DEVELOPING 2D IR SPECTROSCOPY AS A QUANTITATIVE PROBE FOR PROTEIN STRUCTURE

by

ANN MARIE WOYS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(CHEMISTRY)

at the

UNIVERSITY OF WISCONSIN-MADISON

2012

DATE OF FINAL ORAL EXAMINATION: 12/20/11

THE DISSERTATION IS APPROVED BY THE FOLLOWING MEMBERS OF THE FINAL ORAL COMMITTEE:

MARTIN T. ZANNI, PROFESSOR, CHEMISTRY
JAMES L. SKINNER, PROFESSOR, CHEMISTRY
EDWIN L. SIBERT, PROFESSOR, CHEMISTRY
JOHN C. WRIGHT, PROFESSOR, CHEMISTRY
JOHN F. BERRY, PROFESSOR, CHEMISTRY

PREFACE

DEVELOPING 2D IR SPECTROSCOPY AS A QUANTITATIVE PROBE FOR PROTEIN STRUCTURE

Ann Marie Woys, Ph.D.

University of Wisconsin-Madison, 2011

> "Penetrating so many secrets, we cease to believe in the unknowable. But there it sits nevertheless, calmly licking its chops."
>
> H. L. Mencken, "Minority Report," 1956

The goal of my research was two-fold. First, it was to synthesize and characterize model polypeptide systems to act as benchmarks for the testing and development of frequency-field correlations and coupling models. Towards this end, I studied a membrane polypeptide, ovispirin, which is an α-helix that lies on the surface of lipid bilayers and a synthetic polypeptide macrocycle that is designed to adopt a parallel β-sheet conformation. These studies encompassed two of the most common secondary structures found in native protein folds and span a wide range of electrostatic environments. Second, it was to broaden the applicability of 2D IR spectroscopy for studying large protein systems. For this aim, I synthesized a series of novel non-native infrared probes that can be attached to cysteine-mutated proteins. These probes interrogate protein structure via solvent-dependent frequencies and lifetimes. Moreover, they can be utilized in proteins of arbitrary size and at very low concentrations, thus widening the range of systems that FTIR and 2D IR spectroscopy can be applied to for obtaining important structural information.

Ann Marie Woys, April 23rd, 2012

ACKNOWLEDGEMENTS

My graduate work was not only the result of my own motivation and talents; it was also due to the many people who helped me along the way. Because the list includes such a large number of people, I will be concise and tell you only about those who played critical roles. To start, my advisor, Martin Zanni, was pivotal in my development as a scientist. Marty provided me with interesting and challenging projects and pushed me along with a lot of motivation and semi-daily check-ins. He reminded me of my projects' importance. Along with such encouragement, he pushed me to constantly improve my scientific and presentation skills. I am very grateful to have had Martin Zanni as my advisor.

Surrounding me throughout graduate school was the team that Marty built. I thank all the group members for creating an atmosphere where I could flourish. The mentorship I received from Amber Krummel, Prabuddha Mukherjee, and Terry Ding helped me to get started quickly in lab. Sang-Hee Shim with her intelligence, hard-work, and persistence exemplified whom I aspired to be as a graduate student. Dave Strasfeld took time to explain the laser in great detail. Wei Xiong, who was in my year, was kind and jovial, even when I was teasing him. Our postdocs have also been helpful in my development as a scientist. Chris Middleton taught me the value in thinking through a problem slowly and step-by-step. I enjoyed discussing creative and sometimes crazy scientific ideas with Sean Moran. I worked on many projects with my juniors in the Zanni group, I appreciate the positive attitude each one brought to the lab daily. They all were generous, both emotionally and with their smiles. In lab, they were considerate and contributed to the development and versatility of our group's research, so that new ideas could be tried out efficiently. Throughout my six years here, I am glad that I was a part of the Zanni group dynamic.

Aside from the group, many others in the chemistry department were critical to the success of my projects. Often, my projects required me to perform chemical syntheses that the Zanni group had no experience in or was equipped for, forcing me to rely on the expertise and equipment from many generous professors and students. The list of professors whom I consulted and would like to thank includes Charles Casey, John Berry, Jim Skinner, and Tehshik Yoon. In addition to the helpful conversations I had with John Berry and Jim Skinner, I am also grateful for their service on my defense committee, along with John Wright and Ned Sibert. The professors who loaned me equipment and chemicals include Steven Burke, Sam Gellman, Tehshik Yoon, Clark Landis, Mahesh Mahanthappa, Silvia Cavagnero, Song Jin, and Judith Burstyn. While there are too many graduate students, who ran instrumentation for me and who gave advice, to list, the greatest contributions were from Yu-Shan Lin, Alex Clemens, Lu Wang, Chris Shaffer, Andy Schmidt, Aaron Almeida, and Tamas Benkovics. Also in our chemistry department, there are many staff members who kindly assisted with my projects. These included Charlie Fry, Martha Vestling, and Tracy Drier.

Long before graduate school, I appreciate everything my parents did to prepare me during my early childhood. My mom constantly taught me rudimentary academic skills, no matter whether we were in a dump truck or a foundry. My father did the same, but in a much different way. Since he was not an expert in everything, when I was a child, he often left me with people that were experts in their field. These people ranged from sugar beet factory workers to tress drafters to pilots to the engineers and tradesmen that designed and built assembly-line equipment. I was integrated into their job for just a day or two, seeing their factories and offices and learning how to do their jobs. In addition to opening my world to a diverse array of people and their respective professions, my dad ingrained in me a philosophy that no project was impossible.

During high school and my undergraduate training, I received encouragement and guidance from my chemistry teachers. In high school, Sandy Shaffer accepted me and did not judge my questionable life choices. She saw my academic potential, and encouraged me despite the fact that at that time I definitely did not fit the stereotype of a ‘chemist.’ During my undergraduate education, I received support from my physical chemistry teacher and research advisor, Deborah Huntley. She advised my research and spent hours of her week teaching me physical chemistry, solid state physics, electronic structure calculations, and whatever else I was curious about. She explained what graduate school was and encouraged me to apply. Without the guidance of these two people, I am not sure what I would be doing right now.

In Madison, Rachel, Shu, April, and I entered graduate school together. We helped each other through a lot of rough spots and had some crazy times along the way. Finally, when times were difficult, my family was always there to support me. At any time, I knew calling my grandma would certainly make me happy. She was supportive of my hyperactivity and non-stop questions as a child and even my choice to attend graduate school. My brother Robert is another person who supported me hugely throughout graduate school. He drove to Madison and helped me transport my belongings 4 of the 5 times I moved, no matter what state he was currently living in. Aside from geographic issues, he is and always was my best friend. My sister, Sally, was born when I was 19 years old, and she is the reason I decided to do anything with my life, so I acknowledge her as the reason I accomplished all of the work contained in this thesis.

Ann Marie Woys, April 23rd, 2012

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION TO 2D IR SPECTROSCOPY OF PROTEINS

CHAPTER 2 METHODS FOR LABELING PEPTIDES AND PROTEINS

CHAPTER 3 2D IR LINESHAPES PROBE OVISPIRIN PEPTIDE CONFORMATION AND DEPTH IN LIPID BILAYERS

CHAPTER 4 PARALLEL β-SHEET VIBRATIONAL COUPLINGS REVEALED BY 2D IR SPECTROSCOPY OF AN ISOTOPICALLY LABELED MACROCYCLE: QUANTITATIVE BENCHMARK FOR THE INTERPRETATION OF AMYLOID AND PROTEIN INFRARED SPECTRA.

CHAPTER 5 SYNTHESIS OF CYSTEINE-SELECTIVE CYCLOPENTADIENYL RHENIUM CARBONYL LABELS FOR FTIR AND 2D IR PROBES OF PROTEIN STRUCTURE

CHAPTER 6 A NON-NATURAL INFRARED PROBE OF PROTEIN ELECTROSTATICS AND SOLVATION

APPENDIX 1 SUPPORTING INFORMATION FOR CHAPTER 3

APPENDIX 2 SUPPORTING INFORMATION FOR CHAPTER 4

APPENDIX 3 SUPPORTING INFORMATION FOR CHAPTER 5

APPENDIX 4 SUPPORTING INFORMATION FOR CHAPTER 6

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

CHAPTER 1

Introduction to 2D IR Spectroscopy of Proteins

1.1 Introduction

Many diseases are directly linked to molecular mechanisms in the body. By understanding these molecular mechanisms, we may hope to benefit humankind through the development of therapeutics. More often than not, however, therapeutics are not developed in a rational manner, by understanding the disease, but instead, many substances, often random are screened, hoping for some to work.[1] Sometimes screens work well, but a better way is a bottom-up approach to drug design, where we understand the workings of the disease and build molecules that either impede progress of the disease or prevent its manifestation in the first place.

The molecular mechanisms of disease may be understood by understanding the key components that keep it sustained. Membrane proteins are often good targets for therapeutics, since membrane proteins are life's workhorse. They transport energy and regulatory molecules in and out of cells.[2] They maintain salt concentration in cells and induce potentials across cell membranes.[3, 4] As a result, disrupting a single type of membrane protein is sometimes enough to prevent an illness. For this reason, most drugs target membrane proteins.[5, 6] For example, the drug Rimantadine targets a single membrane protein, the M2 proton channel in the influenza virus, which is all that is necessary to prevent the flu from spreading.[7, 8] The goal is to find suitable target proteins and then understand how they work.

To understand how proteins like these work, we need a way to look at the protein as it performs its duty. There are some good methods for doing something close to this.[9] I say close, because each method is artificial or limited in some way.[9] One of the best techniques for solving protein structure is x-ray crystallography, due to its atomic resolution. However, for x-ray crystallography to give information about a protein's structure, the system has to be crystallized.[10, 11] Many proteins cannot be crystallized or lose their interesting characteristics when crystallized, especially if the protein's function is linked to its dynamics. Due to the importance of the topic, there cannot be enough tools. My dissertation is centered on developing a new method for studying protein structure and kinetics that compliments x-ray crystallography and other protein structure techniques.

As a spectroscopist, my expertise is to use light from lasers to study molecular structure. Some laser wavelengths may damage or change the structure of the protein, especially visible and ultraviolet wavelengths, which cause electronic transitions in proteins.[12] My research utilizes infrared radiation that is more innocuous. When a molecule absorbs infrared energy, it becomes vibrationally excited. Vibrational energy levels are quantized. Each time the molecular vibrational mode in a protein absorbs a photon from an infrared laser pulse, the quantum state is changed. The vibrational quantum states give information about the protein's molecular structure. I specialize in a technique that probes the vibrational energy levels, called 2D IR spectroscopy.[13-15] This technique holds great promise for studying proteins, especially those that are aggregating, membrane bound, or dynamical- exactly those conditions that other techniques are not as well adapted for.

2D IR spectroscopy utilizes a series of very short (~femtosecond) infrared laser pulses to interrogate molecules. Typically, three pulses are used to induce vibrational transitions (see Fig. 1.1). The transition can occur to a level higher than the current state, absorption, or a level lower via simulated emission ($v_n \rightarrow v_{n\pm1}$). The first pulse creates a coherence between two levels (0 and 1), which is because the laser pulse creates a wavefunction that consists of a linear superposition of the v=0 and v=1 states. In the notation of the density matrix, this is referred to as |1><0| or |0><1|, depending on the initial phase of the coherence. The second pulse puts the oscillator in a population, either |1><1| or |0><0|, which I show as the next possibility in Fig. 1.1. The third pulse creates a second coherence state with three possible combinations. We measure the field emitted by the molecule during this final coherence state.

The signal that is emitted is a combination of inherent properties to the oscillator and also a record of the oscillator's interaction with its environment during its vibrational coherence time. The environment interacts with the molecule to change the way it vibrates during this time. For example the molecules surrounding the oscillator may form hydrogen bonds at different orientations or induce a fluctuating electric field. From this record, we can learn about the environment of a vibration in the protein.

The record of the oscillator's environment between the 1st and 2nd pulses (which is that 1st coherence time) as interrogated with the 3rd pulse can be illustrated as a 2-dimensional contour plot (Fig. 1.2). The vertical axis is the frequency that we effectively pump with the first pulse (ω_{pump}), and the horizontal axis is the frequency of the signal we are probing (ω_{probe}). Signal shows up on the diagonal when the three pulses only interact with the ground and first vibrational state. Shifted off the diagonal are the pathways that involve v=2. Fig. 1.2 shows the

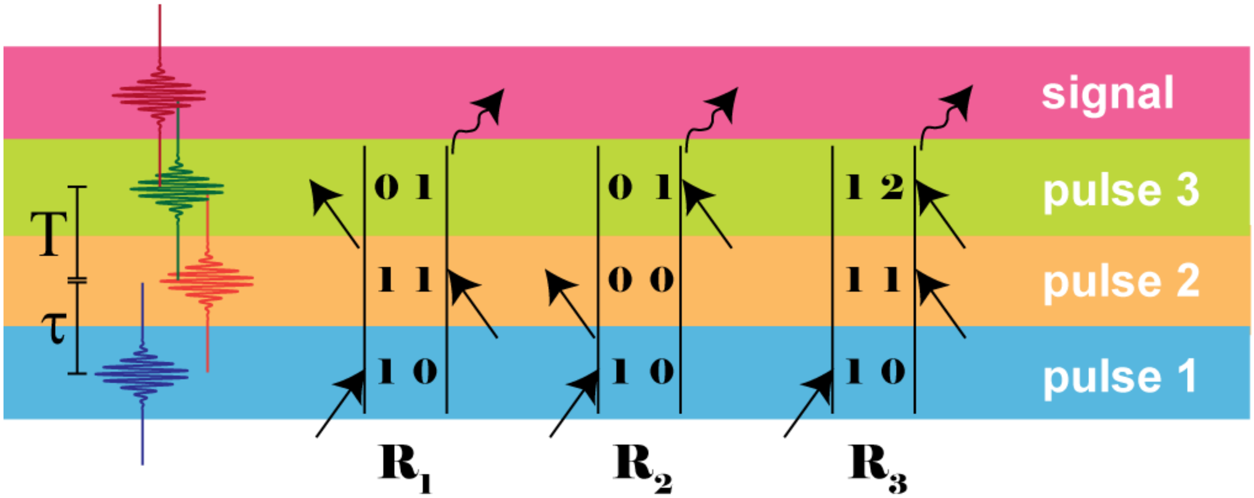


Figure 1.1 - The pulse sequence and three possible vibrational pathways are show in the diagrams (called Feynman diagrams). In most 2D IR spectra, the time between the first and second pulses, τ, is scanned and Fourier transferred to give the y-axis, and the x-axis is the measured signal generated from the free induction decay during the final coherence time, which is often collected in the frequency-domain.

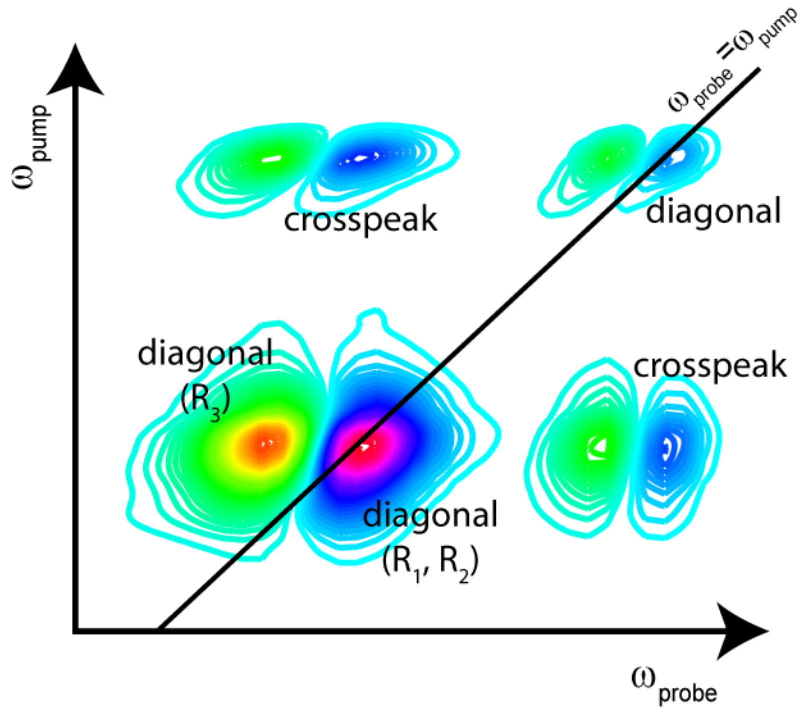


Figure 1.2 - 2D IR spectrum with vibrational pathways that contribute to peaks labeled for the lower frequency diagonal peak.

diagonal peaks labeled with the pathways that contribute to each peak, using the notation from Fig. 1.1.

