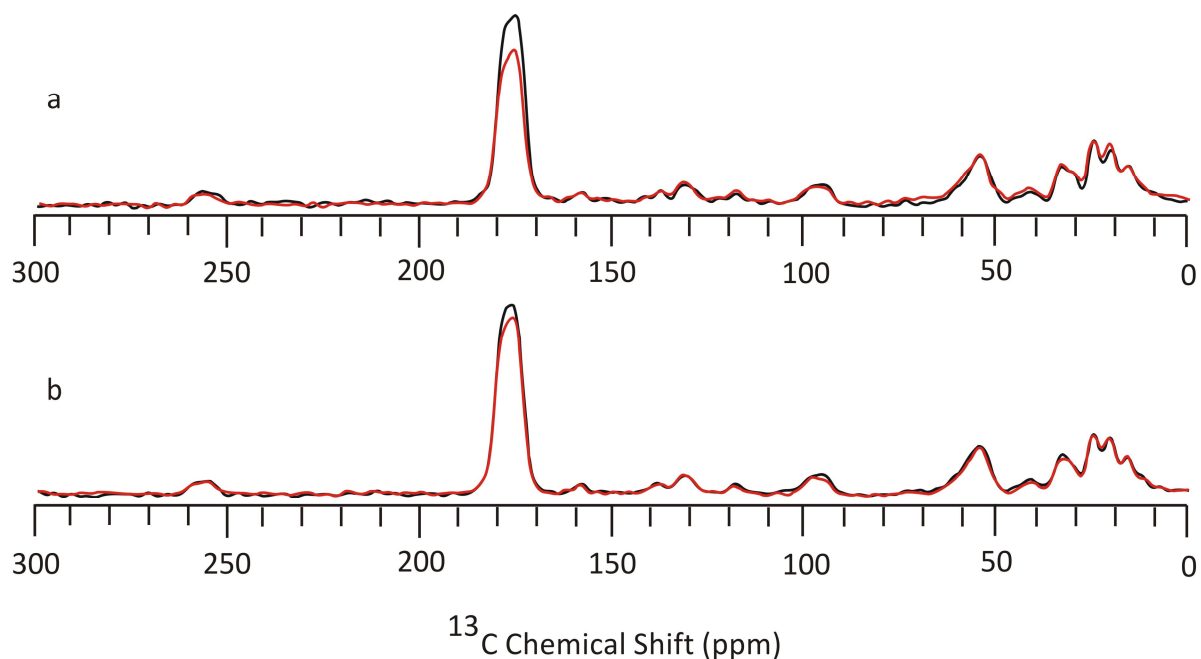


Figure 4-3: $1\text{-}^{13}\text{C}$ Ala S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Ala and ^{15}N Leu, and b) $1\text{-}^{13}\text{C}$ Ala and ^{15}N Gly.

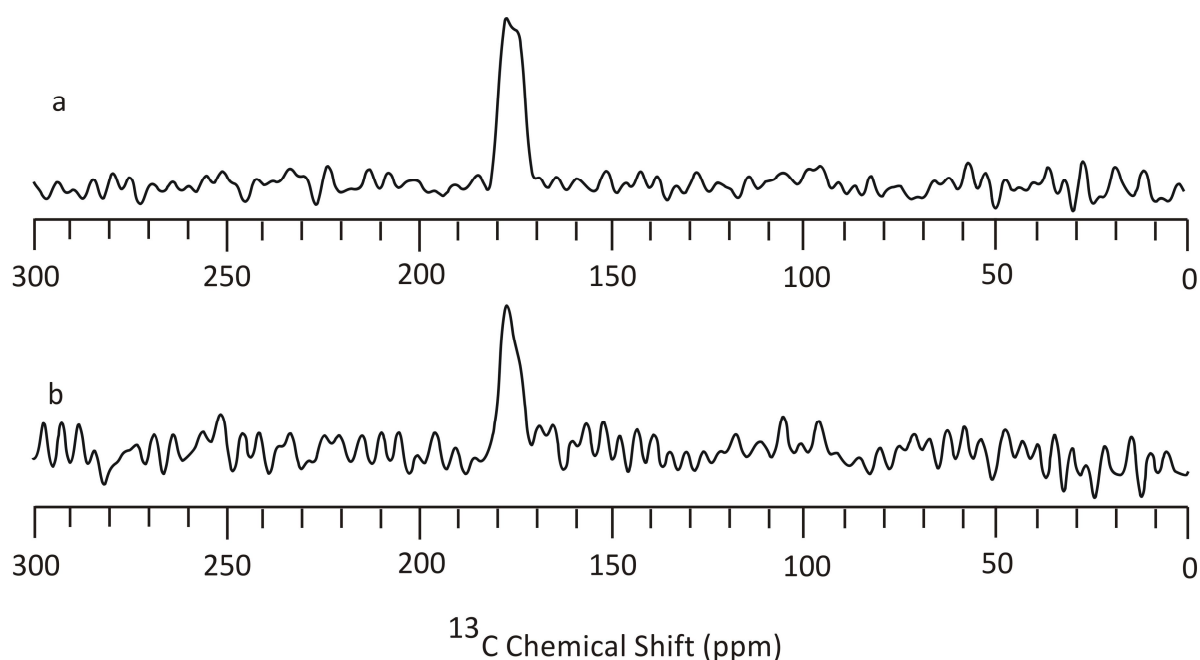


Figure 4-4: $1\text{-}^{13}\text{C}$ Ala REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Ala and ^{15}N Leu, and b) $1\text{-}^{13}\text{C}$ Ala and ^{15}N Gly.

$1\text{-}^{13}\text{C}$ Gly Labeling Schemes

Gly₂₃ (single coil) labeling: $1\text{-}^{13}\text{C}$ Gly, ^{15}N Phe

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH¹¹ LVEALYLVCG²¹ ERGFFYTPKT³¹ RREAEDLQVG
⁴¹QVELGGGPGA⁵¹ GSLQPLALEG⁶¹ SLQKRGIVEQ⁷¹ CTSICSLYQ⁸¹ LENYCN

Gly₄₉ (single coil) labeling: $1\text{-}^{13}\text{C}$ Gly, ^{15}N Ala

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH¹¹ LVEALYLVCG²¹ ERGFFYTPKT³¹ RREAEDLQVG
⁴¹QVELGGGPGA⁵¹ GSLQPLALEG⁶¹ SLQKRGIVEQ⁷¹ CTSICSLYQ⁸¹ LENYCN

Gly₆₆ (single coil) labeling: $1\text{-}^{13}\text{C}$ Gly, ^{15}N Ile

MGSSSHHHHHHSSGLDPVL ¹FVNQHLCGSH¹¹ LVEALYLVCG²¹ ERGFFYTPKT³¹ RREAEDLQVG
⁴¹QVELGGGPGA⁵¹ GSLQPLALEG⁶¹ SLQKRGIVEQ⁷¹ CTSICSLYQ⁸¹ LENYCN