Crosspeaks also appear in Fig. 1.2. The pathways that give rise to crosspeaks involve vibrational states on both oscillators. Crosspeaks appear when vibrational modes are coupled, which depends on their relative orientation and distance. By measuring the coupling from a 2D IR spectrum, we obtain information about these dependencies, which in turn tells us about the molecular structure. Thus vibrational spectroscopy provides a means to probe protein structure.

To gain structural information from just the diagonal peaks, we need to know how the environment influences the 1st vibrational coherence. The environment can affect the oscillator's frequency (ω) with its electric field.[16] For example, water has a greater electric field than oil on average, due to the partial charges on water. Therefore, the frequency alone tells us about the average electric field at any point from which we can infer something about the exposure to water or charge.[17] Since hydrophobicity is an important factor in protein structure, measuring water exposure is very useful.

The environment is not static; it moves. Thus, the electric field around an oscillator changes with time causing a time dependent frequency of the oscillator (Fig. 1.3a). If frequency fluctuations occur much more rapidly than the timescale of our experiment, the linewidth will appear narrower (called motional narrowing) than the actual distribution of frequencies sampled by the molecule.[18] This timescale is measured in the 2D IR spectrum along the antidiagonal (Fig. 1.3b), by the antidiagonal linewidth which is inversely proportional to the dephasing time (neglecting vibrational relaxation).[13] Fluctuations much slower than the measurement appear static, and the distribution of these frequencies causes inhomogeneous broadening on the diagonal in the 2D IR spectrum.[18] The inhomogeneous lineshape convoluted with the

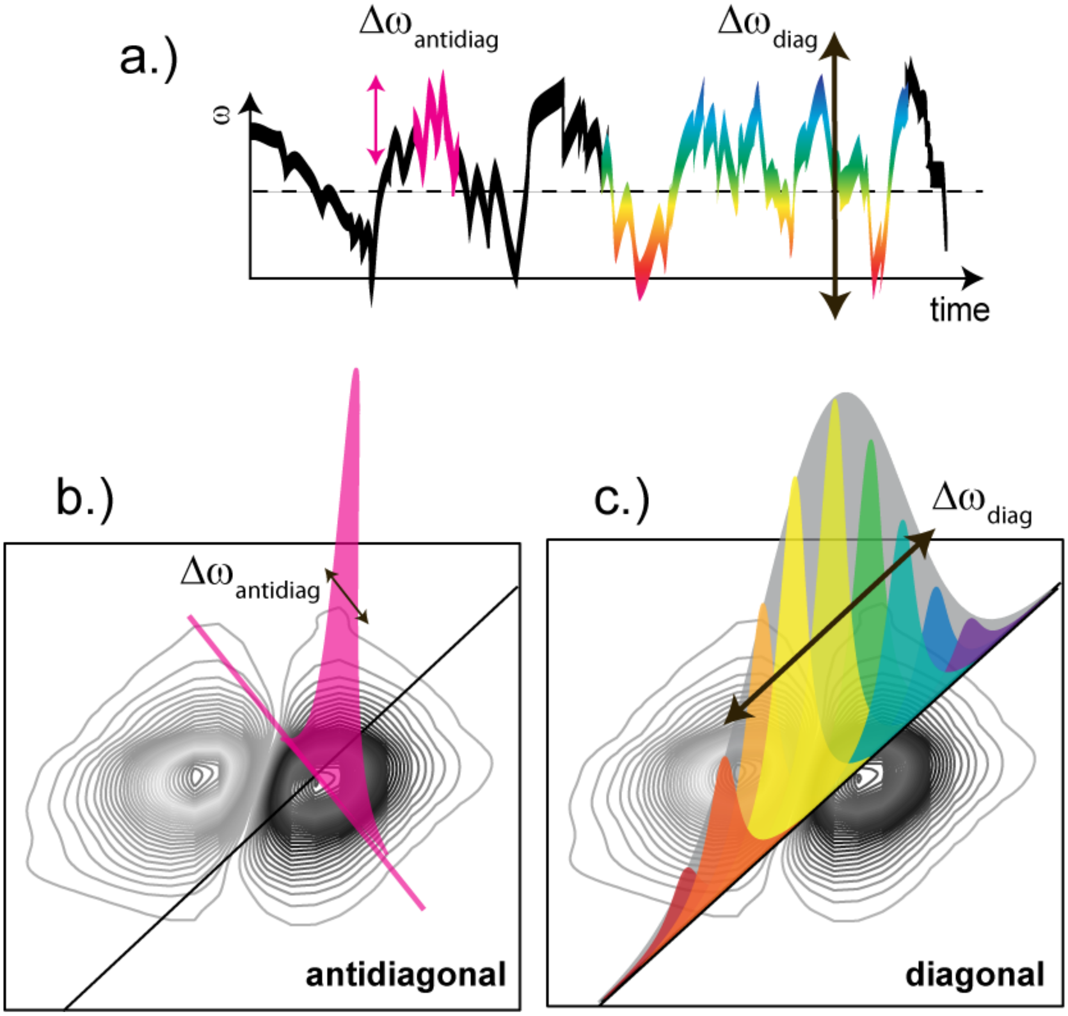


Figure 1.3 - Time trajectory of the frequency (a.) with the frequency fluctuation that are much faster than the timescale of the measurement highlighted in fuchsia and shown along the antidiagonal (b.) and the frequency fluctuations that appear static over the course of the measurement, which broaden the diagonal spectrum (c.).

homogeneous lineshape contribute to the spectrum observed on the diagonal (Fig. 1.3c).[14] From the homogeneous and inhomogeneous linewidth in the 2D IR spectrum, the vibrational dynamics and thus the electric field dynamics are measured. Knowing the electric field fluctuations provide environmental information useful for determining the protein structure.[19]

The information described above is obtained by scanning the time between the first and second pulses, but what happens if we scan the time after the population is created (remember two pulse interactions make a population)? This time is called the waiting time (T, see Fig. 1.1) and so it follows that such experiments are called waiting time experiments. During this time the oscillator has a chance to relax by transferring energy to a different vibration on the molecule or to the solvent. Many factors influence this relaxation, such as the friction exerted by the solvent, which is a combination of the electric field of the solvent and the interaction between the oscillator and the solvent.[20, 21] Energetically accessible modes will also assist relaxation.

To obtain the quantities described above, one would like to know how to make such measurements. In the following sections, I give an overview of how to implement and understand 2D IR experiments on peptides and proteins. First, the setup of the laser is described. Then I discuss how to label peptides and proteins so that individual sites on the protein can be resolved. As each of these techniques for labeling is presented, a brief preview of the data in later chapters is discussed to give an idea of the sort of information that can be gained from each labeling technique.

1.2 Experimental Setup for 2D IR Spectroscopy

To generate the three femtosecond infrared pulses discussed above, we start by pumping a Ti:Sapphire crystal with 527 nm continuous-wave (cw) light from a Nd:YVO_4 laser. The output of the Ti:Sapphire crystal are pulses of 800 nm, 100 femtosecond light. These femtosecond pulses are stretched on a grating and then amplified by combining with 20W, 532 nm, 100 fs pulses from a Nd:YAG laser in a Ti:Sapphire crystal. The emitted pulses are compressed in time with another grating, so that 100 fs pulses are returned. The next step is to convert these 800 nm pulses to mid-IR light. For that we use a two stage optical parametric amplifier (OPA) with a double pass barium borate (BBO) crystal and difference-frequency generation in a $AgGaS_2$ crystal. The result is 4-5 μJ, 100 fs, 5-6 μm light, which we then send into the spectrometer for measuring the 2D IR response of the sample.

For the spectrometer part of the table, the mid-IR pulse is split in three and for that we use mid-IR pulse-shaping.[22] Within this setup, the first two pulses (the pump pulses), are generated from one pulse with the pulse shaper. The pulse shaper starts with about two-thirds of the mid-IR pulse power. The broad band of frequencies contained in the femtosecond pulse ($\Delta E \Delta t \geq \hbar/_2$) is dispersed with a grating onto germanium crystal, so that the frequencies are spread out across ~5 cm. The refractive index of the germanium is modulated with a radio-frequency wave generated from computer input, and as the refractive index changes at different places along the germanium, some mid-IR frequencies are passed and some are diffracted. The diffracted mid-IR light is collimated with a second grating and the output is two pump pulses separated by a delay that can be varied for each pulse.

The two pump pulses are overlapped with the rest of the mid-IR beam, which acts as the probe pulse, on the sample, where the sample response is generated. The signal is dispersed within the monochromator and detected on a strip of 64 Mercury Cadmium Telluride (MCT) pixels. For a 2D IR spectrum, the time between the pump pulses is varied from 0 to 2560 fs in 24 fs steps for the amide I mode work and 0 to 4000 fs in 40 fs steps for the metal carbonyl work. A Fourier transform of this time scan becomes the pump axis on the vertical, and the horizontal axis is what is detected in frequency space on the array detector.

This setup is particularly well-suited to studying membrane and/or aggregating proteins that scatter infrared light, since the phases of the pump pulses are controlled and can be cycled, most spectral scatter is removed.[22, 23] Also, with this implementation, spectra can be collected in much less than a minute. For this reason, the technique is well-suited to study the kinetics of protein structure and dynamics throughout kinetic processes such as aggregation.

1.3 Labeling Peptides

Many vibrational modes in the protein are available for study. The amide I band is particularly attractive for several reasons. To start, the mode is mainly localized to carbonyl stretching on the protein backbone. Also, the oscillator strength is quite good (~500 $M^{-1}cm^{-1}$) relative to other protein vibrational modes.[24] Additionally, due to the strong coupling between amide I modes, the frequency and lineshape are sensitive to the secondary structure of the protein.[25] Finally, and possibly most pertinent to these studies, the lineshape is very sensitive to local electrostatic environment.[26]

To get site-specificity, a particular amide I mode can be isolated without perturbation of protein structure or chemistry with $^{13}C=^{18}O$ isotope labeling of the polypeptide backbone. Such labeling red-shifts the mode ~60 cm^{-1} from the unlabeled band, essentially decoupling it from the other amide I modes and in the case of peptides, removing it from the spectral wing of the unlabeled band.[27] $^{13}C=^{18}O$ isotope labeling may be applied to peptides synthesized with Fmoc protecting chemistry in the solid phase. For this assembly of amino acids, the base-labile Fmoc group inhibits reactivity of the amine in the amino acid, and acid-labile groups protect the side-chain reactivity.[28] The result is well-directed synthesis with each addition of amino acid yielding >99% product.[28]

Peptides isotope labeled with this method require 1-$^{13}C=^{18}O$-labeled amino acid with Fmoc and possibly side chain protecting groups. We found two routes for producing this starting material. A detailed explanation of the labeling procedures is found in Chapter 2. In the first, we start with 1-^{13}C free amino acid, and ^{18}O-label under acidic conditions.[23] From there, the Fmoc group and any other necessary side chains are added. For amino acids with hydrocarbon side chains, only requiring Fmoc protection, labeling efficiency for this method is commonly >96%, and the yield is >85%. The second method is best for amino acids with reactive side chains requiring acid-labile protecting groups. In this method, the oxygen in 1-^{13}C labeled amino acid already bearing the required protecting group is exchanged with N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide (EDC), a reagent that is used for amino acid coupling during peptide synthesis.[29] The reaction conditions are mild enough that protecting groups are not lost, resulting in product with labeling efficiency and yields >98%.

We applied this technique to the 18 residue membrane protein, ovispirin to correlate membrane environment with the vibrational dynamics of the backbone carbonyl.[26] We isotope labeled fifteen samples of ovispirin, each containing a $^{13}C=^{18}O$-labeled backbone carbonyl at a different position. 2D IR spectra were collected for each peptide bound to vesicles. Due to the variance in electrostatic environments across the membrane normal, the lineshape varies with the depth of each carbonyl in the bilayer. With this information, we determined whether the carbonyl was surrounded by the hydrocarbon lipid tails, was partially hydrated, near charged lipid headgroups, or surrounded with buffer.

1.4 Assisting in the Development of 2D IR Calculations

Many factors influence the 2D IR spectrum, making the data rich with information and sometimes difficult to interpret. The 2D IR data is sensitive to the inherent characteristics of the oscillator, the angle at which other atoms are bonded to it, the solvent and charges surrounding it, and coupling to other vibrations in the system. To obtain a detailed molecular picture of the system, we start with a molecular dynamics (MD) simulation. The atomic positions and charges are computed as a function of time, from which we can compute the electric field over time. For calculation of the amide I band, the frequency maps developed by Skinner and coworkers correlate the frequencies with electric fields and provide~~s~~ a way to generate local frequencies directly from MD simulations.[30, 31] Other researchers have developed methods as well.[32, 33] Mixed quantum/classical lineshape theory is used to calculate a 2D IR spectrum.

For example, MD simulations were performed for ovispirin on a lipid bilayer.[26] With the resulting trajectory, 2D IR spectra were calculated and compared to experiment. The comparison enabled us to gain an atomic understanding of the peptide structure in the membrane environment. From calculations and experiments of individual amide I modes, we learned about how the membrane environment affects the 2D IR lineshape and details of the peptide's structure in the membrane.

When calculating 2D IR spectra for many amide I modes on the protein, coupling between those vibrations becomes very important. A commonly used coupling model is the transition dipole coupling (TDC) method.[34, 35] In the TDC model, coupling constants are dependent on the distance and relative orientation between oscillators, and the coupling parameters for different TDC models are often obtained from ab initio calculations. To test existing coupling models, double $^{13}C{=}^{18}O$ isotope labels isolate the pair of amide I vibrations and provide a direct measurement of coupling. Experiments like this have been carried out for α–helices and anti-parallel β–sheets, since small peptides with these secondary structures occur in nature and are easy to synthesize.[36, 37] In our department, the Gellman group recently developed a method to synthesize a macrocyclic peptide with parallel β–sheet structure.[38] In collaboration with their group, we measured the 2D IR spectra of doubly $^{13}C{=}^{18}O$ isotope labeled β–sheet macrocycles. Experiments like this will help guide the development of calculations for proteins, so that the underlying cause of fine detail in the 2D IR experimental spectrum can be understood and utilized for determining protein structure.

1.5 Comparison of 1D and 2D IR Spectroscopy

2D IR spectroscopy is the modern equivalent of FTIR spectroscopy. FTIR and 2D IR spectroscopies are similar in that both the FTIR linewidth and the 2D IR diagonal linewidth result from inhomogeneous and homogenous broadening. From this measurement alone, a comparison of the inhomogeneous contribution for several $^{13}C{=}^{18}O$ labeled samples is not possible with only the FTIR and 2D IR diagonal spectra, since the homogeneous linewidth may not be constant. With the 2D IR spectrum, however, the homogeneous broadening is directly measured from the antidiagonal linewidth. The antidiagonal linewidth also allows one to measure spectral diffusion when the waiting time is scanned.

Another useful property of 2D IR spectroscopy is that it suppresses background.[39] For FTIR spectroscopy, the intensity of the peaks relative to their linewidth is less than for 2D IR diagonal spectrum (Fig. 1.4a). The difference is how the intensities scale with the transition dipole of the oscillator. For FTIR spectroscopy, intensity goes with the square of the transition dipole moment, but for 2D IR spectroscopy, the transition dipole moment is to the 4th power. The effect is that the background peaks that typically have broad absorption and weak oscillator strength are suppressed, as compared to narrow and strongly absorbing peaks. As a result, the 2D IR isotope labeled peak is much more prominent (Fig. 1.4b), so that the frequency and linewidth measurements are more robust. We have shown that labeled peaks in 2D IR spectra provide reliable measurement of linewidth, especially for samples with interference from other IR transitions (such as those with higher frequency $^{13}C{=}^{18}O$ peaks or with many more residues, which increase interference with the unlabeled band wings).

1.6 Labeling Proteins

Even though 2D IR spectroscopy is good for gathering quantitative measurements of the vibrational dynamics in $^{13}C{=}^{18}O$-labeled peptides, some problems arise when applied to much larger proteins. Amino acid side chains absorb at the same frequency of the isotope.[24] For larger systems, a label with greater molar absorbance and/or for the absorption to occur at a less cluttered region of the mid-IR spectrum is preferable. Since many proteins are produced by biological expression, a label selective to react with a specific side chain is also helpful.

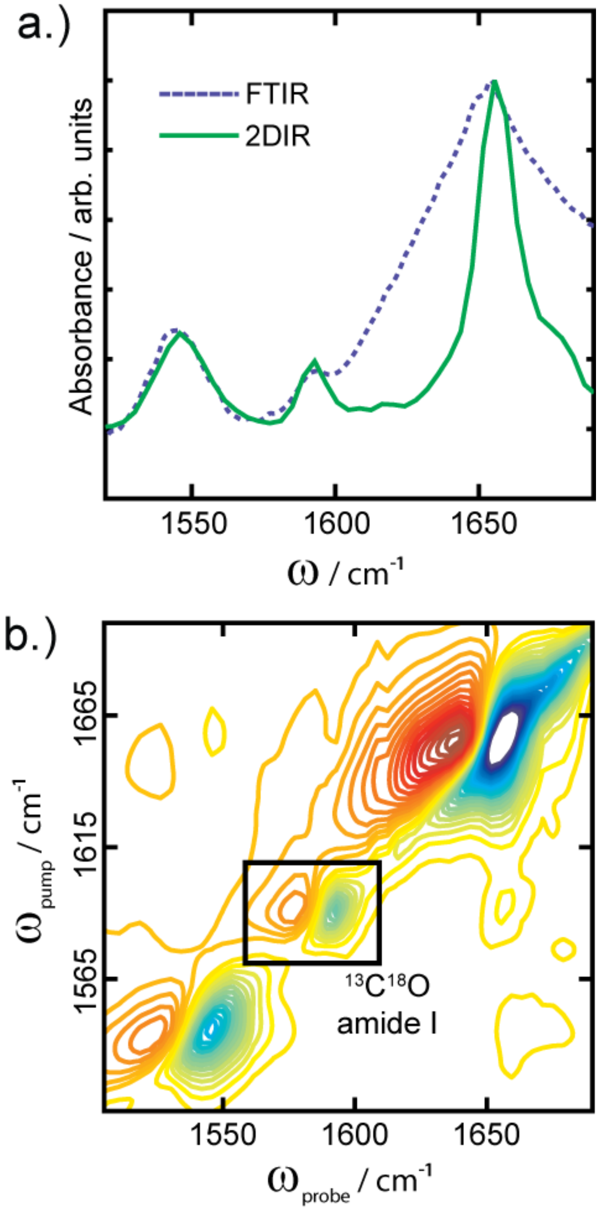


Figure 1.4 - 2D IR diagonal and FTIR spectrum overlaid for a $^{13}C=^{18}O$ labeled peptide (a), and the full 2D IR spectrum to show the resolution of the isotope label lineshape (b). The FTIR spectrum appears broader than the 2D IR diagonal spectrum, because intensity scales with the transition dipole moment differently for each technique.

We designed a non-natural label made from a metal carbonyl that would be strong absorbing, isolated from other protein modes, and amenable to recombinant protein synthesis. For a metal carbonyl stable to protein sample conditions that could be modified for attachment to the protein, we used tricarbonyl(η^5-cyclopentadienyl)rhenium(I). $CpRe(CO)_3$ would be the platform from which Cp-ring substitutions could provide reactivity to the protein.

Some functional groups have already been developed to react only with certain amino acid side chains. For example, N-hydroxysuccinimide esters, imidoesters, pyrylium ions, and Fischer-type metallocarbene complexes label lysine residues, but these are not ideal since lysine occurs frequently in protein sequences.[40] In contrast, the cysteine residue is far less frequent, and there are well established functional groups that are selectively reactive with thiols. One such group is maleimide.[41] We decided not to use maleimide, because the resulting linkage between the label and the protein is rigid and bulky, since it contains a five-membered ring. Ideally, the linkage between the label and the protein should be as small as possible. Another cysteine-specific option, commonly used for site-directed spin-labeling (SDSL) in EPR spectroscopy, is the methanethiosulfonate functional group.[42, 43] For labeling with the methanethiosulfonate, mild conditions (25°C, pH 7 buffer) in 30 min result in a disulfide bond between the label and protein.

To bridge $CpRe(CO)_3$ to methanethiosulfonate, the cyclopentadienyl ring may be activated for substitution with n-BuLi. Alkylation occurs at this step, and then, after a series of substitutions, three labels, $[\eta^5\text{-}(C_5H_4)(CH_2)_nSSO_2CH_3]Re(CO)_3$, where n=1,3, or 4, are synthesized. Details of the syntheses are given in Chapter 4.

As an aside, during this time, we also decided to use click chemistry for attaching an alkyne-containing label to an azide-containing protein, since proteins have recently been expressed with a non-natural azido amino acid via methionine. Also, Stille and coworkers reported their synthesis of alkyne-substituted $CpRe(CO)_3$,[44] which we were able to reproduce. Work is in progress for an azido-modified CTL9 protein to be labeled with $[\eta^5\text{-}(C_5H_4)CCH]Re(CO)_3$, and for the effect of pH-modulated structural changes to be studying with 2D IR spectroscopy.

To utilize this probe, we needed to link the vibrational spectroscopy to the environment. The first model system was two solvents with very different electrostatic properties, 1:1 CH_2Cl_2/CCl_4 and DMSO. The 1:1 CH_2Cl_2/CCl_4 solvent has a dielectric of nearly zero and was intended to simulate a hydrophobic region in a protein or the lipid tail region of a membrane. In contrast, DMSO has dielectric of 3.6 Debye (for reference, ε_{water}=1.8 Debye). Between these two solvents, we found that the label's frequency and lifetime (ranging 15 cm^{-1} and 15 ps, respectively) are very sensitive to the solvent and are correlated. For this label, we observed only a minimal effect on the homogeneous and inhomogeneous linewidths (<0.03 cm^{-1} and <2.5 cm^{-1} variation, respectively). Based on these results, we expect that the frequency and lifetime will provide the most information about the structural dynamics of proteins.

For the second model system, we chose the well-studied and easy-to-handle model protein, ubiquitin. The two mutants, K6C/D77 and K63C, were labeled with $[\eta^5\text{-}(C_5H_4)(CH_2)_nSSO_2CH_3]Re(CO)_3$, where n=1, 3, or 4. The correlation between frequency and lifetime observed for the labels in solvents is not present for the labels attached to ubiquitin. We speculate that when the labels are attached to the protein, more energy pathways are available for relaxation even if the label is somewhat less solvated. All in all, we observe that a plentitude of information is available by performing 2D IR experiments on these labels attached to proteins.

In the upcoming chapters, I will explain how to synthesize these labels and demonstrate how they might be used in peptide or protein structural studies. A detailed description of the synthesis of the labels is given in Chapter 2. Results from $^{13}C=^{18}O$ labeling of a peptide to find its structure in a membrane will be given in Chapter 3. Experiments to measure coupling in a model parallel β–sheet with double $^{13}C=^{18}O$ labels will be presented in Chapter 4. Chapters 5 and 6 are concerned with the development of a metallocarbonyl vibrational label for proteins.

1.7 References

(1) Potyrailo, R.; Rajan, K.; Stoewe, K.; Takeuchi, I.; Chisholm, B.; Lam, H. *ACS Combinatorial Science* **2011**, *13*, 579.

(2) Borst, P.; Elferink, R. O. *Annu. Rev. Biochem.* **2002**, *71*, 537.

(3) Ernst, O. P.; Sakmar, T. P. *Nature (London, U. K.)* **2012**, *482*, 318.

(4) Gonzalez, C.; Contreras, G. F.; Peyser, A.; Larsson, P.; Neely, A.; Latorre, R. *Biophys. Rev.* **2012**, *4*, 1.

(5) Yildirim, M. A.; Goh, K.-I.; Cusick, M. E.; Barabasi, A.-L.; Vidal, M. *Nat Biotech* **2007**, *25*, 1119.

(6) Hubbard, R. E. *Neuropharmacology* **2011**, *60*, 7.

(7) Schnell, J. R.; Chou, J. J. *Nature (London, U. K.)* **2008**, *451*, 591.

(8) Stouffer, A. L.; Acharya, R.; Salom, D.; Levine, A. S.; Di, C. L.; Soto, C. S.; Tereshko, V.; Nanda, V.; Stayrook, S.; De, G. W. F. *Nature (London, U. K.)* **2008**, *451*, 596.

(9) Bill, R. M.; Henderson, P. J. F.; Iwata, S.; Kunji, E. R. S.; Michel, H.; Neutze, R.; Newstead, S.; Poolman, B.; Tate, C. G.; Vogel, H. *Nature Biotechnology* **2011**, *29*, 335.

(10) Fromme, P.; Spence, J. C. H. *Curr. Opin. Struct. Biol.* **2011**, *21*, 509.

(11) Ubarretxena-Belandia, I.; Stokes, D. L. *Adv. Protein Chem. Struct. Biol.* **2010**, *81*, 33.

(12) Hunt, N. T. *Chemical Society Reviews* **2009**, *38*, 1837.

(13) Hamm, P.; Zanni, M. *Concepts and Methods of 2d Infrared Spectroscopy*; Cambridge University Press, 2011.

(14) Remorino, A.; Hochstrasser, R. M. *Acc. Chem. Res.*, Ahead of Print.

(15) Thielges, M. C.; Fayer, M. D. *Acc. Chem. Res.*, Ahead of Print.

(16) Lee, H.; Lee, G.; Jeon, J.; Cho, M. *J. Phys. Chem. A* **2012**, *116*, 347.

(17) Suydam, I. T.; Boxer, S. G. *Biochemistry* **2003**, *42*, 12050.

(18) Schmidt, J. R.; Sundlass, N.; Skinner, J. L. *Chem. Phys. Lett.* **2003**, *378*, 559.

(19) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2009**, *113*, 592.

(20) Cho, M. *J. Chem. Phys.* **1996**, *105*, 10755.

(21) Ladanyi, B. M.; Stratt, R. M. *J. Chem. Phys.* **1999**, *111*, 2008.

(22) Shim, S. H.; Zanni, M. T. *Physical chemistry chemical physics : PCCP* **2009**, *11*, 748.

(23) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* **2010**, *52*, 12.

(24) Barth, A. *Prog Biophys Mol Biol* **2000**, *74*, 141.

(25) Schweitzer-Stenner, R. *The Journal of Physical Chemistry B* **2004**, *108*, 16965.

(26) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de, P. J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832.

(27) Torres, J.; Adams, P. D.; Arkin, I. T. *J. Mol. Biol.* **2000**, *300*, 677.

(28) Chan, W. C.; White, P. D.; Editors *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford Univ Press, 2000.

(29) Seyfried, M. S.; Lauber, B. S.; Luedtke, N. W. *Org Lett* **2009**, *12*, 104.

(30) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2011**, *115*, 3713.

(31) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *The journal of physical chemistry. B* **2009**, *113*, 592.

(32) Hayashi, T.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. A* **2005**, *109*, 9747.

(33) la, C. J. T.; Knoester, J. *J. Chem. Phys.* **2006**, *124*, 044502/1.

(34) Moore, W. H.; Krimm, S. *Proceedings of the National Academy of Sciences of the United States of America* **1975**, *72*, 4933.

(35) Torii, H.; Tasumi, M. *Journal of Raman Spectroscopy* **1998**, *29*, 81.

(36) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R. A. G. D.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.

(37) Setnicka, V.; Huang, R.; Thomas, C. L.; Etienne, M. A.; Kubelka, J.; Hammer, R. P.; Keiderling, T. A. *J. Am. Chem. Soc.* **2005**, *127*, 4992.

(38) Freire, F.; Gellman, S. H. *J. Am. Chem. Soc.* **2009**, *131*, 7970.

(39) Kim, Y. S.; Hochstrasser, R. M. *The Journal of Physical Chemistry B* **2009**, *113*, 8231.

(40) Salmain, M.; Jaouen, G. *Comptes Rendus Chimie* **2003**, *6*, 249.

(41) Haquette, P.; Salmain, M.; Svedlung, K.; Martel, A.; Rudolf, B.; Zakrzewski, J.; Cordier, S.; Roisnel, T.; Fosse, C.; Jaouen, G. *ChemBioChem* **2007**, *8*, 224.

(42) Berliner, L. J.; Grunwald, J.; Hankovszky, H. O.; Hideg, K. *Anal. Biochem.* **1982**, *119*, 450.

(43) Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 19666.

(44) Lo Sterzo, C.; Stille, J. K. *Organometallics* **1990**, *9*, 687.

CHAPTER 2

Methods for Labeling Peptides and Proteins[*]

2.1 Introduction

In this chapter, I review the experimental procedures for labeling peptides and proteins. Since labeling peptides and proteins is not a common skill for many spectroscopists, these procedures are written with the physical chemist in mind. With detail, I will explain two methods for labeling peptides, and two methods for labeling proteins.

The first method I describe for labeling peptides is the most cost-effective for labeling specific residues with non-reactive side chains and starts from 1-^{13}C labeled free amino acid. The second method for labeling amino acids works well for labeling residues with reactive side chains, which require protecting groups on the side chains during synthesis. Both of these methods provide material ready for peptide synthesis via Fmoc-chemistry.

On the other hand, proteins are often labeled post-recombinant synthesis at a specific residue. One of the methods targets the cysteine thiol side chain, whose synthesis is described in Chapter 5. While Chapter 5 would be sufficient for an experienced inorganic chemist to reproduce, I would like to provide useful information here for a physical chemist attempting to reproduce that synthesis. Therefore, in this chapter, I will explain some basic skills used to successfully synthesize $CpRe(CO)_3$ derivatives. The other method for labeling proteins utilizes

---

[*] Some portions of this chapter were originally published in *Methods* **2010**, *52*, 12-22, in collaboration with Chris T. Middleton, Sudipta S. Mukherjee, and Martin T. Zanni and also in *Org. Lett.* **2010**, *12*, 4848-4851, with Peter Marek, Kelvin Sutton, Martin T. Zanni, Daniel P. Raleigh.

azido-alkyne click-chemistry to attach the labels to the proteins. The procedures for the alkyne label synthesis will be discussed in this chapter, as well as azido-modified protein labeling.

2.2 $^{13}C^{18}O$-Labeling of Peptides: A Procedure for Amino Acids with Non-Reactive Side Chains

Due to common use for NMR spectroscopy, most ^{13}C-labeled amino acids are commercially available but only induce a 40 cm^{-1} in the amide-I frequency. A 40 cm^{-1} shift is not always sufficient to resolve the isotope labeled residue, especially in peptides with large β-sheet character. This is also problematic for large peptides because the presence of natural abundance (1.1%) ^{13}C atoms becomes significant. Arkin and coworkers pioneered the use of $^{13}C{=}^{18}O$ labels which induces a 60 cm^{-1} shift and largely alleviates these problems.[1,2] For example, a 300 residue protein will naturally have, on average, three $^{13}C{=}^{16}O$ backbone amide carbonyls but a <1% probability of having a single natural $^{13}C{=}^{18}O$ carbonyl.

To site-specifically label a backbone carbonyl group in a peptide, the labels are incorporated very early in the synthesis. Free amino acids with ^{13}C at the backbone carbonyl position are purchased from Cambridge Isotope Laboratories, Inc. The two carboxylic oxygen atoms are exchanged to ^{18}O under refluxing acidic conditions similar to the method of Torres, et al.:[1]

1) The apparatus is comprised of a reaction vessel and an HCl gas generator, as shown in Fig. 2.1.
2) All glassware is flame-dried under vacuum and washed with argon three times.
3) ^{13}C-labeled amino acid (10 mmol) is added to the reaction vessel and dried with flame under vacuum. Steps 4–6 are performed under Argon atmosphere.