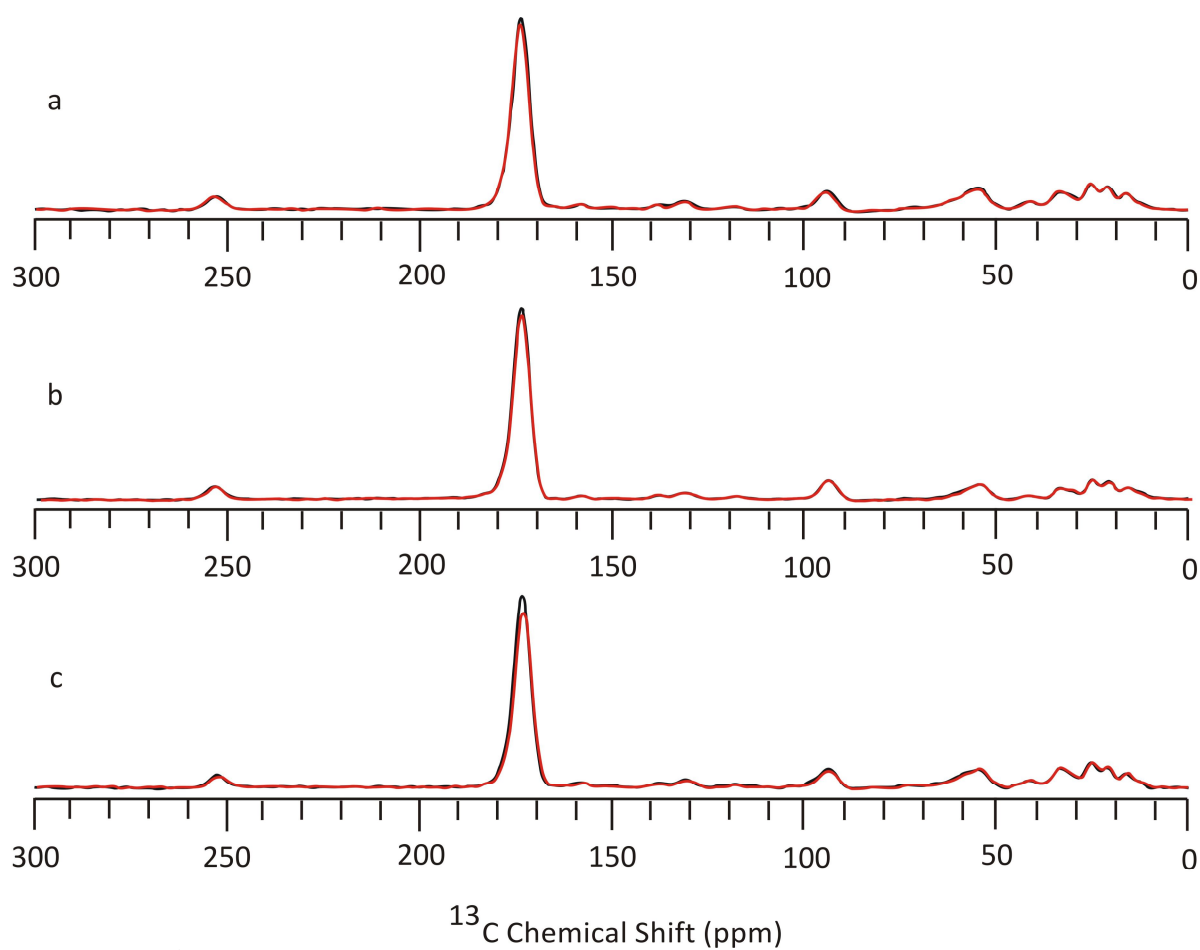


Figure 4-5: $1\text{-}^{13}\text{C}$ Gly S_0 (black) and S_1 (red) REDOR spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 5th order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Phe, b) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala, and c) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ile.

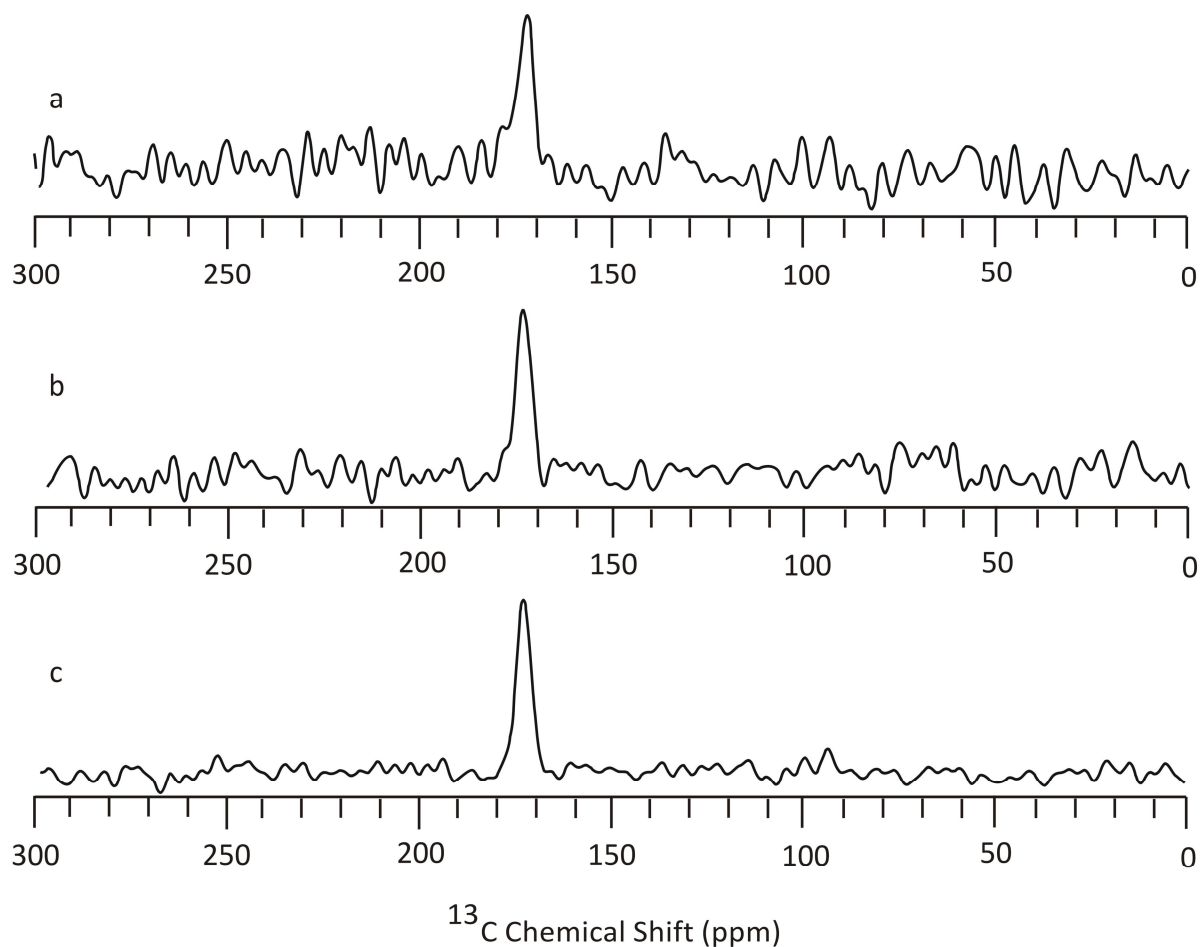


Figure 4-6: $1\text{-}^{13}\text{C}$ Gly REDOR ΔS spectra of human proinsulin inclusion body samples. Each spectrum is the result of 50,000 S_0 scans – 50,000 S_1 scans. The spectra are processed with 100 Hz of Gaussian line broadening and a 3rd order baseline correction. The spectra correspond to fully hydrated insoluble cell pellets from *E. coli* induced to express human proinsulin labeled with a) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Phe, b) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala, and c) $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ile .

Summary of experimental results

Table 4-1: Analysis and deconvolution of ΔS SSNMR spectra of human proinsulin labeled with 1- ^{13}C Leu (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Spectral deconvolution was conducted for Leu_{11,17} and Leu_{15,78} with two Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 4-7. Deconvolution was not meaningful for the Leu₄₄ and Leu₅₆ samples because the ΔS spectra were broad and relatively featureless. The conformations designated are assigned based on characteristic ^{13}CO chemical shifts for different Leu secondary structures which have Gaussian distributions as follows: coil = 176.9 ± 1.7 ppm, helical = 178.5 ± 1.3 ppm, β strand = 175.7 ± 1.5 ppm (4). In refDB, “helical” is defined as $[-120^\circ < \phi < -34^\circ \text{ AND } -80^\circ < \psi < 6^\circ]$. “beta” or β as presented in the table is defined as $[-180^\circ < \phi < -40^\circ \text{ OR } 160^\circ < \phi \leq 180^\circ] \text{ AND } [70^\circ < \psi < 180^\circ \text{ OR } -180^\circ < \psi < -170^\circ]$. “coil” is defined as “everything else”(5).

Position	Peak Information			
	Chemical Shift (ppm)	FWHM (ppm)	Integrated Signal Intensity	Secondary Structure
Leu _{11,17}	174.8	4.1	266	β
	178.0	3.5	148	helical
Leu _{15,78}	174.3	3.5	191	β
	178.1	3.7	69	helical
Leu ₄₄	177.0	6.2	128	coil
Leu ₅₆	176.8	5.8	148	coil

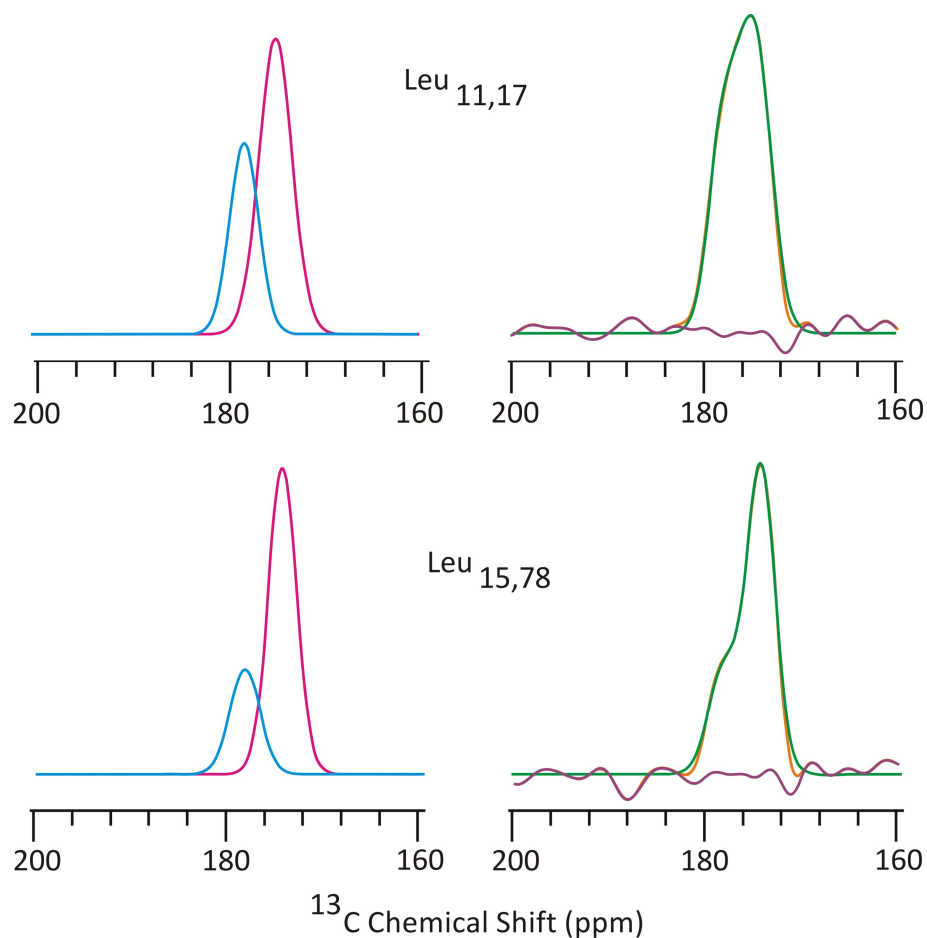


Figure 4-7: Deconvolutions of ΔS spectra are displayed for human proinsulin ICP samples labeled with $1\text{-}^{13}\text{C}$ Leu. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.

Table 4-2: Analysis and deconvolution of ΔS SSNMR spectra of human proinsulin labeled with 1- ^{13}C Ala (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Spectral deconvolution was conducted for Ala_{14,57} and Ala₅₀ with two Gaussian line shapes whose peak shifts, line widths, and intensities were independently varied until there was minimal difference between the sum of the line shapes and the experimental line shape. For both cases, there was excellent agreement between the best-fit deconvolution sum line shape and the experimental line shape, as illustrated in Figure 4-8. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Ala secondary structures which have Gaussian distributions as follows: coil = 177.7 ± 1.6 ppm, helical = 179.4 ± 1.3 ppm, β strand = 176.1 ± 1.5 ppm (4). Please see the caption for Table 4-1 for an explanation of helical, β strand, and coil in terms of dihedral angles.

Position	Peak Information			
	Chemical Shift (ppm)	FWHM (ppm)	Integrated Signal Intensity	Secondary Structure
Ala _{14,57}	174.5	4.2	90	β
	178.3	4.0	92	helical
Ala ₅₀	174.0	3.2	23	β
	177.5	4.0	59	coil

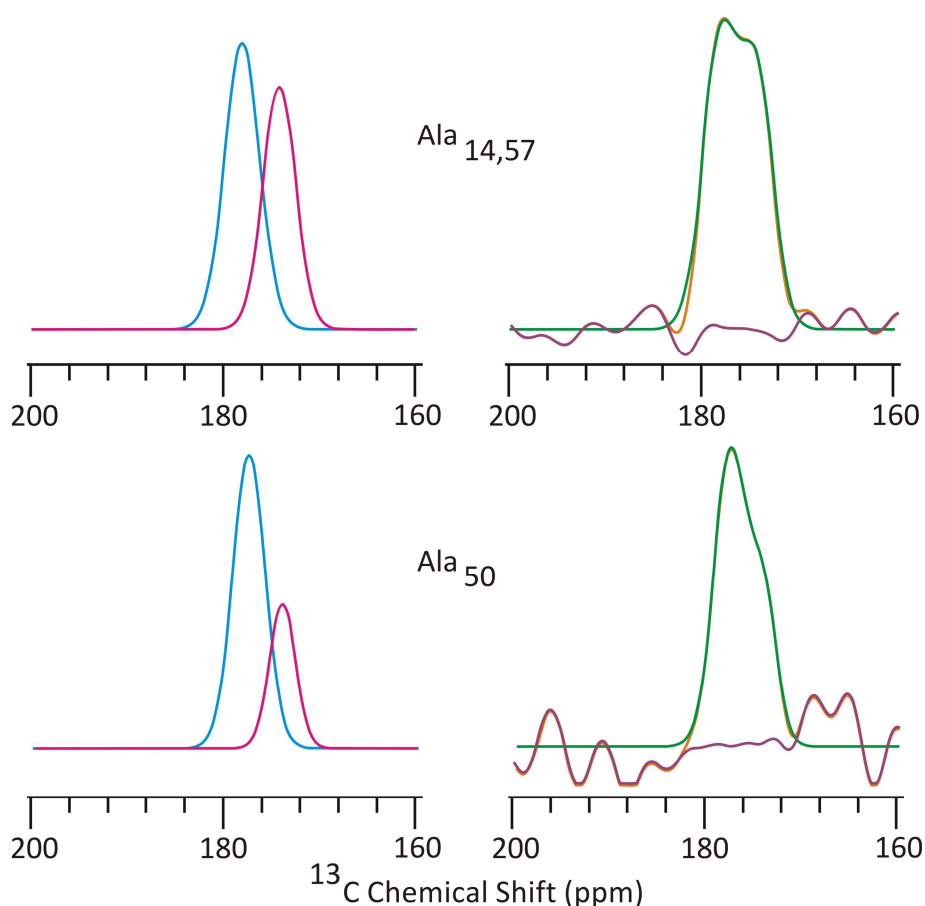


Figure 4-8: Deconvolutions of ΔS spectra are displayed for human proinsulin ICP samples labeled with 1- ^{13}C Ala. The fitting of each deconvolution is shown on the right, where orange represents the experimental line, green is the best-fit deconvolution sum, and purple is the difference between the two.

Table 4-3: Analysis of ΔS SSNMR spectra of human proinsulin labeled with 1- ^{13}C Gly (and various ^{15}N labeling, as indicated previously) within insoluble cell pellets. Deconvolution was not meaningful for the spectra as the peaks are relatively featureless. The conformations designated are assigned based on characteristic ^{13}C CO chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (4). Please see the caption for Table 4-1 for an explanation of helical, β strand, and coil in terms of dihedral angles.

Position	Peak Information			
	Chemical Shift (ppm)	FWHM (ppm)	Integrated Signal Intensity	Secondary Structure
Gly ₂₃	172.5	5.9	56	β
Gly ₄₉	173.6	5.1	80	coil
Gly ₆₆	172.9	4.8	137	β

In the DKP-proinsulin structure, insulin-like structure was observed by solution NMR for residues within the A and B chains, and much less ordered structure was observed for the C chain(2). The SSNMR results obtained on human proinsulin within bacterial inclusion bodies provides a similar result, with all of the chemical shifts that correlate with random coil conformation being obtained on samples labeled to observe residues within the C chain of proinsulin. Residues Leu₄₄, Leu₅₆, Ala₅₀, and Gly₄₉ were all observed to have random coil correlated chemical shifts.

Also interesting is the observation of chemical shifts indicative of β -strand secondary structure for many of the samples, including the samples that are labeled to observe Leu_{11,17} and Leu_{15,78}. Both Leu_{11,17} and Leu_{15,78} are expected to have α -helical secondary structure according to the solution NMR structure. For the sample observing Leu_{11,17}, 100% of the signal is indicative of β -strand conformation, and for Leu_{15,78}, >70% of the signal lies within the β -strand region of the spectrum. Other samples that showed peaks in the β -strand region of the spectrum included Ala_{14,57} (a mixture of β -strand and helical shifts), Ala₅₀ (a mixture of β -strand and coil shifts), Gly₂₃ (β -strand), and Gly₆₆ (β -strand).

The SSNMR results of the study of human proinsulin within inclusion bodies yielded quite different results than previous structural studies of recombinant protein in inclusion bodies in the Weliky group. Previous studies on the influenza fusion protein FHA2 yielded highly helical structure when studied in both whole *E. coli* cells and insoluble cell pellets(6). Studies of

the Hairpin protein which represents the helix-loop-helix region of HIV-1 gp41 ectodomain also yielded highly helical structure within whole *E. coli* cells and insoluble cell pellets(7). The Fgp41 construct which represents most of the ectodomain of HIV-1 gp41 including the fusion peptide through the C-terminal helix also adopts a highly helical structure within whole *E. coli* cells(8). The results from the study of proinsulin are the first results in our group that have shown non-helical structure within inclusion bodies. We now have evidence that the structure of recombinant proteins within inclusion bodies varies greatly between different proteins. From the studies of Hairpin, Fgp41, and FHA2 within inclusion bodies, a large amount of helical structure was retained, suggesting natively folded protein was present. The data presented in this chapter for human proinsulin is evidence for mostly unfolded protein within the inclusion bodies. Our group's work suggests that there are different types of inclusion bodies, with either (at least partially) folded or unfolded protein, or a mixture of both. In conclusion, SSNMR and the REDOR pulse sequence provides some insight into the structure of recombinant protein within bacterial inclusion bodies, an area which has been highly speculative until now.