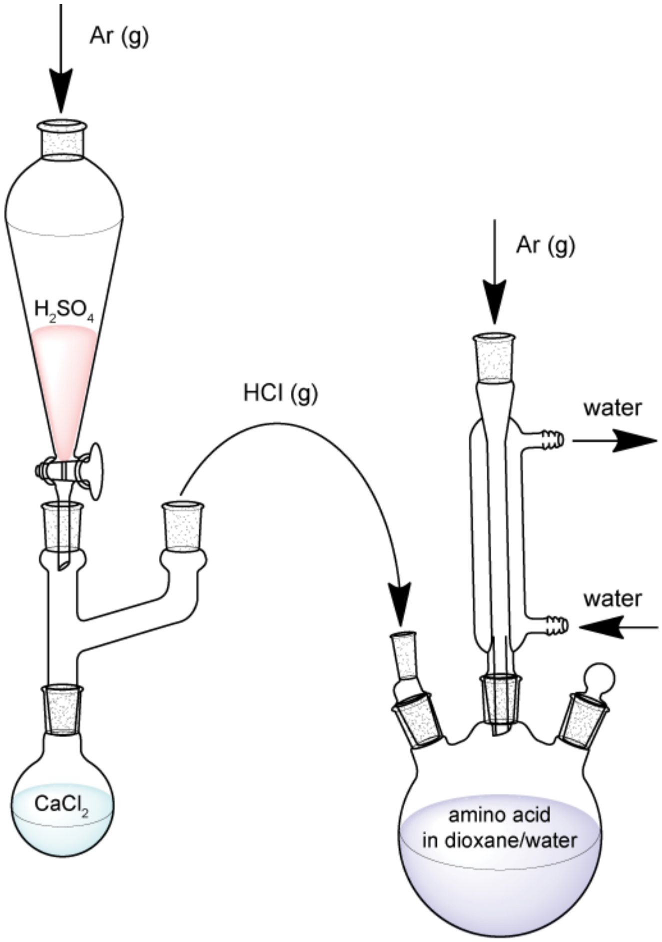


Figure 2.1 - Apparatus for acid-catalyzed ^{18}O-exchange.

4) ^{18}O-enriched water (95% ^{18}O, 1 mL) and 1,4-dioxane (10 mL) are added to the amino acid. Hydrophobic amino acids may require up to 20 mL of 1,4-dioxane. The resulting suspension is stirred and cooled with an ice bath.
5) ^{18}O-exchange occurs at mildly acidic conditions and room temperature; however, to reduce the time required for efficient ^{18}O uptake, the reaction is performed at low pH and 100^{o}C. Dropwise addition of concentrated anhydrous sulfuric acid onto anhydrous calcium chloride produces dry HCl gas, which flows into the dioxane/water suspension via tubing attached to a hollow glass rod. The amino acid dissolves more readily as the suspension pH decreases. Bubbling of HCl is performed for approx. 20 min until the pH of the suspension is 1–2.[1]
6) The HCl port is closed and the solution is refluxed with stirring for 3–6 h. Exchange time depends on steric hinderance caused by the amino acid side chain.[3]
7) The suspension is brought to room temperature and neutralized with 1 M NaOH (aq) while stirring.
8) The dioxane and water are removed under vacuum and collected with a small liquid nitrogen trap.
9) The enrichment is determined by mass spectrometry and the above procedure reiterated with fresh ^{18}O-enriched water until the desired enrichment is achieved.

At this point, protecting groups must be added for Fmoc solid-phase peptide synthesis. Amino acids with nonreactive side chains, specifically glycine (G), alanine (A), leucine (L), isoleucine (I), valine (V) and phenylalanine (F), only require Fmoc protection,[4] and we use the one-pot procedure reported by Cruz et al. with yields of 81–96%.[5] Alternatively, the Fmoc protecting group can be added first, followed by ^{18}O exchange for 3–10 h (longer for β-branched amino acids).[3] Most amino acids require both Fmoc and side-chain protection groups for peptide synthesis. Since the Fmoc group is base labile while the side-chain protection groups are acid labile, orthogonal protection requires multiple steps and is generally impractical. For this reason, isotope labeling has been almost exclusively limited to amino acids with nonreactive side chains (listed above). The lone exception is lysine, which has been successfully $^{13}C=^{18}O$-labeled and incorporated into a peptide.[6] In this case the side-chain protecting group was added through a multi-step synthesis via a copper intermediate as described by Wiejak.[7]

2.3 $^{13}C^{18}O$-Labeling of Peptides: A Procedure for Amino Acids with Reactive Side Chains

For amino acids with reactive side chain groups, an acid-labile group is necessary for controlling reactivity in addition to the base-labile Fmoc group added to the amino acid amine. While the addition of these orthogonal groups is performed routinely by chemical suppliers, duplication of these reactions is not a trivial endeavor. Although I found some success with orthogonally protecting lysine with Boc and Fmoc via a copper intermediate,[7] I was unable to add the Pbf or the Pmc group to arginine. Due to the difficulty of these reactions, we prefer a different route.

To obtain ^{18}O-exchanged amino acids bearing the appropriate protecting groups for peptide synthesis, I utilize the method by Seyfried et al.[8] In this reaction, the repetitive coupling of EDC with the carboxylic acid in conjunction with the heavy isotope effect replaces ^{16}O with ^{18}O to >98%. The conditions are mild enough that no protecting groups are lost. The greatest disadvantage to this method is that obtaining 1-^{13}C amino acids with protecting groups is cost prohibitive. Some amino acids are not sold without custom synthesis. However, for the amino acids available and for residues whose labeling will provide important structural/dynamical information, this method is simple and yields nearly quantitative product with >98% ^{18}O incorporation.

To provide an example of the use of this reaction, 1-^{13}C Fmoc-*O*-*tert*-butyl-L-serine was ^{18}O isotope labeled as outlined in the literature.[8] Isotope labeled chemicals were obtained from Cambridge Isotope Laboratories, Inc. Acetonitrile was distilled from CaH_2 and used shortly thereafter. All other materials were used without further purification. Glassware and stirbars were dried either by flame or in an oven. The isotope labeling reaction was performed under argon atmosphere. NMR data was collected with a Bruker AC 300 MHz spectrometer, which is supported by NSF (CHE 9208463) and NIH (1 S10 RR0 8389-01). ^{1}H NMR spectra were referenced with trimethylsilane. Efficiency of ^{18}O isotope labeling was determined by liquid chromatography-mass spectrometer (LC-MS) with a Shimadzu LCMS-2010A employing for the stationary phase, a Supelco (Bellefonte, PA) 15 cm x 2.1 mm C-18 wide-pore reverse-phase column, and mobile phase, 0.1% formic acid in water and 0.1% formic acid in acetonitrile. Eluents were mass analyzed by electrospray ionization and with a single quadrupole analyzer. The instrument was purchased with funds from a Keck grant. Both the NMR and LC-MS instruments are supported by the University of Wisconsin Chemistry Department.

Oxygen-18 isotope exchange of Fmoc-*O*-*tert*-butyl-L-serine (1-^{13}C, 99%) is briefly summarized. 3,5-dimethylpyridine hydrobromide was prepared by dripping 3,5-lutidine into 2 M hydrobromic acid in 1:1 molar ratio. The solvent was evaporated. The resulting solid, along with the Fmoc-*O*-*tert*-butyl-L-serine and *N*-Ethyl-*N′*-(3-dimethylaminopropyl)carbodiimide (EDC), was dried with three cycles of adding anhydrous acetonitrile, sonicating the suspension, and removing the solvent. The labeling reaction was performed at the 250 mg scale twice.

3,5-dimethylpyridine hydrobromide (2.46 g, 20 eq.) was dissolved into anhydrous *N*,*N*-dimethylformamide (DMF, 6 mL). EDC (1.25 g) was added, and then 750 μL of water (^{18}O, 97%). Fmoc-*O*-*tert*-butyl-L-serine (250 mg) in DMF (8 mL) was added by syringe, which was washed with 2-3 mL of DMF. The reaction was stirred for 18 hours at room temperature. Another addition of EDC (1.25 g) was followed by stirring for another 8 hours. More EDC (1.25 g) was then added. After 15 hours, the reaction mixture was diluted in ethyl acetate (90 mL) and washed with 0.1 M citric acid (75 mL) three times and once with 0.1 M citric acid in brine (75 mL). The aqueous layers were back-extracted with ethyl acetate (90 mL). The organic phase was dried over $MgSO_4$(s) and filtered. After evaporating the solvent, the product was further purified by HPLC. Reverse phase prep scale purification was performed on a Jasco HPLC (Quaternary Gradient Pump) with C18 Vydac column (218TP1022) stationary phase and 0.05% trifluoroacetic acid in water / 0.05% trifluoroacetic acid in acetonitrile gradient mobile phase. More than ten runs were required for purification of the 960 mg crude Fmoc-*O*-*tert*-butyl-L-serine obtained from the reaction. NMR and LC-MS data were collected before and after HPLC purification (Fig. 2.2). The isotope peaks for the sodium adduct (Fig. 2.2e and 2.2j) in the product elutents were used to determine the isotope label uptake. The isotope labeling efficiency was calculated for the coupled amino acid assuming 50/50 carboxylic oxygen preference for the singly ^{18}O-labeled amino acid during peptide bond formation. Isotope enrichment was 93%, and no back-exchange was observed after HPLC purification.

Since purification by HPLC is costly and time consuming, we also explored product purification by gradient flash chromatography as well as an additional extraction. Due to the carboxylic acid moiety in Fmoc-L-alanine, all solvents contained 0.5% acetic acid. Isocratic solvent conditions were first optimized by thin-layer chromatography (TLC). The R_f value was

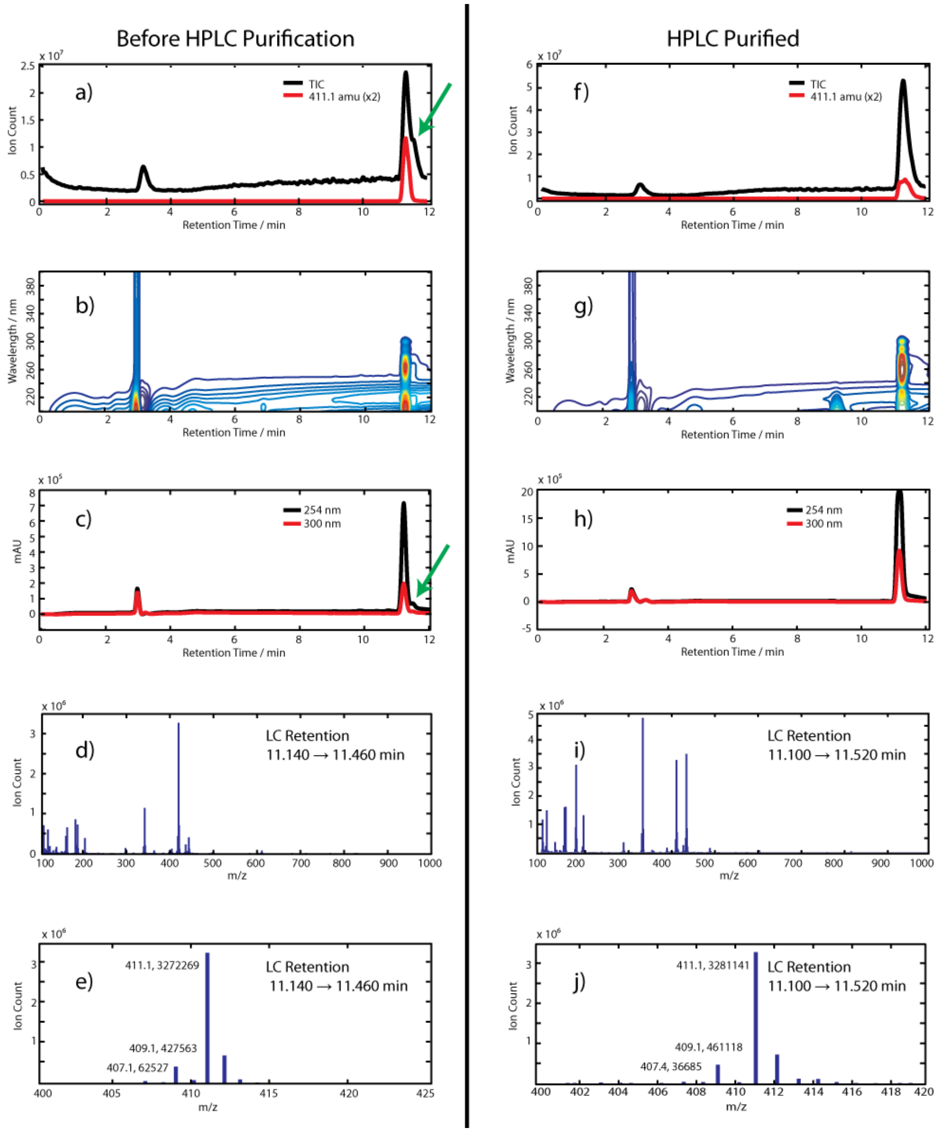


Figure 2.2 - Crude (a-e) and HPLC purified (f-j) product from the ^{18}O labeling reaction of 1-^{13}C Fmoc-*O*-*tert*-butyl-L-serine was analyzed by LC-MS (gradient of 5:95 to 95:5 acetonitrile:water with 95:5 acetonitrile:water wash (continuous 0.1% formic acid). a) and f) The product peak is the major peak in the Total Ion Count (TIC) and 411.1 m/z. The shoulder (green arrow) is a byproduct removed by HPLC purification. b) and g) UV-Vis for the elution is tracked, and c) and h) the product and byproduct absorb at 254 and 300 nm. d) and i) The integrated mass spectrum of the product has most intense mass at 411.1 m/z (product-sodium adduct). e) and j) The isotope pattern for the 411.1 m/z was used to calculate isotope-labeling efficiency.

adjusted to optimize separation of Fmoc-L-alanine and other impurities observed with TLC. A program using the TLC spot position for Fmoc-L-alanine (R_f=0.13) and the nearest impurity (R_f=0.26) as well as the mass of crude product assisted in the choosing the isocratic solvent composition of 3% MeOH in dichloromethane (DCM).[9] Crude mass determined the column dimensions[10] and amount of silica gel.[9] A brief description of the column preparation follows. A small piece of glass wool was placed at the bottom of a 4.4 cm diameter column and covered with enough sand to reach the diameter of the column. 109.1 g of 40-63 μm silica gel was loaded as a slurry using dichloromethane (DCM) with 0.5% acetic acid as the initial solvent. Sand was added, and the column was washed with DCM. A gradient program was designed as outlined by Stevens and Hill.[11] Once the crude material was dissolved in DCM and loaded onto the column, methanol composition was incrementally adjusted from 0 to 6% over ten column volumes. Eluted product was analyzed by LC-MS and NMR (Fig. 2.3).