REFERENCES

REFERENCES

1. Cowley, D. J., and Mackin, R. B. (1997) Expression, purification and characterization of recombinant human proinsulin, *Febs Letters* 402, 124-130.
2. Yang, Y., Hua, Q.-x., Liu, J., Shimizu, E. H., Choquette, M. H., Mackin, R. B., and Weiss, M. A. (2010) Solution Structure of Proinsulin CONNECTING DOMAIN FLEXIBILITY AND PROHORMONE PROCESSING, *Journal Of Biological Chemistry* 285, 7847-7851.
3. Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR, *J. Magn. Reson.* 162, 479-486.
4. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts, *J. Biomol. NMR* 25, 173-195.
5. Willard, L., Ranjan, A., Zhang, H. Y., Monzavi, H., Boyko, R. F., Sykes, B. D., and Wishart, D. S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality, *Nucleic Acids Research* 31, 3316-3319.
6. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy, *J. Am. Chem. Soc.* 130, 12568-12569.
7. Curtis-Fisk, J. (2009) Structural studies of the Influenza and HIV viral fusion proteins and bacterial inclusion bodies, Ph. D. Thesis, Michigan State University.
8. Vogel, E. P., Curtis-Fisk, J., Young, K. M., and Weliky, D. P. (2011) Solid-State Nuclear Magnetic Resonance (NMR) Spectroscopy of Human Immunodeficiency Virus gp41 Protein That Includes the Fusion Peptide: NMR Detection of Recombinant Fgp41 in Inclusion Bodies in Whole Bacterial Cells and Structural Characterization of Purified and Membrane-Associated Fgp41, *Biochemistry* 50, 10013-10026.

APPENDICES

APPENDIX A

The Entire Ectodomain of gp41 – Fgp41:Fragment2

There has been considerable interest in recent years in determining the importance of the “membrane proximal external region” or “MPER” of gp41 in the process of membrane fusion, as it has been recognized as a target of several broadly neutralizing antibodies(1). It has also been hypothesized that the hydrophobic residues in the MPER interact with the viral membrane, inducing curvature(2). More recent studies have suggested that the C-terminus of the MPER in tandem with the N-terminus of the transmembrane domain are responsible for membrane disruption of the viral particle(3).

To investigate the MPER in the context of the ectodomain of gp41, a construct that is merely an extension of Fgp41 was studied. The construct “Fragment2” contains the entire ectodomain of gp41, and the sequence is from the same patient sera as Fgp41. Initial attempts at working with this construct yielded no discernable recombinant protein even after many attempts at purification under a variety of conditions. For this reason, mutations were performed to mutate the Cys residues to Ala using the same primers as were used to create Fgp41_{noCys}. Successful mutations were confirmed by DNA sequencing. Information regarding the constructs is shown below.

DNA sequence of Fgp41:Fragment2

```
ATGGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGC
GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTGCACCAGCAAAGCAATTTGCT
GAAGGCTATAGAGGCTCAACAGCATCTGTTGAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAG
AGTCCTGGCTGTGGAAAGATACCTACAGGATCAACAGCTCCTGGGAATTTGGGGCTGCTCTGGAAAAC
CATCTGCACCTCTTTTGTGCCCTGGAACAATAGTTGGAGTAACAAGACTTATAATGAGATTTGGGACAAC
ATGACCTGGTTGCAATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTACTTGAAGAC
TCGCAGAACCCAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAGATAAATGGGCAAATTTGTGGAA
TTGGTTTAGCATAACAACTGGCTGTGGTATATAAAGCTCGAGCACCACCACCACCACCACTGA
```

Protein Sequence of Fgp41:Fragment2

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVHQQSNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGC SGKLICTSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKWANLWNWFSITNWLWYIKLEHHHHH

First C to A mutation:

Forward primer: GAATTTGGGGCGCCTCTGGAAAAC

Reverse primer: GTTTCCAGAGGCGCCCCAAATTC

DNA sequence of Fgp41:Fragment2 after first C to A mutation:

ATGGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGC
GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTGCACCAGCAAAGCAATTTGCT
GAAGGCTATAGAGGCTCAACAGCATCTGTTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAG
AGTCCTGGCTGTGGAAGATACTACAGGATCAACAGCTCCTGGGAATTTGGGGCCCCTCTGGAAAAC
CATCTGCACCTCTTTGTGCCCTGGAACAATAGTTGGAGTAACAAGACTTATAATGAGATTTGGGACAAC
ATGACCTGGTTGCAATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTACTTGAAGAC
TCGCAGAACCGAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAGATAAATGGGCAAATTTGTGGAA
TTGGTTTAGCATAACAACTGGCTGTGGTATATAAAGCTCGAGCACCACCACCACCACTGA

Protein sequence of Fgp41:Fragment2 after first C to A mutation:

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVHQQSNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGASGKLICTSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNQQEKNEQDL
LALDKWANLWNWFSITNWLWYIKLEHHHHH

Second C to A mutation:

Forward primer: CTCATCGCCACCTCTTTTGTGC

Reverse primer: GCACAAAAGAGGTGGCGATGAG

DNA sequence of Fgp41:Fragment2 after second C to A mutation:

ATGGCAGTTGGACTAGGAGCTGTCTTCCTTGGGTTCTTGGGAGCAGCAGGGAGCACTATGGGCGCGGC
GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGCATAGTGCACCAGCAAAGCAATTTGCT
GAAGGCTATAGAGGCTCAACAGCATCTGTTGAAACTCACGGTCTGGGGTATTAAACAGCTCCAGGCAAG
AGTCCTGGCTGTGGAAGATACTACAGGATCAACAGCTCCTGGGAATTTGGGGCGCCTCTGGAAAAC
CATCGCCACCTCTTTTGTGCCCTGGAACAATAGTTGGAGTAACAAGACTTATAATGAGATTTGGGACAAC
ATGACCTGGTTGCAATGGGATAAAGAAATTAGCAATTACACAGACACAATATACAGGCTACTTGAAGAC

TCGCAGAACAGCAGGAAAAGAATGAACAAGACTTATTGGCATTAGATAAATGGGCAAATTTGTGGAA
TTGGTTTAGCATAACAACTGGCTGTGGTATATAAAGCTCGAGCACCACCACCACCACCACTGA

Protein sequence of Fgp41:Fragment2 after second C to A mutation:

AVGLGAVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVHQQSNNLLKAIEA
QQHLLKLTVWGIKQLQARVLAVERYLQDQQLLGIWGASGKLIATSFVPWN
NSWSNKTYNEIWDNMTWLQWDKEISNYTDTIYRLLEDSQNNQKEKNEQDL
LALDKWANLWNWFSITNWLWYIKLEHHHHHH

The preceding protein sequence will be referred to as Fgp41:Fragment2_{noCys} for the remainder of this appendix. All of the work following was done utilizing the Fgp41:Fragment2_{noCys} plasmid transformed into BL21(DE3) Rosetta2 *E. coli* cells. Expression parameters established for Fgp41 were utilized, including inducing protein expression with [IPTG] = 2 mM and expression at 37 °C for a period of 6 hours.

Purification #1

5.0 grams of cells induced to express Fgp41:Fragment2_{noCys} were sonicated in 40 mL of buffer containing 50 mM sodium phosphate at pH 8.0, 300 mM NaCl, and 20 mM imidazole. The lysate was centrifuged at 50000g for 20 minutes at 4°C. The soluble material was utilized in a purification (same as Purification #1 for Fgp41_{noCys}) however there was no Fgp41:Fragment2_{noCys} present in the eluents. This is in line with lane 1 shown in the SDS-PAGE gel below in Figure A-1, where there is not a band corresponding to Fgp41:Fragment2_{noCys} in the soluble material after sonication in phosphate buffer. The insoluble material was sonicated in 40 mL of urea lysis buffer, which contained 50 mM sodium phosphate at pH 8.0, 300 mM NaCl, 20 mM imidazole, and 8 M urea. The lysate was centrifuged at 50000g for 20 minutes at

4°C, and the supernatant was combined with 0.50 mL of prepared His-Select cobalt resin. The insoluble material after sonication in urea was saved to run in SDS-PAGE, and can be seen below in lane 2, Figure A-1. After one hour of mixing at room temperature, the resin was loaded onto a column and washed with 6 mL of fresh lysis buffer, and the last of these washes was run on the SDS-PAGE (lane 4, Figure A-1). The washes were done until the A₂₈₀ reading of the eluent was small and constant (about 0.2 mg/mL). Protein was eluted from the resin with urea elution buffer (50 mM sodium phosphate at pH 8.0, 300 mM NaCl, 250 mM imidazole, and 8 M urea). The elution can be seen in lane 8, Figure A-1 below.

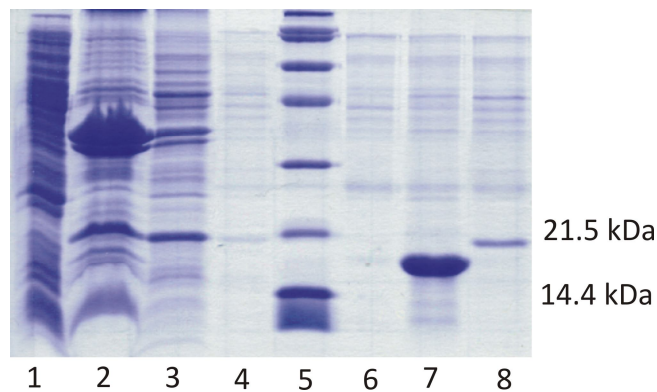


Figure A-1: Examination of the solubility of Fgp41:Fragment2_{noCys} under different conditions. The lanes are as follows: 1) proteins soluble in sodium phosphate buffer, 2) insoluble material after sonication in urea, 3) unbound protein in “flow through”, 4) protein eluted with wash buffer, 5) Broad Molecular Weight Standards with important mass markers on the right-hand side of the figure, 6) proteins present in an eluent from the purification of cells containing the empty pET24a+ plasmid as a control, 7) purified Fgp41_{noCys} (as shown in Figure 2-11), 8) protein eluted in 250 mM imidazole containing buffer. The darkest band corresponds to Fgp41:Fragment2_{noCys}.

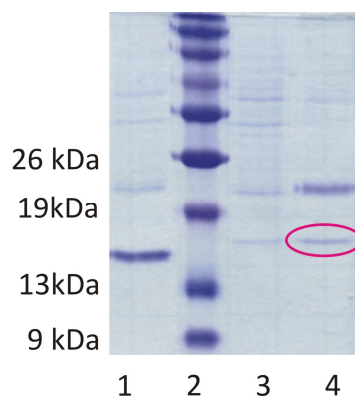


Figure A-2: Comparison of Fgp41_{noCys} and Fgp41:Fragment2_{noCys} both purified using urea. The lanes are as follows: 1) Fgp41_{noCys} elution fraction, 2) Spectra Molecular Weight Standards, 3) Fgp41:Fragment2_{noCys} elution fraction, and 4) Fgp41:Fragment2_{noCys} elution fraction. The gel shift due to the molecular weight difference is clearly observed in this gel. The band that corresponds to Fgp41:Fragment2_{noCys} can be seen most clearly in circled in Lane 4.

It is clear from the SDS-PAGE of Fgp41:Fragment2_{noCys} shown in Figure A-1 that for some reason, Fgp41:Fragment2_{noCys} is both difficult to solubilize (as shown by a large amount of the protein present in the insoluble fraction after sonication in 8M urea) and difficult to purify by affinity chromatography (as shown by a distinct band present in lane 3 – proteins which had not bound to the resin, as well as the distinct, yet faint band shown in lane 8 – proteins present at the end of the purification protocol). The observation that some of the protein has not bound to the resin could indicate that the polyhistidine tag is protected from the bulk solution and inaccessible to the Cobalt resin.

Purification #2

This approach utilized 6 M guanidine hydrochloride as the denaturant in the lysis buffer. Guanidine hydrochloride is a common denaturant utilized in protein purification. The drawback is that in the presence of SDS, guanidine hydrochloride precipitates. This is an issue because SDS PAGE is usually used to analyze the effectiveness of protein purification protocols.

2.5 grams of cells induced to express Fgp41:Fragment2_{noCys} were sonicated (4 rounds of 1 minute, 80% amplitude, 0.8 sec on, 0.2 sec off, on ice) in 40 mL of buffer containing 6 M guanidine HCl, 50 mM sodium phosphate, 300 mM NaCl, and 20 mM imidazole at pH 8.0. The lysate was centrifuged at 50000g for 20 minutes at 4°C. The supernatant was combined with 0.25 mL of prepared His-Select cobalt resin and allowed to incubate at room temperature for one hour. The resin was loaded back onto the column and washed with 10 × 0.25 mL lysis buffer. The protein was then eluted from the column with 6 × 0.25 mL elution buffer (6 M guanidine HCl, 50 mM sodium phosphate, 300 mM NaCl, and 250 mM imidazole at pH 8.0). All elution fractions were placed into a 3500 MWCO dialysis cassette and dialyzed in 0.5 L 1X SDS/Tris/Glycine running buffer for ~10 minutes. This caused precipitation of guanidine hydrochloride. The cassette was removed from the dialysis buffer, a small amount of the protein solution was removed to run on a gel; the cassette was then rinsed and placed into 0.5 L 8M urea in PBS overnight with stirring. The protein solution was removed from the cassette, a small amount set aside to run a gel, and 500 µL was concentrated to 10 µL to run on a gel.

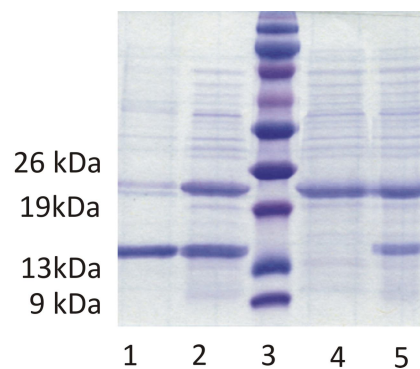


Figure A-3: Results of the purification of Fgp41:Fragment2_{noCys} with guanidine HCl as the denaturant. Lane 1) Fgp41 purified with urea, Lane 2) Fgp41 purified with guanidine HCl, Lane 3) Spectra Molecular Weight Standards, 4) concentrated Fgp41:Fragment2_{noCys} elution fractions after dialysis into 8M urea, Lane 5) mixture of Fgp41:Fragment2_{noCys} and Fgp41_{noCys}

after dialysis into 8M urea. The large band between molecular weight markers 19 and 26 kDa can most likely be attributed to the chloramphenicol resistance protein.

Despite the lack of a developed purification protocol that yields a large enough amount of purified Fgp41:Fragment2_{noCys}, I believe it is worth pursuing. SSNMR data has indicated that Fgp41:Fragment2_{noCys} is produced in amounts approximately equal to Fgp41_{noCys}. By inducing cells to produce isotopically labeled recombinant protein in the exact same manner, and running NMR experiments on the insoluble cell pellets (details of which are described in Chapter 3), the relative levels of recombinant protein expression can be examined. Figure A-4 displays the REDOR S_0 spectra of 1- ^{13}C , ^{15}N Leu labeled insoluble cell pellets induced to express either Fgp41_{noCys} or Fgp41:Fragment2_{noCys}; the spectra represent the signal from all ^{13}C present in the samples.

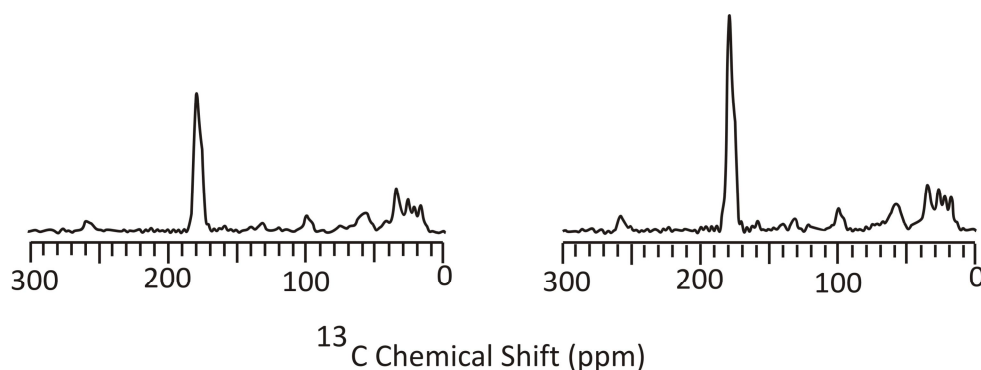


Figure A-4: REDOR S_0 spectra for 1- ^{13}C , ^{15}N Leu labeled Fgp41:Fragment2_{noCys} insoluble cell pellet (left) and 1- ^{13}C , ^{15}N Leu labeled Fgp41_{noCys} insoluble cell pellet (right). The spectra are each the sum of 50,000 REDOR S_0 scans and were both processed with 100 Hz Gaussian line broadening and a 5th order baseline correction. The spectra are scaled so that the intensity in the 0 to 90 ppm region is the same (as this should be unaffected by isotopic labeling and recombinant protein production).

By comparing the integrated signal intensities from the spectra displayed in Figure A-4, a relative level of expression for each protein construct can be calculated. Table A-1 below displays numerical data obtained from processing of the spectra.

Table A-1: Numerical data obtained from the REDOR S_0 spectra of $1\text{-}^{13}\text{C}, ^{15}\text{N}$ Leu labeled Fgp41_{noCys} and Fgp41:Fragment2_{noCys} insoluble cell pellets. To calculate the “scaling factor”, the integrated signal intensity in the 0 to 90 ppm region of the spectrum was divided by 1000. This number was then multiplied by the integrated signal intensity in the carbonyl region and the value from the same process for pET24a(+) sample was subtracted to yield the “reduced carbonyl signal”. The reduced carbonyl signal was divided by the number of Leu residues present in the protein constructs to give the “normalized signal”.

Protein Construct	Integrated Signal Intensity (170 → 185 ppm)	Integrated Signal Intensity (0 → 90 ppm)	Scaling Factor	Reduced Carbonyl Signal	Number of Leu Residues in Construct	Normalized Signal
Fgp41 _{noCys}	1041	926	1.08	684	24	28.5
Fgp41:F2 _{noCys}	918	857	1.17	632	26	24.3

From the data presented in Table A-1, we can conclude that Fgp41:Fragment2_{noCys} is being produced at a level comparable to Fgp41_{noCys}, and thus should be amenable to recovery in yields similar to that of Fgp41_{noCys}. Other possibilities for protein purification schemes could include acid or base denaturation, a combination of detergents to solubilize the protein from inclusion bodies, or HPLC purification following a harsh denaturing step such as sonication in glacial acetic acid (4-6).