However, not all impurities were removed. A shoulder appeared on the product peak in the 300 nm chromatogram (Fig. 2.3a-f). For the serine-labeling reaction, this impurity was also present, but eluted farther from the product and therefore was separated by HPLC and could be separated by flash chromatography. However, elution of the alanine-labeling reaction product overlapped with its impurity. In order to purify the Fmoc-L-alanine product, the LC-MS data was more closely analyzed, resulting in a proposed structure for the impurity (Fig. 2.3k) that contains an additional amide rather than the carboxylic acid group in Fmoc-L-alanine. Using this hypothesis, we performed a basic extraction as an alternative to HPLC or flash chromatography. The following workup is similar to that used by Cruz for Fmoc-L-alanine.[12] Product, obtained by performing the labeling reaction with Fmoc-L-Ala-OH, was dissolved in Na_2CO_3(aq), so that the pH of solution was 8-10. The solution was extracted with *tert*-butyl

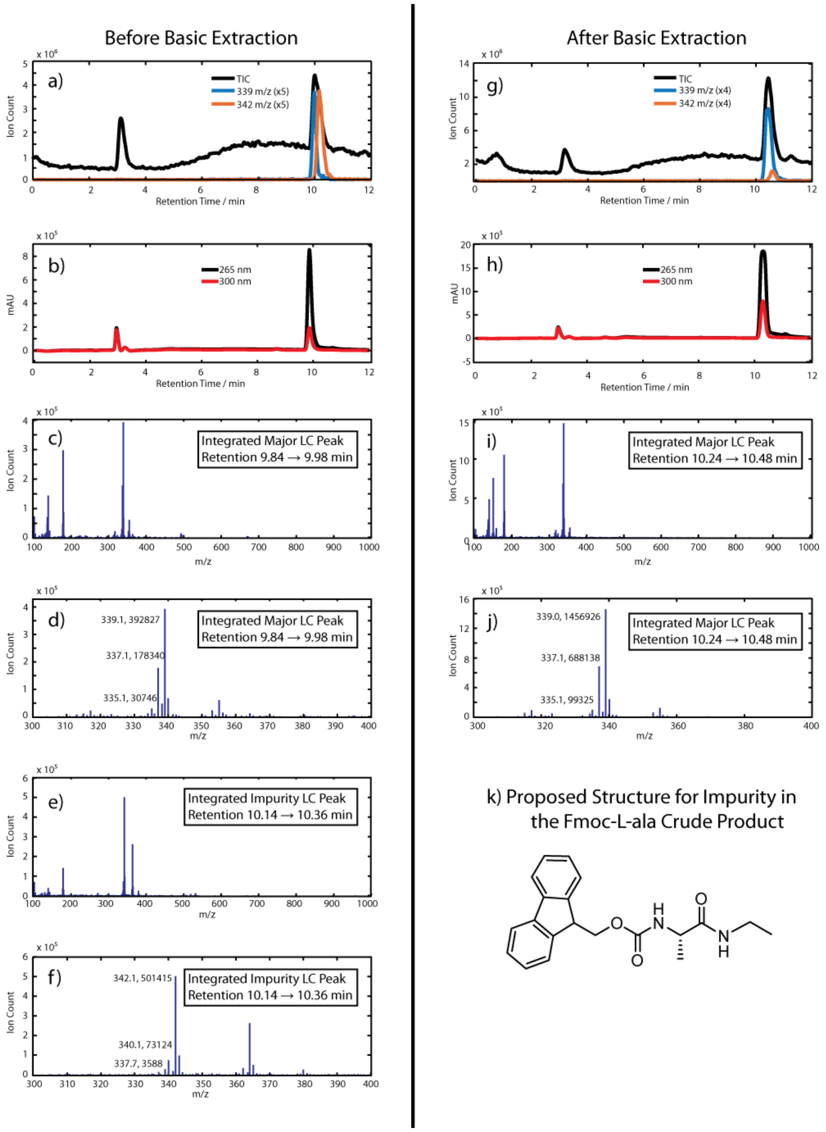


Figure 2.3 - LC-MS data was collected for the ^{18}O labeling reaction product 1-^{13}C Fmoc-L-alanine before and after the extraction performed at basic pH (same gradient as for Fig. 2.2). a) and g) The extraction was successful in removing the byproduct as indicated by monitoring the TIC, product peak (339.1), and impurity (342.1) during LC. b) and h) The product and byproduct overlap closely such that no shoulder is apparent in the UV-Vis data. c) and e) The parent mass in the integrated mass spectra for the first portion and last portion of the unpurified elution are respectively 339.1 and 342.1 m/z. d) and f) A closer look of these peaks revealed the isotope pattern due to ^{18}O labeling. i) and j) The integrated mass spectrum of the major elution has the parent peak at 339.1 m/z and is used to calculate labeling efficiency. k) Proposed structure of the byproduct removed by extraction at basic pH.

methyl ether three times, while maintaining a basic pH of the aqueous layer. The aqueous layer was chilled on ice, and the pH was lowered to 2 with 6 N HCl(aq) at which time a white precipitate formed. The precipitate was filtered and washed with dilute HCl (pH 5) solution. The impurity once apparent in the LC-MS data was removed, indicating that the extracted side-product (Fig. 2.3g and 2.3h) was consistent with our proposed structure. Because the extraction was performed quickly, the ^{18}O labeling efficiency was not altered (Fig. 2.3d and 2.3j). Thus, pH-adjusted extraction appears to be a promising method for alleviating the need for gradient flash chromatography or HPLC.

2.4 General Information for Synthesizing CpRe(CO)$_3$ Derivatives

All of the necessary information for repeating the synthesis of $[\eta^5\text{-}(C_5H_4)(CH_2)_nSSO_2CH_3]Re(CO)_3$ is in Chapter 5. There are, however, a few guidelines and notes that I would give to someone with perhaps more spectroscopy experience than wet lab skills. The skills necessary for these reactions include: running silica flash columns, the characterization of this type of compound and techniques for air-sensitive chemistry.

To start, purification and characterization are important when attempting to synthesize and isolate a new molecule. For the derivatives of CpRe(CO)$_3$, the methods of characterization are limited. Given the mass of these derivatives, LCMS-ESI may appear a good technique at first, but the detected mass may not be anything like the product. This is because when the CpRe(CO)$_3$-derivative exits the ESI nozzle, which acts as a mini-reactor, the CO ligands exchange with CN in odd proportions.[13] EI mass spectrometry is much better at detecting the parent mass of these compounds, but that may not be helpful when troubleshooting the reaction.

The best way I found to troubleshoot these reactions was to run a column and collect NMR spectra for each compound that eluted. With this technique, the products and byproducts were much more clearly identified, even though the process can be time consuming. A detailed guide for setting up flash columns is given in Section 2.6. To interpret NMR data of the eluting peaks, a good resource is Hans Reich's website, which provides peak shift data for many functional groups. If peaks assignments still seem ambiguous, 2D NMR experiments are extremely useful. To set up 2D NMR experiments, Charlie Fry is a good resource. Since CpRe(CO)3 derivatives are difficult to characterize with ESI and MALDI mass spectrometry, proficient column skills and good NMR data interpretations are requirements for obtaining derivatives of $CpRe(CO)_3$.

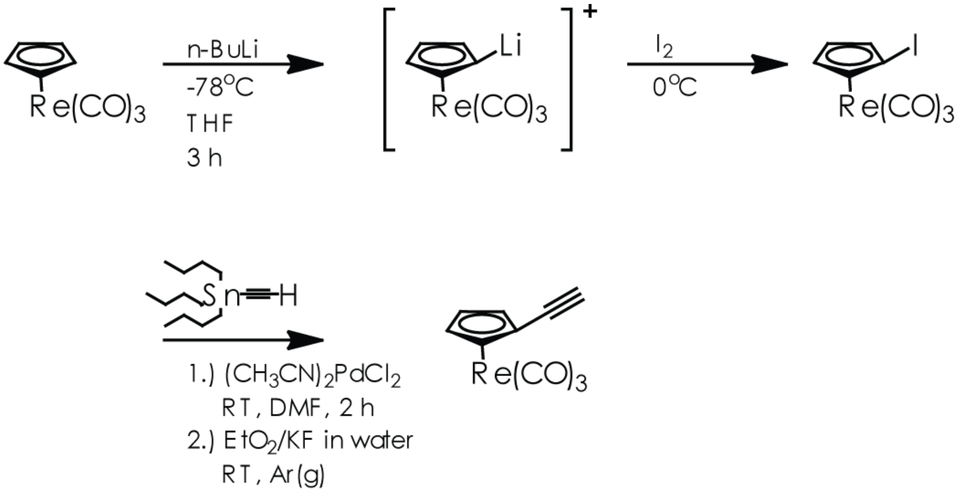
Another necessary skill for these reactions is Schlenk-line technique. Since these reactions are sensitive to both oxygen and water, they will not work if any air or water is present or is exposed to the reaction. To avoid oxygen and water contamination in the reaction, several techniques are required. A good first step is to work out an agreement with a professor that has a good solvent system, so that the reactions can be performed with solvent free of water or impurities. Before obtaining the solvent, all glassware and needles should be dried in an oven, and the Schlenk line should pump down to mTorr range and hold it when the vacuum line is shut. Resolving leaks in the Schlenk line may also be time consuming, but these reactions will fail without performing this step. To prepare a flask for solvent, remove the flask from the oven, allow it to cool down to room temperature in a desiccator, cover with a septa, and tightly wire shut. Attach the flask on the line and pump down to mTorr. Cannula the solvent from the solvent system in the flask following the proper procedures for the solvent system. Always have a syringe handy, so that the option to release pressure is nearby. Once the solvent is obtained, the solvent could also be degased by sparging or Freeze-Pump thaw the solvent several times to be sure that all oxygen is replaced by N_2 or Ar. These measures to prevent any water or oxygen contamination are required to perform the reaction. When performed correctly, Schlenk line techniques will ensure that the reaction is at no time exposed to atmosphere.

Although Schlenk line technique is incredibly useful, use caution and be aware of the hazards that coincide with the technique. A major hazard is the condensation of gases that is possible in the liquid nitrogen cooled trap. If a solid is present in the trap, argon has likely solidified. If this occurs, pressure can build up and cause the glass line to explode. To prevent argon from condensing, turn on the vacuum pump before placing the liquid nitrogen cold trap, and also check for leaks at all the valves by checking that compartments hold vacuum. Also, change the trap regularly, and do not warm the trap without opening a line to atmosphere. Another gas that can condense is oxygen, which is a pretty blue, magnetic liquid. This may occur when the Schlenk line is left open with the vacuum pulling air through the cold trap. If condensed oxygen is observed, evacuate the area immediately, and call safety. By knowing and taking care with the dangers that accompany Schlenk-line usage, the technique can negate the need for a glove box for many air-sensitive reactions.

2.5 Future Work in the Zanni Group: Labeling Azido-Modified Proteins with $[\eta^5\text{-}(C_5H_4)C{\equiv}CH]Re(CO)_3$

Because alkyne/azido-labeling applied to proteins is fast, efficient, and selective enough to be performed in cellular environment, developing a labeling scheme with this chemistry is desirable. I synthesized $[\eta^5\text{-}(C_5H_4)C{\equiv}CH]Re(CO)_3$, which is an alkyne-substituted derivative of

Scheme 2.1


Re(CO)3
n-BuLi
-78°C
THF
3 h
Li
Re(CO)3
+
I2
0°C
I
Re(CO)3
Sn≡H
1.) (CH3CN)2PdCl2
RT, DMF, 2 h
2.) EtO2/KF in water
RT, Ar(g)
Re(CO)3

$CpRe(CO)_3$, by reproducing the synthesis reported by Stille et al (see Scheme 2.1).[14] The synthesis used the Pd catalyst, $(CH_3CN)_2PdCl_2$, which has optimal reactivity under inert atmosphere. With the label prepared, an azido-modified protein was obtained from the Raleigh group at the State University of New York at Stony Brook.

Ivan Peran in the Raleigh group sent us CLT9 protein containing an azido group and the protocol for "clicking" the label onto the protein. I have not labeled the protein yet. This project will be finished by Sean Moran. The directions for adding this label were developed and provided by Ivan Peran in the Raleigh group:

> First, prepare the buffer and stock solutions were prepared. The buffer was 20mM MOPS, 100mM NaCl, pH 7, in which CTL9 is soluble at <1 mM. For the click reaction, Tris buffer should not be used as it binds to copper, thereby inhibiting it. The five tubes were prepared with the following solutions/materials: 1.) 0.5 mg of CTL9 in 100 μL of buffer, 2.) 2.77 mg of the copper ligand, bathophenanthroline disulfonic acid (MW= 554.6g/mol) in 100 μL of H_2O for a final concentration of 50 mM, 3.) 2.82 mg bathocuprionedisulfonic acid (BCS, MW= 564.54g/mol) in 100 μL of H_2O for a final concentration of 50 mM. (This binds to the copper at the end of the reaction), 4.) 7.18 mg of CuBr (MW= 143.4g/mol), 5.) 1 mL DMSO.
>
> For the reaction, a solution with 100 μL total volume of 1mM bathophenanthroline disulfonic acid, 0.25 mM azido-CTL9 protein, 0.6 mM alkyne rhenium label in 20 mM MOPS, 100 mM NaCl, pH 7 was prepared. Afterward, the pH was confirmed to as seven and transferred into a Schlenck flask. Due to the oxygen-sensitive nature of the reaction, all of the reaction mixtures were left in the glove box for six hours, and all operations from here out, were performed in a glove box. DMSO (1 mL) was added to the dry

CuBr. The resulting solution (1 mM) was added to the protein solution. The reaction proceeded for 16 hours. BCS (3 mM) was added to stop the reaction. Purification of the labeled protein should be similar to the protein without the label. The unlabeled protein was purified over a C18 column by HPLC. Buffer A was 100% H_2O and 0.1% TFA, and buffer B was 90% Acetonitrile, 10% H_2O and 0.1% TFA. The gradient used was 100% A for 10min, 20-90% B in 70min.

2.6 Purification by Silica Gel Column

Purification by silica gel column is a very useful tool, especially for synthesizing novel molecules. First, begin by finding the proper isocratic solvent conditions. For the conditions for the appropriate separation, the spreadsheet by Fair and Kormos is helpful.[9] Start testing separation with different solvents using TLC plates. If the compound being purified is acidic or basic, be sure to buffer the solvent to maintain only one conjugate species population. Find the solvent giving an Rf value for the product of ~.3 and that separates the impurities from the product. For an example, see the purification of $^{13}C^{18}O$-labeled Fmoc-L-alanine in Section 2.3. To save time and solvent, design your gradient solvent conditions, for which the article authored by Stevens and Hill is helpful.[11]

Once the gradient solvent conditions are chosen, the column dimensions and silica amounts must be decided. Based on the mass of the crude product, use the table by Still et al. to determine column dimensions.[15] The amount of silica gel was calculated from the spreadsheet,[9] and weighed into an Erlenmeyer flask. As a loose stopper, a few threads (a small pinch) of glass wool were pushed down near the column stopcock with a plastic pipette and prevented the subsequently-added sand from running out of the column. About 0.5-1 cm of sand was added to the column to prevent the distortion from the shape of the end of the column from causing separated bands from mixing. For the addition of silica, a sure-fire technique is the slurry method. For this method, the silica is dissolved in the starting solvent and added to the column, rather than adding it dry (which is a good method for chemists that do this often). As the silica settles and the solvent flows out, do not let the column run dry until the purification is complete. Once ~10 cm of silica is added, add another layer (~.5 cm) of sand. At this point, the crude product is finally added.

To add the crude product, dissolve it in as little of the starting solvent possible, and add as all at once. Once a column volume has eluted, collect the elutants and monitor with TLC continuously. When the product begins to emerge, reserve the most pure aliquots for NMR spectroscopy, mass spectrometry, and possible elemental analysis. If performed properly, the gradient flash column results in 95-100% return of clean product.

2.7 References

(1) Torres, J.; Adams, P. D.; Arkin, I. T. *J. Mol. Biol.* **2000**, *300*, 677.

(2) Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T. *Biopolymers* **2001**, *59*, 396.

(3) Marecek, J.; Song, B.; Brewer, S.; Belyea, J.; Dyer, R. B.; Raleigh, D. P. *Org. Lett.* **2007**, *9*, 4935.

(4) Chan, W. C.; White, P. D.; Editors *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford Univ Press, 2000.

(5) Cruz, L. J.; Beteta, N. G.; Ewenson, A.; Albericio, F. *Org. Process Res. Dev.* **2004**, *8*, 920.

(6) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de, P. J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832.

(7) Wiejak, S.; Masiukiewicz, E.; Rzeszotarska, B. *Chem. Pharm. Bull.* **1999**, *47*, 1489.

(8) Seyfried, M. S.; Lauber, B. S.; Luedtke, N. W. *Org. Lett.* **2010**, *12*, 104.

(9) Fair, J. D.; Kormos, C. M. *Journal of Chromatography A* **2008**, *1211*, 49.

(10) Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923.

(11) Stevens, W. C., Jr.; Hill, D. C. *Mol. Diversity* **2009**, *13*, 247.

(12) Cruz, L. J.; Beteta, N. G.; Ewenson, A.; Albericio, F. *Organic Process Research & Development* **2004**, *8*, 920.

(13) Hop, C. E. C. A.; Brady, J. T.; Bakhtiar, R. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 191.

(14) Lo Sterzo, C.; Stille, J. K. *Organometallics* **1990**, *9*, 687.

(15) Still, W. C.; Kahn, M.; Mitra, A. *The Journal of Organic Chemistry* **1978**, *43*, 2923.