REFERENCES

REFERENCES

1. Shi, W., Bohon, J., Han, D. P., Habte, H., Qin, Y., Cho, M. W., and Chance, M. R. (2010) Structural Characterization of HIV gp41 with the Membrane-proximal External Region, *Journal Of Biological Chemistry* 285, 24290-24298.
2. Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., and Weissenhorn, W. (2010) Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions, *Plos Pathogens* 6, e1000880.
3. Apellaniz, B., Ivankin, A., Nir, S., Gidalevitz, D., and Nieva, J. L. (2011) Membrane-Proximal External HIV-1 gp41 Motif Adapted for Destabilizing the Highly Rigid Viral Envelope, *Biophysical Journal* 101, 2426-2435.
4. Frankel, S., Sohn, R., and Leinwand, L. (1991) The Use Of Sarkosyl In Generating Soluble-Protein After Bacterial Expression, *Proceedings Of The National Academy Of Sciences Of The United States Of America* 88, 1192-1196.
5. Tao, H., Liu, W., Simmons, B. N., Harris, H. K., Cox, T. C., and Massiah, M. A. (2010) Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS, *Biotechniques* 48, 61-64.
6. Sackett, K., Nethercott, M. J., Shai, Y., and Weliky, D. P. (2009) Hairpin folding of HIV gp41 abrogates lipid mixing function at physiologic pH and inhibits lipid mixing by exposed gp41 constructs, *Biochemistry* 48, 2714-2722.

APPENDIX B

Studies of FHA2 – dependence of secondary structure within membranes on sample pH and the presence of cholesterol

Introduction

The Influenza virus starts the process of viral infection after it enters the target cell through endocytosis after interaction of HA1 (the receptor binding unit of the envelope protein hemagglutinin) with sialic acid receptors(1). Virus/Endosome membrane fusion occurs after a restructuring of HA2 (the fusion subunit of hemagglutinin), and the proposed fusion trigger for the conformational change of HA2 is a drop in pH as is experienced in the late endosome. HA2 is a Type I fusion protein, which has an N-terminal “fusion peptide” region and refolds into a low-energy coiled-coil post fusion(2). FHA2 is a protein construct that represents the entire ectodomain of the Influenza A X31 strain HA2 fusion protein. For a comprehensive introduction to FHA2, including optimization of the expression and purification, as well as structural and functional studies, please refer to Jaime Curtis-Fisk’s dissertation(3). Additionally, it has been suggested that viruses, including influenza, tend to bud from ordered lipid “raft domains” which include higher than average concentrations of membrane components such as sphingolipids and cholesterol(4).

The project presented in this appendix of my dissertation aimed to investigate two questions regarding membrane associated FHA2 structure. 1) How does the structure of FHA2 change with respect to a change in pH (“active” pH of 5.0 vs. physiological pH of 7.4) when it is associated with membranes? 2) How does the structure of FHA2 change with the presence of cholesterol in the membrane?

Unfortunately, the method of protein expression and the inclusion of isotopic labels within the expressed protein had not been entirely understood at the time of these studies. I have since learned (and presented in detail in Chapter 3) that without proper precautions, E.

coli will break down and reincorporate labeled amino acids into other residues. This becomes a problem if REDOR filtering is to be used to determine structural information at specific sites within a protein. However, the results are presented in what follows.

FHA2 Expression

The protocol to produce isotopically labeled influenza virus fusion protein ectodomain FHA2 for NMR experiments was previously developed in the Weliky lab (5). Following is a summary of the methods used. One key feature was initial bacterial growth in rich medium (LB) to high cell densities. Relative to initial growth in minimal medium, protein production was augmented by the cell densities and by the larger number of ribosomes per cell. Bacterial cell cultures were grown in media containing 15 mg/L kanamycin because the pET24a(+) vector contains a gene for kanamycin resistance. Bacterial cells in 1 mL of 80/20 (v/v) H₂O/glycerol were added to two 2.8 L baffled fernbach flasks which each contained 1 L of LB and were capped with a foam plug. Bacterial growth to OD₆₀₀ ~4 occurred during overnight incubation at 37 °C with shaking at 140 rpm. The cell suspensions were centrifuged (10000g, 10 min) and the cell pellets were harvested and then resuspended in a single flask containing 1 L of fresh medium with M9 minimal salts, 2.0 mL of 1.0 M MgSO₄, and 5.0 mL of 50% glycerol solution. Growth resumed after approximately one hour of incubation at 37 °C. At this time, 100 mg/L of 1-¹³C amino acid and 100 mg/L of ¹⁵N amino acid (or 100 mg/L of 1-¹³C, ¹⁵N amino acid) were added to the medium. IPTG was then added to a final concentration of 0.2 mM which induced

expression of FHA2 (6 hours, 23 °C). The cell pellet was harvested after centrifugation and stored at -80 °C. The wet cell mass was ~8 g.

FHA2 purification

Buffers for the purification of FHA2 were as follows:

Lysis Buffer / Wash 1 Buffer: 0.5% N-lauroylsarcosine, 50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH = 8.

Wash 2 Buffer: 0.5% N-lauroylsarcosine, 50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 0.5% β -thio-octyl-glucoside, 0.4% C8E5, pH = 8.

Wash 3 Buffer: 50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 0.5% β -thio-octyl-glucoside, 0.4% C8E5, pH = 8.

Elution Buffer: 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, 0.5% β -thio-octyl-glucoside, 0.4% C8E5, pH = 8.

Optimal purity was obtained using 5.0 grams of cells induced to express FHA2 and 0.5 mL of prepared His-Select Co resin. Cells were sonicated on ice in ~40 mL of lysis buffer using four 1-minute cycles at 80% amplitude with 0.8 seconds on/0.2 seconds off. The cell lysates were then centrifuged at 20,000 rpm for 20 min (at 4°C). The clarified supernatant was combined with 0.5 mL of prepared resin and allowed to mix at room temperature for 1 hour. The resin was loaded onto a column and washed with 3 column volumes each of wash buffers 1,2, and 3. After the washes, the FHA2 was eluted from the resin using buffer containing [imidazole] = 250 mM.

Membrane Reconstitution

For studies of FHA2 using Solid-State NMR, purified FHA2 was reconstituted into lipid vesicles so that the protein could be studied in a biologically relevant environment. The composition of the lipid vesicles utilized in these studies was designed to include a 4:1 ratio of choline : negatively charged lipid headgroups.

A homogeneous mixture of the POPC (27 mg) and POPG (7 mg) lipids and the bTOG (136 mg) detergent was made by: (1) dissolution in chloroform; (2) removal of chloroform by nitrogen gas and overnight vacuum; and (3) dissolution in HEPES/MES buffer. FHA2 (~10 mg) was added to the solution. Dialysis of the bTOG/lipid/FHA2 solution against HEPES/MES buffer removed bTOG with consequent liposome formation with bound FHA2. The lipid mixtures used were either 4:1 POPC:POPG or 8:2:5 POPC:POPG:Chol. Dialysis parameters included: (1) bTOG/lipid/FHA2 solution in 10 KDa MWCO tubing (~15 mL initial volume); (2) 3L buffer volume; and (3) 3 day duration at 4 °C while stirring with one buffer change. The proteoliposome pellet was harvested after centrifugation (50000g, 3 hours) and unbound FHA2 did not pellet under these conditions. The pellet was packed into a 4 mm diameter magic angle spinning (MAS) rotor with ~5 mg FHA2 and ~20 mg total lipid in the 40 μ L active sample volume.

SSNMR Experimental Parameters

Data were obtained with a 9.4 T instrument (Agilent Infinity Plus) and a triple-resonance MAS probe whose rotor was cooled with nitrogen gas at -10 °C. Because of heating from MAS and RF radiation, we expect that water in the sample was liquid rather than solid. Experimental parameters included: (1) 8.0 kHz MAS frequency; (2) 5 μ s ^1H $\pi/2$ pulse and 2 ms cross-polarization time with 50 kHz ^1H field and 70-80 kHz ramped ^{13}C field; (3) 1 or 2 ms rotational-

echo double-resonance (REDOR) dephasing time with a $9\ \mu\text{s}\ ^{13}\text{C}\ \pi$ pulse at the end of each rotor period except the last period and for some data, a $12\ \mu\text{s}\ ^{15}\text{N}\ \pi$ pulse at the center of each rotor period; and (4) ^{13}C detection with 90 kHz two-pulse phase modulation ^1H decoupling (which was also on during the dephasing time); and (5) 0.8 sec pulse delay(6). Data were acquired without (S_0) and with (S_1) $^{15}\text{N}\ \pi$ pulses during the dephasing time and respectively represented the full ^{13}C signal and the signal of ^{13}C s not directly bonded to ^{15}N nuclei. The $S_0 - S_1$ (ΔS) difference signal was therefore dominated by the labeled ^{13}CO s in the sequential pairs targeted by the labeling. Spectra were externally referenced to the methylene carbon of adamantane at 40.5 ppm so that the ^{13}CO shifts could be directly compared to those of soluble proteins(7).

NMR Results

Presented in Figures B1 – B6 are the ΔS spectra corresponding to membrane associated FHA2 samples. The specific details of NMR sample preparation for each sample including labeling scheme, lipid composition, and pH of the sample is presented in the figure captions, as are the number of scans for each experiment.

Tables B1 – B6 contain peak data obtained from the ΔS spectra presented in Figures B1 – B6. The reported secondary structure for each residue is obtained by comparing the peak chemical shift to RefDB (8).

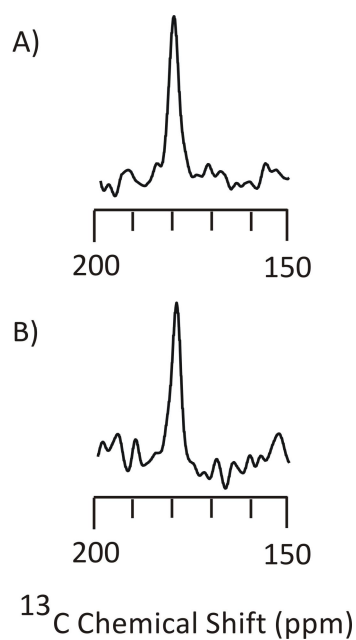


Figure B1: ΔS spectra corresponding to labeling at Phe3 of FHA2 (FHA2 was labeled with 1- ^{13}C Phe and ^{15}N Gly). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [52996 S_0 – 52996 S_1] scans. B) The sample was made in the same was as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [58080 S_0 – 58080 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

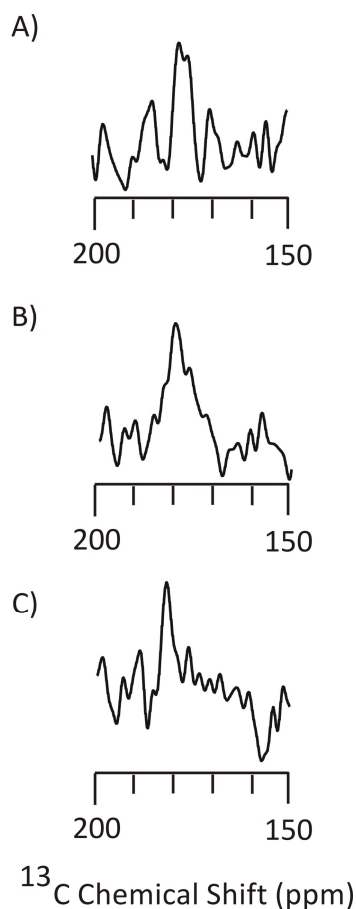


Figure B2: ΔS spectra corresponding to labeling at Gly4 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Gly and ^{15}N Ala). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [149040 S_0 – 149040 S_1] scans. B) The sample was made in the same way as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [167904 S_0 – 167904 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [102928 S_0 – 102928 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

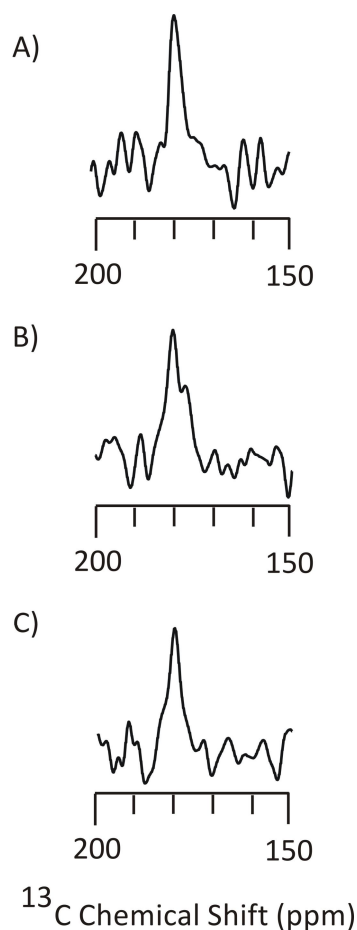


Figure B3: ΔS spectra corresponding to labeling at Ala7 of FHA2 (FHA2 was labeled with 1- ^{13}C Ala and ^{15}N Gly). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [49328 S_0 – 49328 S_1] scans. B) The sample was made in the same way as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [55408 S_0 – 55408 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [96240 S_0 – 96240 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

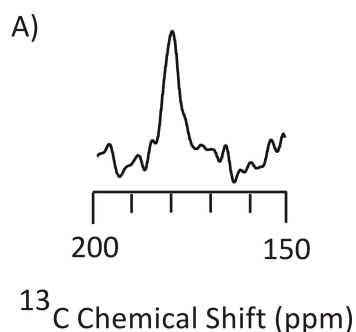


Figure B4: ΔS spectrum corresponding to labeling at Gly16 of FHA2 (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Gly and ^{15}N Met). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [139296 S_0 – 139296 S_1] scans. Spectrum was processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

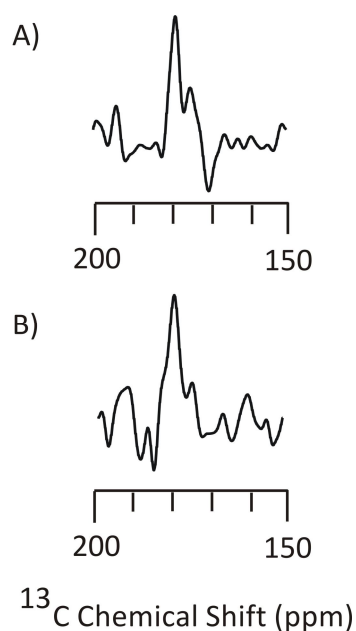


Figure B5: ΔS spectra corresponding to labeling at Phe70 of FHA2. (FHA2 was labeled with $1\text{-}^{13}\text{C}$ Phe and ^{15}N Ser). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [172679 S_0 – 172679 S_1] scans. B) The sample was made in the same way as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [200192 S_0 – 200192 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

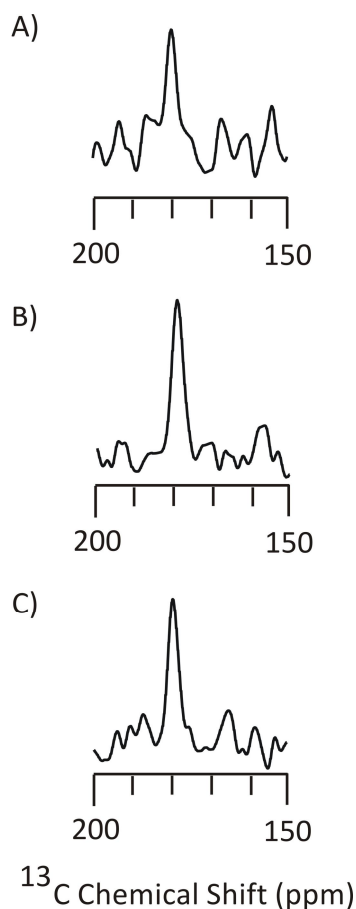


Figure B6: ΔS spectra corresponding to labeling at Leu98 of FHA2. (FHA2 was labeled with $1\text{-}^{13}\text{C}$, ^{15}N Leu). A) Purified FHA2 protein was combined with a lipid film containing a 4:1 POPC:POPG mixture and dialyzed at pH 5.0. ΔS spectrum is [101408 S_0 – 101408 S_1] scans. B) The sample was made in the same way as described in A, but after the initial dialysis at pH 5.0, the sample was then dialyzed at pH 7.4. ΔS spectrum is [111552 S_0 – 111552 S_1] scans. C) The sample was made in the same way as described in A, but the lipid film contained a 8:2:5 mixture of POPC:POPG:chol. ΔS spectrum is [92800 S_0 – 92800 S_1] scans. All spectra are processed with 200 Hz Gaussian line broadening and 5th order baseline correction.