CHAPTER 3

2D IR Lineshapes Probe Ovispirin Peptide Conformation and Depth in Lipid Bilayers*

3.1 Abstract

We report a structural study on the membrane binding of ovispirin using 2D IR lineshape analysis, isotope labeling and molecular dynamics simulations. Ovispirin is an antibiotic polypeptide that binds to the surfaces of membranes as an α-helix. By resolving individual backbone vibrational modes (amide I) using 1-$^{13}C=^{18}O$ labeling, we measured the 2D IR lineshapes for 15 of the 18 residues in this peptide. A comparison of the lineshapes reveals an oscillation in the inhomogeneous linewidth that has a period equal to that of an α-helix (3.6 amino acids). The periodic trend is caused by the asymmetric environment of the membrane bilayer that exposes one face of the α-helix to much stronger environmental electrostatic forces than the other. We compare our experimental results to 2D IR lineshapes calculated using the lowest free energy structure identified from molecular dynamics simulations. These simulations predict a periodic trend similar to the experiment and lead us to conclude that ovispirin lies in the membrane just below the headgroups, is tilted, and may be kinked. Besides providing insight into the antibiotic mechanism of ovispirin, our procedure provides an infrared method for studying peptide and protein structures that relies on the natural vibrational modes of the backbone. It is a complementary method to other techniques that utilize lineshapes, such as

---

* This chapter was originally published in *J. Am. Chem. Soc.* **2010**, *132*, 2832-2838. It was prepared in collaboration with Y.-S. Lin, A. S. Reddy, W. Xiong, J. J. de Pablo, J. L. Skinner, and M. T. Zanni.

fluorescence, NMR and ESR spectroscopies, because it does not require mutations, the spectra can be quantitatively simulated using molecular dynamics, and the technique can be applied to difficult-to-study systems like ion channels, aggregated proteins, and kinetically evolving systems.

3.2 Introduction

Protein structure and dynamics are often studied using lineshape analysis of spectra from native and non-native probes. Spectroscopic lineshapes show signatures of relaxation processes that transfer energy away from the spectroscopic observable (e.g. population transfer) and dephasing that is a result of structural fluctuations of the environment surrounding the observable.[1-3] Fluorescence, electron spin resonance (ESR), nuclear magnetic relaxation (NMR), and other spectroscopies use these processes to study protein structures,[4-15] although each method relies on distinct relaxation and dephasing processes. For example, ESR linewidths are determined by the collision frequency of non-natural nitroxide spin labels with paramagnetic reagents that are added to the solution, and thus are proportional to solvent accessibility, among other factors.[6,8,12] By synthesizing a series of nitroxide mutants, secondary structures are identified by the measured linewidths. For example, this approach has been utilized with much success to identify α-helices, which display an oscillatory trend in the ESR linewidths with a period of 3.6 amino acids.[8] Much work has gone into making ESR and these other approaches viable, and now that the mechanisms of energy transfer and dephasing are understood, these methods provide an established and powerful set of tools for studying protein structures, especially those bound to membranes.

This paper reports that the natural infrared linewidths of peptide and protein backbones can be used to probe structure like those other techniques, but is sensitive to a unique set of environmental forces and dynamical timescales. Infrared linewidths measure dynamics on the femtosecond to few picosecond timescale during which time the only dynamical processes come from hydrogen bond motions and small fluctuations in the side chain and backbone structure. Dynamics like these are of interest at the active site of proteins,[16-18] for example, but it is not immediately apparent how IR linewidths could be used to probe protein secondary and tertiary structures. However, we and others have found that the inhomogeneous linewidth of the amide I band (i.e. the carbonyl stretch of the backbone) is a good structure indicator, because it measures the amount of structural disorder of the backbone and surrounding electrostatic environment that is not averaged out after a few picoseconds.[6,8,12,19-26] Thus, disorder that appears homogeneous in ESR (or other) measurements still contributes to the inhomogeneous infrared linewidth. Thus, this IR approach provides a complimentary perspective to more established techniques on the hydrogen bonding, electrostatics, and dynamics of protein. In addition, it has the following six useful attributes. First, isotope labeling does not perturb molecular structure like large fluorescent or non-natural nitroxide spin labels. Second, it requires only micrograms of material whereas NMR usually requires milligrams. Third, 2D IR spectroscopy may be applied to a broad range of systems, including membrane-bound, aggregated and soluble proteins. Fourth, lipid bilayers can be used rather than micelles as some techniques require, which is especially important for conformationally flexible systems like ion channels. Fifth, 2D IR spectroscopy can be used to study protein dynamics and folding across all relevant timescales.[27-29] Finally, and perhaps most importantly, 2D IR spectra can be modeled from molecular dynamics simulations to provide a direct link between experiments and simulations.

The ability to calculate 2D IR spectra directly from molecular dynamics simulations is a particularly attractive feature, because precise details about the structure can be tested by comparison to experiments. In principle, generating infrared spectra from molecular dynamics simulations just requires a simple Fourier transform of classical dipole time-correlation functions, but in practice such an approach is not accurate enough for quantitative comparison to experiments. However, in the past few years, mixed quantum-classical methods have been developed for obtaining highly accurate 1D and 2D IR spectra from lineshape theory using information from molecular dynamics simulations.[24,25,30-34] There are a number of different methods, but the basic premise of them all is that for each step in a molecular dynamics trajectory, the structure is used to generate vibrational frequencies with an electrostatic map derived from high-level ab initio calculations of model systems, and then IR spectra are calculated using lineshape theory. The comparison of experiment and theory in this paper is an important benchmark in the development of these semi-empirical methods. The methodology is exciting because it provides a direct means of testing molecular dynamics simulations and for interpreting the experimental results based on first principles.

In this paper, we use $^{13}C=^{18}O$ isotope labeling, 2D IR spectroscopy and molecular dynamics simulations to probe the membrane-bound structure of the ovispirin peptide. Ovispirin is an 18-residue polypeptide that resembles the N-terminal domain of the antimicrobial peptide SMAP-29.[35,36] It is α-helical in aqueous buffer and when bound to a lipid bilayer.[37] According to solid-state NMR experiments that measure the angles of N-H bonds in macroscopically aligned bilayers, the helical axis of ovispirin lies parallel to the membrane surface.[38] However, the solid-state NMR measurements only provided angular measurements and, thus, the depth at which ovispirin lies in the membrane bilayer is unknown. Its rotational orientation is also unverified, although a helix wheel diagram puts all of the charged amino acids on one side of the helix (Fig. 3.2c), as is typical for amphipathic peptides that bind parallel to the surface of negatively charged membranes.[4,7,15,39] The equilibrium depth is an important parameter to know, because it is related to the mechanism of antimicrobial action.[40,41] By isotope labeling nearly every residue in ovispirin and measuring the corresponding 2D IR spectra, we have observed an oscillatory pattern in the inhomogeneous linewidths that is consistent with peripheral binding of ovispirin to the membrane in an α-helical structure. To characterize the depth of ovispirin in the membrane, its residue-specific secondary structure, and to understand the structural and electrostatic origin for this lineshape trend, we calculated the free energy profile of ovispirin as a function of membrane depth, which predicts that the peptide lies just below the membrane headgroups. Using this structure, we calculated the 2D IR lineshapes, which agree well with experiment. These simulations point to the strength and disorder of the electrostatic forces surrounding the amide I probe as the most important factors that contribute the most to the observed linewidth trends. Thus, we find that 2D IR spectroscopy can be used to probe the locations of residues in membrane-bound peptides and proteins based on their electrostatic environment.

3.3 Experimental Section

Ovispirin samples were produced by incorporating individually isotope labeled amino acids into the peptide sequence using standard FMOC peptide synthesis protocols. Peptides were purified and combined with 3:1 POPC/POPG lipid vesicles in an aqueous phosphate buffer using H_2O as the solvent, such that the lipid to peptide ratio was at least 21:1. Measurements were performed in H_2O solvent so that arginine sidechain absorption does not interfere with the isotope labeled peaks. Excess bulk H_2O was removed with a nitrogen stream to reduce absorption at 1650 cm^{-1} by the water bend vibrational mode. The bilayers are still fully hydrated as monitored by the water stretching band at 3300 cm^{-1}. At ~1595 cm^{-1} the optical density (OD) of the isotope label was about 0.05. At the frequency of the isotope label, the H_2O bend added an additional OD of 0.25, but because 2D IR scales as the transition dipole, $|\mu|^4$, whereas FTIR scales as $|\mu|^2$, the water does not appreciably contribute to the 2D IR spectrum. 2D IR spectra were collected using a mid-IR pulse shaper and a pump-probe beam geometry. Simulations of ovispirin in a POPC/POPG bilayer were performed with GROMACS using the GROMOS87 force field.[42] After a 100 ns equilibration, trajectories were calculated from production runs of 100 ns for non-mutated and 50 ns for mutated peptides. From the production runs, 2D IR spectra were calculated every 5 ns over a 2 ns window. At each 2 fs timestep, the electrostatic field on the carbon and nitrogen atoms in the amide bond is calculated and converted to a frequency using a map, from which the 2D IR spectra are generated. Details of our methods are given in the Supplemental.

3.4 Results and Discussion

Ovispirin has the sequence H_2N-KNLRR IIRKI IHIIK KYG-COOH. For 11 of the residues, we isotope labeled the naturally occurring amino acid (N, K, L, G or I) using acid-catalyzed ^{18}O exchange on the amino acid followed by protection of the primary amine and side chain for standard FMOC peptide synthesis. For residues 4, 5, 8 and 12, a $^{13}C{=}^{18}O$ labeled glycine mutation was performed instead, because the sidechain groups of R and H are synthetically more difficult to protect. The 2D IR spectrum of each of the 15 peptides was measured after mixing with POPC/POPG lipid vesicles. Under our sample conditions, the peptides are monomeric, α-helical and peripherally bound to the bilayers.

Shown in Fig. 3.1a is a 2D IR spectrum collected with ovispirin labeled at R8G. As is typical for 2D IR spectra, each vibrational mode creates a pair of out-of-phase peaks along the diagonal and cross peaks between coupled modes. The diagonal peak pair at 1747 cm^{-1} is due to the ester stretch of the lipid headgroups, the pair at 1655 cm^{-1} is the unlabeled amide I band, and the pair at 1546 cm^{-1} is the amide II band. The $^{13}C={}^{18}O$ labeled band that we are the most interested in lies in a largely transparent region of the spectrum at 1590 cm^{-1} in between the unlabeled amide I and II bands (square box). There are cross peaks between many of these modes, including the isotope labeled peak and the unlabeled amide I, amide II and lipid ester stretches. These cross peaks provide information on the secondary structure, protonation state, and vicinity to the lipid headgroups, all of which contain useful structural information, but are not the focus of this paper. For this study, we analyze the 2D lineshapes of the isotope labeled mode at 1590 cm^{-1}.

Shown in Fig. 3.1b and 3.1c are the 2D IR spectra of I10 and K15, respectively, for just the frequency range spanning the $^{13}C={}^{18}O$ label. These 2D IR spectra have been collected using a mid-IR pulse shaper which automatically generates absorptive lineshapes and properly phased spectra. Phase cycling was used to remove scatter from the membranes and work in the rotating frame. There was no time delay between the mixing and probing pulses (t_2=0) so that the vibrational modes did not undergo spectra diffusion. We have published a recent review article on using pulse shaping to collect 2D IR spectra.[43] Since these spectra are line-narrowed, the width of the absorption along the anti-diagonal gives the homogeneous linewidth whereas the

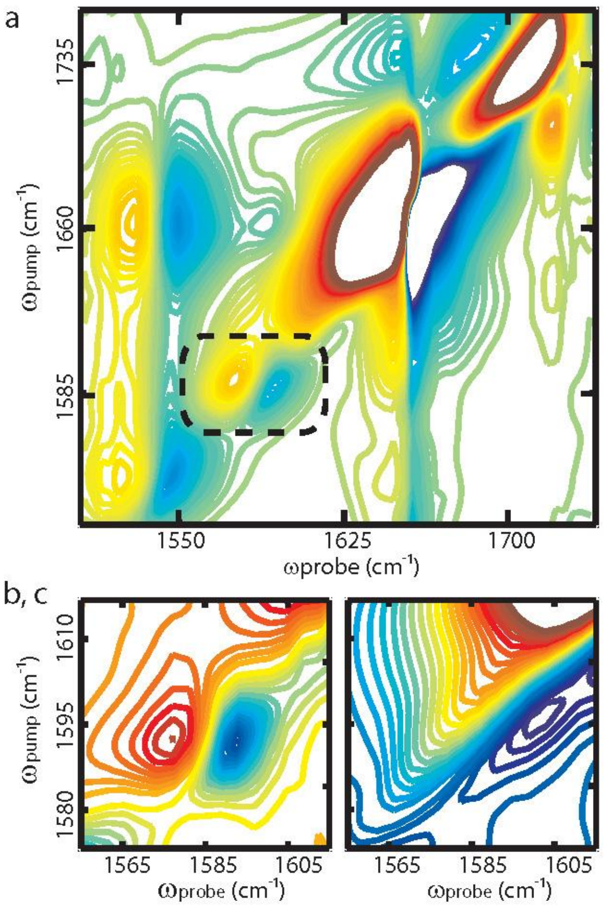


Figure 3.1. 2D IR spectra of $^{13}C{=}^{18}O$ labeled ovispirin. (a) 2D IR spectrum of ovispirin labeled at R8G, which spans the amide I through the lipid headgroup absorptions. The box highlights the labeled mode at ~1590 cm^{-1}. (b) 2D IR spectrum of just the region around the isotope label for I10 and (c) for K15. K15 is much more inhomogeneously broadened than I10 as indicated by its larger ellipticity along the diagonal.

diagonal width is the total linewidth. Thus, the ellipticity of the peaks measures the amount of inhomogeneous broadening. For I10, the peaks are quite round, indicating that they are largely homogeneous. In contrast, the peaks for K15 are very elongated, indicating that they are extremely inhomogeneous. Thus, by simple visual inspection it is clear that the amount of inhomogeneous character is very different between these two residues. The 2D IR spectra for all 15 residues are presented in the Supplemental (Fig. A.1.1).

Isotope labeling with $^{13}C{=}^{18}O$ shifts the amide I band by about 60 cm^{-1} from the unlabeled peaks so that it is spectroscopically resolved, but the wings of the unlabeled amide I and the amide II bands still have non-negligible contributions in this frequency range. To remove these small but unwanted contributions, we fit the spectra using lineshapes for the desired isotope labeled peaks, the amide II band, and the unlabeled amide I band. Crosspeak contributions to the diagonal lineshapes are negligible and were ignored. The functional form of the lineshapes is given in the Supplemental and involves homogeneous and inhomogeneous contributions following Bloch dynamics. We have found in our previous studies that the amide I band of single isotopically labeled residues largely follow Bloch dynamics.[19,21,22] The fits to the spectra are shown in the Supplemental (Fig. A.1.1), and in Table A.1.1, we report the total diagonal, inhomogeneous, and homogeneous linewidth for each residue that we measured.

Of the parameters obtained from the fits, the most interesting are the diagonal linewidths, which are plotted in Fig. 3.2a. They display a periodic trend, with residues 4, 8, 12 and 15 at the crests of a sinusoidal wave that has troughs at residues 6, 11 and 13. The period of the oscillation matches the periodicity of an α-helix, which we ascertain by comparison to a sine wave with a period of 3.6 residues (dashed). Thus, the periodicity indicates that ovispirin is indeed an α-helix when bound to POPC/POPG membranes. Shown in Fig. 3.2c is a helical wheel diagram of

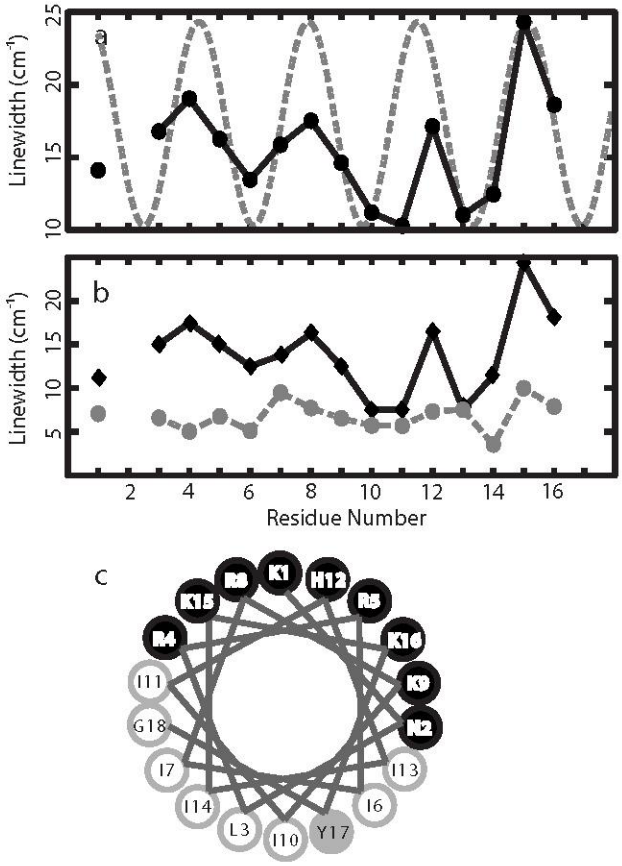


Figure 3.2. 2D IR linewidths taken from Table A.1.1. (a) Diagonal linewidths for each labeled residue. (b) Inhomogeneous (solid) and homogeneous linewidths (dashes). (c) Helical wheel diagram showing ovispirin amphipathy. Hydrophilic and hydrophobic residues are colored dark and light, respectively.

ovispirin. A comparison of the diagonal widths to the helical wheel diagram reveals that the residues with the broadest diagonal linewidths (4, 8, 12 and 15) all fall on the same side of the helix whereas the residues with the narrowest diagonal linewidths are located on the opposite side. There is a ~140% difference in diagonal linewidth between the largest and narrowest diagonal linewidth in ovispirin (24.3 and 10.3 cm^{-1}, respectively). To ensure that the variations in linewidth are caused by membrane binding of monomers and not from interactions between peptides, we also measured the 2D IR spectra of 3 labeled peptides at half the peptide concentration. These dilution experiments produced negligible changes in linewidths (Fig. A.1.2 and Table A.1.2) and confirm that ovispirin is bound to the membrane as monomers.