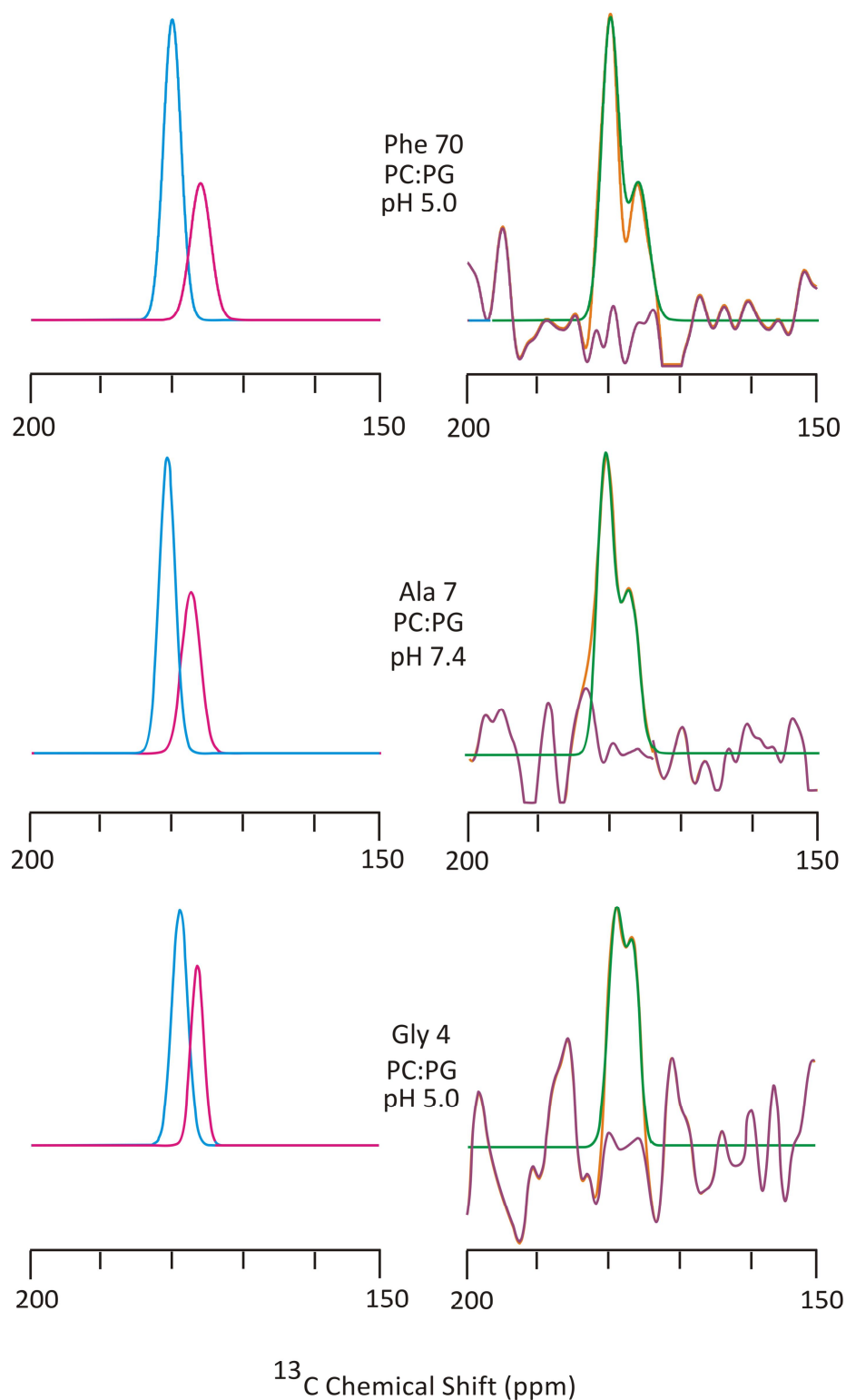


Figure B-7: Deconvolutions of ΔS are displayed for select samples of FHA2 in membranes. The position observed in FHA2 as well as the sample conditions are given in the figure. The fitting of each deconvolution is shown on the right, where orange represents the experimental data, green is the best-fit deconvolution sum, and purple is the difference between the two.

Table B1: Information obtained from analysis of ΔS spectra observing Phe3 of FHA2 in membranes. The ΔS spectra are shown in Figure B1. Deconvolution was not meaningful because the ΔS spectra were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Phe secondary structures which have Gaussian distributions as follows: coil = 175.6 ± 1.6 ppm, helical = 177.3 ± 1.4 ppm, β strand = 174.3 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	179.8	314	103	α
7.4	PC:PG	179.7	196	64	α

Table B2: Information obtained from analysis of ΔS spectra observing Gly4 of FHA2 in membranes. The ΔS spectra are shown in Figure B2. Deconvolution of the pH 5 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the other spectra because they were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	178.7	272	25	α
		176.1	215	15	α
7.4	PC:PG	179.3	454	58	α
5.0	PC:PG:chol	183.0	295	22	α

Table B3: Information obtained from analysis of ΔS spectra observing Ala7 of FHA2 in membranes. The ΔS spectra are shown in Figure B3. Deconvolution of the pH 7.4 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the other spectra because they were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C chemical shifts for different Ala secondary structures which have Gaussian distributions as follows: coil = 177.7 ± 1.6 ppm, helical = 179.4 ± 1.3 ppm, β strand = 176.1 ± 1.5 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	180.6	351	30	α
7.4	PC:PG	180.4	309	31	α
		176.9	285	15	β
5.0	PC:PG:chol	180.6	329	37	α

Table B4: Information obtained from analysis of ΔS spectrum observing Gly16 of FHA2 in membranes. The ΔS spectrum is shown in Figure B4. Deconvolution was not meaningful for the spectrum because it was relatively featureless. The conformation designated is assigned based on characteristic ^{13}C chemical shifts for different Gly secondary structures which have Gaussian distributions as follows: coil = 173.9 ± 1.4 ppm, helical = 175.5 ± 1.2 ppm, β strand = 172.6 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	179.1	425	89	α

Table B5: Information obtained from analysis of ΔS spectra observing Phe70 of FHA2 in membranes. The ΔS spectra are shown in Figure B5. Deconvolution of the pH 5.0 sample was done with two Gaussian lineshapes, whose frequency, width and intensity were independently varied until there was minimal difference between the experimental lineshape and the best fit sum lineshape. Deconvolution was not meaningful for the pH 7.4 sample because the ΔS spectrum was relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Phe secondary structures which have Gaussian distributions as follows: coil = 175.6 ± 1.6 ppm, helical = 177.3 ± 1.4 ppm, β strand = 174.3 ± 1.6 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	179.8	278	30	α
		175.6	283	14	coil
7.4	PC:PG	179.8	318	29	α

Table B6: Information obtained from analysis of ΔS spectra observing Leu98 of FHA2 in membranes. The ΔS spectra are shown in Figure B6. Deconvolution was not meaningful because the ΔS spectra were relatively featureless. The conformations designated are assigned based on characteristic ^{13}C O chemical shifts for different Leu secondary structures which have Gaussian distributions as follows: coil = 176.9 ± 1.7 ppm, helical = 178.5 ± 1.3 ppm, β strand = 175.7 ± 1.5 ppm (8). The peak width reported is the full width at half maximal value.

pH	Lipid film composition	Chemical shift (ppm)	Peak width (Hz)	Integrated signal intensity	Secondary structure
5.0	PC:PG	179.7	298	30	α
7.4	PC:PG	180.2	332	77	α
5.0	PC:PG:chol	180.0	305	54	α

Conclusions

In summary, it is possible that the sample pH (5.0 vs. 7.4) as well as presence/absence of cholesterol within the membranes could affect the secondary structure observed at certain positions in FHA2. Now that the isotopic labeling method has been studied more in depth, and control experiments have been run (as discussed in Chapter 3), this project could be completed.

As we don't fully understand how the scrambling of the isotopic labels have changed what signal we observe in the ΔS spectra, I cannot comment further on interpretations of the spectra.

REFERENCES

REFERENCES

1. Matlin, K. S., Reggio, H., Helenius, A., and Simons, K. (1981) Infectious Entry Pathway Of Influenza-Virus In A Canine Kidney-Cell Line, *J. Cell Biol.* 91, 601-613.
2. White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme, *Crit. Rev. Biochem. Mol. Biol.* 43, 189-219.
3. Curtis-Fisk, J. (2009) Structural studies of the Influenza and HIV viral fusion proteins and bacterial inclusion bodies, Ph. D. Thesis, Michigan State University.
4. Scheiffele, P., Rietveld, A., Wilk, T., and Simons, K. (1999) Influenza viruses select ordered lipid domains during budding from the plasma membrane, *Journal Of Biological Chemistry* 274, 2038-2044.
5. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in E. coli, purification, and refolding of the full ectodomain of the Influenza virus membrane fusion protein, *Prot. Expr. Purif.* 61, 212-219.
6. Gullion, T., and Schaefer, J. (1989) Rotational-echo double-resonance NMR, *J. Magn. Reson.* 81, 196-200.
7. Morcombe, C. R., and Zilm, K. W. (2003) Chemical shift referencing in MAS solid state NMR, *J. Magn. Reson.* 162, 479-486.
8. Zhang, H. Y., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts, *J. Biomol. NMR* 25, 173-195.

APPENDIX C

Locations of NMR Files

Locations of NMR files organized by relevant chapter are shown below. There are additional files in the directory: mb4b/data/Erica/ which are organized by month and year, and details for these NMR files can be found in the corresponding lab notebooks by date. There is also a complete listing of all NMR files in notebook # 5, page 137, 139, and 142 – 146.

NMR Files Grouped by Chapter:

Chapter 2 Figures

Figure 2-6

- a) 011009
- b) 072709
- c) 071909
- d) 010809
- e) 122008
- f) 121808

Chapter 3 Figures

Figure 3-1

- a) 19jun2011
- b) 21jun2011

Figure 3-3, 3-4, 3-5, 3-6

- a) 05012011_redor
- b) 05022011_redor
- c) 5july2011
- d) 2November2011

Figure 3-7, 3-8

17November2011
19November2011
2December2011

Figure 3-9

23May2012
25May2012
27May2012

Figure 3-10

11November2011

15November2011
5November2011

Figure 3-11
1December2011
4December2011
5December2011
11November2011
17November2011
19November2011

Figure 3-12
1December2011
4December2011
5December2011
11November2011
17November2011
19November2011
15November2011
5November2011

Chapter 4 Figures

11November2011
13November2011
15November2011
10November2011
8November2011
12November2011
16November2011
5November2011
7November2011

Appendix A Figures

Figure A-4
1December2011
18November2011

Appendix B Figures

Figure B-1
a) 010609
b) 010709

Figure B-2

- a) 082608
- b) 082908
- c) 090208

Figure B-3

- a) 072408
- b) 072508
- c) 073008

Figure B-4

- a) 100308

Figure B-5

- a) 022608
- b) 030108

Figure B-6

- a) 070908
- b) 071108
- c) 071308