The trend observed in the diagonal linewidths is almost entirely due to changes in inhomogeneity. The homogeneous and inhomogeneous linewidths from Table A.1.1 are plotted in Fig. 3.2b. The homogeneous linewidths vary by just a few wavenumbers and show no discernable correlation to the peptide structure. In comparison, the inhomogeneous linewidths closely track the diagonal linewidths. In our previous work on the transmembrane peptide bundles CD3ζ and M2, we also found that the homogeneous linewidths are roughly invariant, as have other studies.[19-22,24,25] Thus, we conclude that the homogeneous linewidths are largely intrinsic for peptides and proteins and that the relevant information for 2D IR structural studies comes primarily from differences in inhomogeneous linewidths caused by variations in the environments of residues facing into the membrane versus those facing outwards.

To understand the origin of the trends observed in the 2D IR data and to draw more detailed structural conclusions from the experiments, we turn to molecular dynamics simulations (details are given in the Supplemental). Shown in Fig. 3.3a is the free energy profile of ovispirin as a

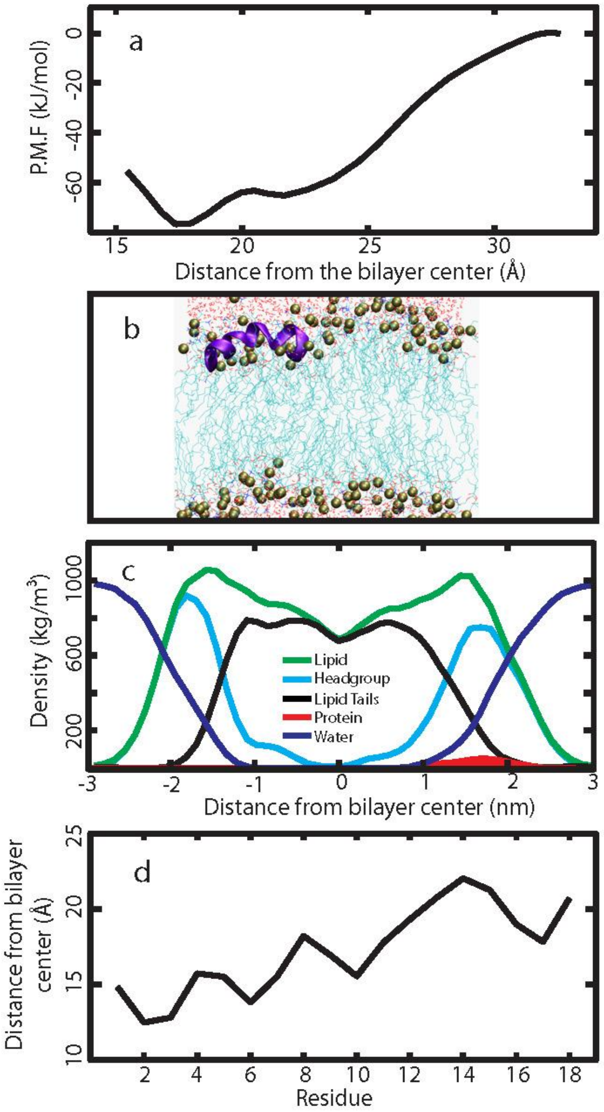


Figure 3.3. Results from molecular dynamics simulations. (a) Potential mean force (PMF) as a function of distance from the bilayer center. (b) Representative snapshot of structure at 17.5 Å from bilayer center. (c) Density profile of the various groups (as defined in the figure) along the bilayer normal. (d) Average depth of α-carbons in each residue over the course of the simulation trajectory.

function of the peptide distance from the center of the membrane bilayer (also see the Supplemental). We observed a free energy minimum at a depth of 17.5 Å from the center of the bilayer. A representative snapshot of the structure is shown in Fig. 3.3b. At this depth, the peptide lies just below the headgroups, which is illustrated by the number density profiles in Fig. 3.3c. It is also tilted and loses its helicity for a few amino acids near residue 14. This is illustrated in Fig. 3.3d, which contains a plot of the average depth of each backbone carbonyl calculated over the course of the 100 ns trajectory. The carbonyl depth oscillates with a 3.6 residue period and is deeper at the N- than the C-terminus, which is a consequence of the α-helix being tilted in the bilayer. At residues 13 and 14, the helix unravels for half of the trajectory, which destroys the periodic trend at this point and indicates that the α-helical secondary structure is less stable in this region of the peptide.

To compare directly the molecular dynamics simulations with experiment, we generated ensemble-averaged (over the 100-ns trajectory) 2D IR lineshapes for each residue in ovispirin. To calculate the spectra for the mutant peptides, independent trajectories were run for each, starting with the initial conformation of the natural peptide, performing the mutation, and then equilibrating the structure for 15 ns before the simulating the 2D IR spectra during the final production run. To convert the molecular dynamics simulation trajectories into 2D IR spectra, at each time step the projection of the electric field on the carbon and nitrogen atoms in the direction of the carbonyl bond was calculated, and an empirically derived map was used to determine the transition frequency.[24] Mixed quantum-classical lineshape formalism was then used to convert the frequency trajectories into 2D IR spectra. Our methodology is described in detail in the Supplemental.

The 2D IR diagonal linewidths calculated from the molecular dynamics simulations are shown in Fig. 3.4a. For residues that were mutated, the linewidth of the simulated mutant is given for that mutated residue, so that the experimental points are directly comparable to the experiment. These are seen to be in quite good agreement with the experimental linewidths in Fig. 3.2a. They also closely mimic the carbonyl depths in Fig. 3.3d. The largest linewidths predicted by simulation is 23. cm^{-1}; the smallest is 13.7 cm^{-1}, which gives a dynamic range of about 80%. For residues 1-11, the simulations predict an oscillation in the linewidths that scales about linearly with the carbonyl depth. The periodicity is lost beginning at residue 12, just before the position in which the peptide helicity breaks down (due to changes in H-bonding). This is most obvious for residue 14, which has conspicuously large frequency fluctuations for a residue that would be expected to face the bilayer interior if the peptide was a perfectly formed α-helix. To understand better the origin of these simulated linewidths, we also calculated the variance of the frequency fluctuations (σ^2), which measures the distribution of frequency fluctuations, but does not include the dynamics that contribute to linewidth narrowing, and broke them down into their respective contributions from the peptide, water, and lipids, which are shown in Fig. 3.4b (the ions make negligible contributions, which are not shown). Also shown is the total variance and the sum of the cross terms. The total variance closely tracks the diagonal linewidths (Fig. 3.4a), indicating that the fast frequency fluctuations that cause line narrowing and determine the homogeneous linewidth are about the same for all the residues, as was observed in the experiments (Fig. 3.2b). All three components (peptide, lipid and water) contribute about equally to the frequency fluctuations of each residue, except for residues that reside on the bottom of the helix, which have smaller contributions from (largely absent) water. However, the observed infrared linewidths cannot be straightforwardly decomposed into their individual components because

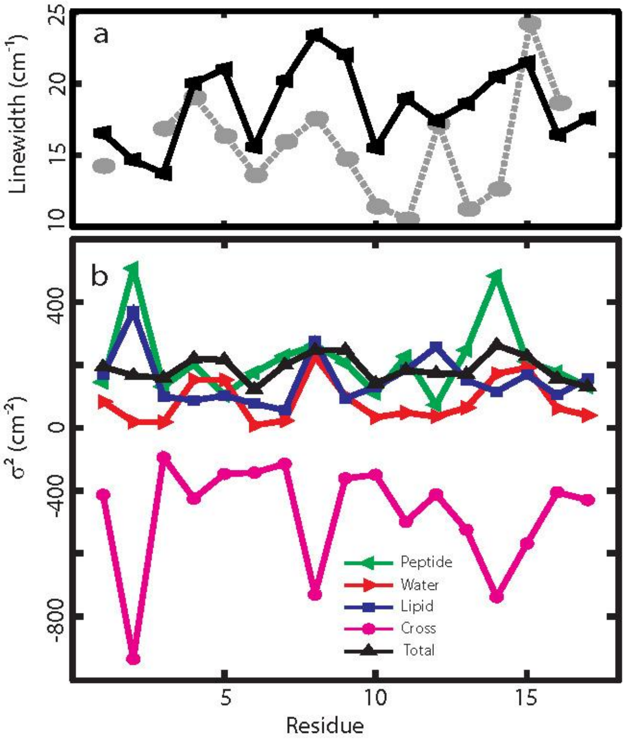


Figure 3.4. 2D IR data calculated from molecular dynamics simulation. (a) Diagonal linewidth of calculated amide I mode for each residue (solid black line) with experimental linewidths (dashed grey line) reproduced from Figure 2a for comparison. (b) Contributions to the frequency variance for each residue.

their frequency fluctuations are correlated. The cross terms are a measure of their correlation, which are negative and nearly equal in magnitude to the individual contributions themselves. Thus, on average, the frequency fluctuations caused by one component is strongly offset from the others due to correlated structural motions.

The 2D IR measurements performed here provide a nearly instantaneous snapshot of the structural distribution of ovispirin and its environment. On the 1 to 2 ps timescale with which the amide I free induction decay lasts, very little motions occurs. Water hydrogen bond dynamics occur on this timescale as do high frequency backbone fluctuations, but little else. Thus, the 2D IR frequencies and linewidths of ovispirin are a measure of the intrinsic structural disorder of the peptide and its surrounding membrane environment. However, the lineshapes do not reflect all types of structural disorder equally. We have found that the residues on the helix that face the interior of the membrane bilayer are much less inhomogeneously broadened than residues facing the headgroups and water, even though the lipid tails of bilayers are highly disordered [44]. Thus, disordered but non-polar environments have little effect on the amide I lineshape, because the amide I mode is mostly sensitive to electrostatic forces. In our previous work on the CD3ζ and M2 transmembrane peptide bundles that span the membrane, we observed narrower amide I linewidths for residues in the center of the membrane than for residues near the membrane/water interface, which is consistent with our results here.[19,21,22,25] For ovispirin, a surface-bound peptide, the change in electrostatics across the region of the membrane at which ovispirin lies causes a 14 cm^{-1} change in the amide I linewidth, which is large enough to be easily quantified with 2D IR spectroscopy. Thus, we have now shown that 2D IR lineshapes can be used to probe the three most common structural motifs of α-helical bound membrane peptides, which are transmembrane peptides, ion channels, and peripherally bound peptides.

The fact that the 2D IR linewidths scale with the strength and disorder of the electrostatic environment provides an intuitive manner in which to interpret the experimental data; the more homogeneous the 2D IR lineshape, the deeper the backbone carbonyl lies in the bilayer. However, because the strongest contributions to the electrostatic forces are from atoms in the immediate vicinity of the vibrational mode (e.g. the first solvent shell), the exact linewidths are determined by the atomic-level details of the peptide structure and bilayer environment. In this regard, the ability to simulate 2D IR spectra from molecular dynamics simulations provides a means to quantitatively investigate specific details about the peptide structure and its lipid environment. In fact, the comparison between simulation and experiment here leads to new insights about ovispirin structure and membrane binding. NMR studies indicate that ovispirin is an α-helix that lies parallel to the plane of the bilayer with the hydrophobic residues facing the interior, but these studies could not determine the depth of the peptide in the bilayer.[38] In our work, the simulated 2D linewidths exhibit an oscillatory trend between residues 1-11 that has a period, phase and amplitude that closely resembles the experiment. Thus, the secondary structure, the packing of the lipids and water around the peptide, as well as its average depth and rotational orientation in the bilayer are well-described by molecular dynamics. However, the simulations do not precisely match the experiments. It appears that the simulations and experiments differ on the tilt of the helix and the secondary structure of ovispirin near the C-terminus. The simulations place the peptide at an average depth of 17.5 Å from the bilayer center with a tilt that puts the N-terminal deeper in the bilayer than the C-terminal by about 4 Å. The overall depth of ovispirin predicted by our molecular dynamics simulations is reasonable since the depth is in agreement with neutron scattering experiments and fluorescent quenching studies on similar peptides.[4,5,7,9,14] However, the experimental 2D IR linewidths *narrow* by about 5 cm^{-1} (on average) from residues 3 to 11, whereas the simulated linewidths *broaden* about 5 cm^{-1} over these same residues. Thus, it appears that the simulations do not capture the proper tilt of ovispirin in the bilayer. Such a difference is not unexpected, because a 100 ns molecular dynamics simulation is not long enough to sample fully low-frequency structural modes like the tilt. There are, of course, also the inevitable questions are about force-field accuracy. Regarding the structure of ovispirin at the C-terminus, the experiments exhibit a sinusoidal trend in linewidths across the entire length of the peptide, suggesting continuous α-helical secondary structure for all measured residues. In comparison, the simulations predict a breakdown in helicity for residues 13-15 so that an oscillatory trend is not predicted for these residues. It is possible that the helix is kinked at this position, which could reconcile the experiment and simulations. We know from our work on the CD3ζ peptide that kinks in α-helices increase the inhomogeneous linewidth,[21,22] which may explain the abrupt increase in the measured linewidths of residues 15 and 16. In fact, solid-state NMR experiments that measure the orientation of NH bonds in oriented bilayers also suggest there is a kink near the C-terminus. It is likely that the molecular dynamics trajectory is too short to fully sample the range of peptide conformations near residue 14. The presence of a kink could be tested, as could the helix tilt, with steered molecular dynamics simulations in which the 2D IR spectra are calculated for a peptide that is constrained to the structure inferred from the experiments. These are important issues for the antimicrobial action of ovispirin, because the peptide depth influences the curvature of the membrane[40,41] and its secondary structure degrades the integrity of the bilayer.[44]

Of course the accuracy with which the molecular dynamics simulations can be converted into 2D IR spectra depends on the quality of the mixed quantum-classical method of calculating the lineshapes. To calculate the frequency trajectories from the molecular dynamics simulations, we used a correlation map derived from calculations of a single peptide bond (N-methylacetamide) in water.[24] The map is expected to work well for small peptides in water, but not as well for hydrophobic environments for which it was not parameterized. Other maps are currently being developed. Even so, the parameterization is already very good, since we correctly predict the linewidths to less than 5 cm^{-1} for all residues, and the variation in linewidth from one side of the helix to the other matches the experiment. The ability to convert molecular dynamics simulations into 1D or 2D IR spectra is very promising, because the conversion only requires simulations of a few hundred picoseconds. As a consequence, the problem reduces to generating the appropriate ensemble average of the structural distribution, but does not require unfeasibly long trajectories for the spectral calculation itself. Thus, infrared spectroscopy provides a direct comparison between simulations and experiment grounded on first principles that is not readily available for other spectroscopies.

The periodic trend in the 2D IR linewidths is caused by the α-helical structure of ovispirin and the gradient of electrostatic forces that exist in bilayers. The interfacial region in which the bilayer changes from hydrophobic to hydrophilic is illustrated in the relative partial density plots of Fig. 3.3c that are calculated from the molecular dynamics simulations. For this POPC/POPG bilayer composition, the bilayer half-thickness is about 30 Å of which the headgroups span 10 Å. The water concentration transitions to the bulk over a distance of about 15 Å. At a depth of 17.5 Å, the peptide lies just below the headgroups so that one side of the helix is hydrated while the other resides in the hydrophobic interior of the bilayer. If the peptide resided in the very center of the bilayer or was only weakly bound to the membrane surface, then we would expect to see very small differences in linewidth since opposite sides of the α-helix would have similar environments. But since the diameter of an α-helix (~9 Å) is only slightly smaller than the width of the interfacial region, we expect that the amide I vibrational modes are sensitive to depth at all relevant positions in the bilayer. This conclusion is supported by our work on the CD3ζ and M2 peptide bundles, which showed continuous differences in linewidth across the bilayer interior.[19,21,22,25] Other secondary structures, such as β-hairpins, should exhibit unique linewidth trends as well. Moreover, since the 2D IR linewidths are mostly determined by the local electrostatic environment of the amide I groups, it might be possible to enhance or suppress the linewidths by adding reagents to the solvent, like is done in EPR or NMR paramagnetic relaxation experiments,[8,12,45] or by mutations that cause local changes in hydration. In fact, molecular dynamics simulations could be used to help design reagents and test the effect of possible mutations.

Along these lines we have used molecular dynamics to ascertain the effect of the mutations on the structure of the peptide and changes in their environment. We calculated the α-carbon depth for each of the equilibrated mutant peptides and found that the backbone depth is unchanged from the native sequence (Fig. A.1.3a). We also looked at the effect of mutated side chains on hydration structure around the labeled backbone and found little or no difference in the standard deviation of hydrogen bonded water molecules. Although one can always argue that the molecular dynamics simulations are too short to determine the effects of the mutations on the global structure, the results are reasonable since the experimental trend we observe is consistent with the α–helical backbone of ovispirin. Thus, the specific nature of the side chains makes a smaller contribution to the linewidths than does the depth or backbone in the bilayer.

3.5 Conclusion

This paper reports that the natural vibrational modes of membrane peptides can be exploited to study their structure and mode of binding. Our approach is complementary to other structural techniques for a variety of reasons, one of which is the ability to compare quantitatively 2D IR spectra to molecular dynamics simulations. Due to longer timescales and other factors, a directly comparison of molecular dynamics simulations to other spectroscopies can be problematic. However, the fast free induction decay time of vibrations makes calculation of infrared spectra easily tractable with simulations. For ovispirin, we found that the free energy minima predicted by simulations agrees well with the experiment, except for the tilt of the helix and the amount of α-helical content near the C-terminal end of the peptide. These results have implications for the anti-microbial mechanism of ovispirin in addition to providing a bench mark system for improving the conversion of molecular dynamics trajectories to infrared spectra. We think that this infrared approach will have important applications to systems that cannot tolerate spin or fluorescence labels, such as the interiors of ion channels or the active sites of enzymes, and in systems that cannot be studied using micelles. Of course the 2D IR method is not limited to equilibrium studies but can be used to monitor changes in linewidths in kinetically evolving systems as well, such as amyloid aggregation. In principle, equivalent linewidth information could be obtained with simple FTIR spectroscopy, since it appears for all systems studied so far that the homogeneous linewidth of amide I vibrational modes are largely invariant to peptide structure and environment. Thus, while 2D IR spectroscopy has many advantages over FTIR spectroscopy such as its ability to suppress background absorptions, the approach of isotope labeling and linewidth analysis could become a widely used tool since FTIR spectrometers are commonly available in many biophysics laboratories. Finally, we note that non-natural infrared probes, such as nitrile groups, are also being developed in an analogous manner.[46-49] Nitrile groups can perturb protein structures, but are much smaller than EPR or fluorescence labels. Moreover, they exhibit frequency shifts and linewidth changes in peptides bound to membranes.[46] Taken all together, site-specific natural and non-natural labeling, 2D IR spectroscopy, and molecular dynamics simulations are helping to develop infrared spectroscopy into a more quantitative structural tool.

3.6 References

(1) Kubo, R. *Adv. Chem. Phys.* **1969**, *15*, 101.

(2) Berne, B. J.; Pecora, R. *Dynamic light scattering : with applications to chemistry, biology, and physics*; Wiley: New York, 1976.

(3) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford Univ Press, New York, 1991.

(4) Chung, L. A.; Lear, J. D.; Degrado, W. F. *Biochemistry* **1992**, *31*, 6608.

(5) Johnson, J. E.; Cornell, R. B. *Biochemistry* **1994**, *33*, 4327.

(6) Hubbell, W. L.; Altenbach, C. *Curr. Opin. Struct. Biol.* **1994**, *4*, 566.

(7) Mishra, V. K.; Palgunachari, M. N. *Biochemistry* **1996**, *35*, 11210.

(8) Hubbell, W. L.; Gross, A.; Langen, R.; Lietzow, M. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 649.

(9) Hristova, K.; Dempsey, C. E.; White, S. H. *Biophys. J.* **2001**, *80*, 801.

(10) London, E.; Ladokhin, A. S. *Curr. Top. Membr.* **2002**, *52*, 89.

(11) Nielsen, R. D.; Che, K. P.; Gelb, M. H.; Robinson, B. H. *J. Am. Chem. Soc.* **2005**, *127*, 6430.

(12) Fanucci, G. E.; Cafiso, D. S. *Curr. Opin. Struct. Biol.* **2006**, *16*, 644.

(13) McDermott, A. *Annual Review of Biophysics* **2009**, *38*, 385.

(14) Orioni, B.; Bocchinfuso, G.; Kim, J. Y.; Palleschi, A.; Grande, G.; Bobone, S.; Park, Y.; Kim, J. I.; Hahm, K. S.; Stella, L. *Biochimica Et Biophysica Acta-Biomembranes* **2009**, *1788*, 1523.

(15) Franzmann, M.; Otzen, D.; Wimmer, R. *ChemBioChem* **2009**, *10*, 2339.

(16) Lim, M. H.; Hamm, P.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15315.

(17) Rella, C. W.; Kwok, A.; Rector, K.; Hill, J. R.; Schwettman, H. A.; Dlott, D. D.; Fayer, M. D. *Phys. Rev. Lett.* **1996**, *77*, 1648.

(18) Bandaria, J. N.; Dutta, S.; Hill, S. E.; Kohen, A.; Cheatum, C. M. *J. Am. Chem. Soc.* **2008**, *130*, 22.

(19) Mukherjee, P.; Krummel, A. T.; Fulmer, E. C.; Kass, I.; Arkin, I. T.; Zanni, M. T. *J. Chem. Phys.* **2004**, *120*, 10215.

(20) Fang, C.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109*, 18652.

(21) Mukherjee, P.; Kass, I.; Arkin, I.; Zanni, M. T. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8571.

(22) Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T. *J. Phys. Chem. B* **2006**, *110*, 24740.

(23) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7720.

(24) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2009**, *113*, 592.

(25) Manor, J.; Mukherjee, P.; Lin, Y. S.; Leonov, H.; Skinner, J. L.; Zanni, M. T.; Arkin, I. T. *Structure* **2009**, *17*, 247.

(26) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17751.

(27) Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14237.

(28) Hamm, P.; Helbing, J.; Bredenbeck, J. *Annu. Rev. Phys. Chem.* **2008**, *59*, 291.

(29) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6614.

(30) Schmidt, J. R.; Roberts, S. T.; Loparo, J. J.; Tokmakoff, A.; Fayer, M. D.; Skinner, J. L. *Chem. Phys.* **2007**, *341*, 143.

(31) Jansen, T. L. C.; Knoester, J. *Biophys. J.* **2008**, *94*, 1818.

(32) Zhuang, W.; Hayashi, T.; Mukamel, S. *Angew. Chem.* **2009**, *48*, 3750.

(33) Mukamel, S.; Abramavicius, D.; Yang, L. J.; Zhuang, W.; Schweigert, I. V.; Voronine, D. V. *Acc. Chem. Res.* **2009**, *42*, 553.

(34) Jeon, J.; Yang, S.; Choi, J. H.; Cho, M. *Acc. Chem. Res.* **2009**, *42*, 1280.

(35) Kalfa, V. C.; Jia, H. P.; Kunkle, R. A.; McCray, P. B.; Tack, B. F.; Brogden, K. A. *Antimicrob. Agents Chemother.* **2001**, *45*, 3256.

(36) Bartlett, K. H.; McCray, P. B.; Thorne, P. S. *Int. J. Antimicrob. Agents* **2004**, *23*, 606.

(37) Sawai, M. V.; Waring, A. J.; Kearney, W. R.; McCray, P. B.; Forsyth, W. R.; Lehrer, R. I.; Tack, B. F. *Protein Eng.* **2002**, *15*, 225.

(38) Yamaguchi, S.; Huster, D.; Waring, A.; Lehrer, R. I.; Kearney, W.; Tack, B. F.; Hong, M. *Biophys. J.* **2001**, *81*, 2203.

(39) Hristova, K.; White, S. H. *Biophys. J.* **1999**, *76*, A219.

(40) Zemel, A.; Ben-Shaul, A.; May, S. *J. Phys. Chem. B* **2008**, *112*, 6988.

(41) Almeida, P. F.; Pokorny, A. *Biochemistry* **2009**, *48*, 8083.

(42) Van Gusteren, W. F. B., H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Groningen, The Netherlands, 1987.

(43) Shim, S.-H.; Zanni, M. T. *PCCP* **2009**, *11*, 748.

(44) Jaud, S.; Tobias, D. J.; Falke, J. J.; White, S. H. *Biophys. J.* **2007**, *92*, 517.

(45) Livshits, V. A.; Dzikovski, B. G.; Marsh, D. *J. Magn. Reson.* **2001**, *148,* 221.

(46) Mukherjee, S.; Chowdhury, P.; DeGrado, W. F.; Gai, F. *Langmuir* **2007**, *23*, 11174.

(47) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200.

(48) Oh, K. I.; Choi, J. H.; Lee, J. H.; Han, J. B.; Lee, H.; Cho, M. *J. Chem. Phys.* **2008**, *128*.

(49) Kozinski, M.; Garrett-Roe, S.; Hamm, P. *J. Phys. Chem. B* **2008**, *112*, 7645.

CHAPTER 4

Parallel β-sheet vibrational couplings revealed by 2D IR spectroscopy of an isotopically labeled macrocycle: Quantitative benchmark for the interpretation of amyloid and protein infrared spectra.[*]

4.1 Abstract

Infrared spectroscopy is playing an important role in the elucidation amyloid fiber formation, but the coupling models that link spectra to structure are not well tested for parallel β-sheets. Using a synthetic macrocycle that enforces a two stranded parallel β-sheet conformation, we measured the lifetimes and frequency for six combinations of doubly $^{13}C{=}^{18}O$ labeled amide I modes using 2D IR spectroscopy. The average vibrational lifetime of the isotope labeled residues was 550 fs. The frequencies of the labels ranged from 1585 to 1595 cm^{-1}, with the largest frequency shift occurring for in-register amino acids. The 2D IR spectra of the coupled isotope labels were calculated from molecular dynamics simulations of a series of macrocycle structures generated from replica exchange dynamics to fully sample the conformational distribution. The models used to simulate the spectra include through-space coupling, through-bond coupling, and local frequency shifts caused by environment electrostatics and hydrogen bonding. The calculated spectra predict the linewidths and frequencies nearly quantitatively.

---

[*] This chapter will be submitted to *J. Am. Chem. Soc*. in April 2012. It was prepared in collaboration with A. M. Almeida, L. Wang, C.-C. Chiu, M. McGovern, J. J. de Pablo, J. L. Skinner, S. H. Gellman, and M. T. Zanni.

Historically, the characteristic features of β-sheet infrared spectra have been attributed to through-space couplings such as transition dipole coupling. While these couplings are important, especially for the intensity distribution, we find that frequency shifts of the local carbonyl groups due to nearest neighbor couplings and environmental factors are more important. Through-space couplings dictate the spectral intensities. As a result, the characteristic absorption spectra empirically used for decades to assign parallel β-sheet secondary structure arises because of a redistribution of oscillator strength, but the through-space couplings do not themselves dramatically alter the frequency distribution of eigenstates much more than already exists in random coil structures. Moreover, solvent exposed residues have amide I bands with >20 cm^{-1} linewidth. Narrower linewidths indicate that the amide I backbone is solvent protected inside the macrocycle. This work provides calculated and experimentally verified couplings for parallel β-sheets that can be used in structure-based models to simulate and interpret the infrared spectra of β-sheet containing proteins and protein assemblies, such as amyloid fibers.

4.2 Introduction

Vibrational spectroscopy is one of the most commonly used techniques for assessing the secondary structure content of proteins.[1,2] It is often the first tool that a researcher uses to assess structure in membrane bound proteins and to monitor dynamics in protein folding, to name only two applications.[3-5] More than 50 years ago it was recognized that the relationship between protein secondary structure and the observed amide I absorption band is caused by vibrational coupling between the amide groups of the protein backbone.[6] That realization enabled researchers to begin modeling the infrared spectra of proteins using a local mode or excitonic-type Hamiltonian in which only the amide I vibrational modes need to be considered.[7] Excitonic models have become a key tool in the interpretation of the infrared spectra of proteins, because they enable a very straightforward and intuitive structure-spectra relationship, in contrast to the vibrational modes of most other molecules, and even the other protein amide modes, for which few simple models currently exist.[8-12]

In the last few years, improvements to the methods for calculating the parameters of these excitonic Hamiltonians is enabling a transition from semi-quantitative to quantitative simulations of protein amide I spectra.[13-19] The two primary ingredients for exciton Hamiltonians are the frequencies for each of the individual amide groups and the coupling between pairs. Amide I frequencies are now being calculated from mixed quantum/classical theories using parameters developed from ab initio or empirical fitting parameters.[13,20] These methods not only account for hydrogen bonding, but also frequency shifts caused by microsolvation, local dielectric effects, and the dependence of the vibrational mode itself on the dihedral angles. To calculate the couplings, models now exist that include ab initio, transition charge, and transition dipole density models.[7,15,16,19,21] With an exciton Hamiltonian in place, mixed quantum/classical lineshape theory is used to simulate the infrared spectra.[22,23] The result is an infrared spectrum that includes the molecular structure, its surrounding environment, and the structural dynamics of both. Thus, with a quantitative theory in hand, structural and dynamical information can be interpreted from infrared spectroscopy and 2D IR spectroscopy in particular.

While exciton models are improving and simulations of proteins are abundant in the literature, the accuracy with which infrared spectra can be calculated is not fully tested. Pieces of the modeling have been developed quite accurately. For example, the frequency shifts caused by electrostatics and hydrogen bonding of the environment have been developed using empirical fits to model system.[13,24] The frequency shift caused couplings and dihedral angles is less well tested. Part of the problem is of an experimental nature. The infrared spectra of proteins and even small peptides are too congested to resolve individual amide I absorption bands.[25] As a result, when a simulation does not match an experiment, it is difficult to ascertain which part of the exciton model failed. One way to relieve congestion is through the use of isotope labeling. $^{13}C^{18}O$-labeling of backbone carbonyl groups enables individual or subsets of amide I modes to be spectroscopically resolved and vibrationally decoupled from the unlabeled residues.[23,26-28] As a result, one can measure the frequencies of individual residues or pairs of amide I modes. These measurements provide the fundamental quantities necessary to test and calibrate well-defined elements in the excitonic Hamiltonian.

It is especially important to benchmark the accuracy for common secondary structures and in typical environments. Using various combinations of isotope labeling and either 1D or 2D IR spectroscopy, the fundamental coupling constants and environmental frequency shifts for many of the most common protein secondary structures have been determined in a variety of environments. Soluble helices have been isotopically labeled at single, double and quadruple positions to infrared spectra that probe the most important coupling constants in an α-helix.[29-31] Isotope labeling has been used to study the coupling in 3-10 helices.[32] Frequency shifts associated with environments have been characterized in soluble and membrane bound α-helices, including surface bound monomers and α-helical bundles with and without water pores.[23,27] Many of the coupling constants for β-turns and anti-parallel β-sheets have been measured with isotope labeling and 2D IR spectroscopy.[33-35] As a result of these experiments, there is reasonable certainty in the accuracy to which α-helices, turns and anti-parallel β-sheets are modeled. Thus, for these secondary structures and environments, there exist benchmark calculations and experiments that quantify the accuracy of the structure/spectra relationship.

Bench marking parallel β-sheets has proven to be more difficult, because small peptides do not naturally adopt parallel β-sheet structure in water.[36,37] Parallel β-sheets are less common than antiparallel β-sheets,[38] and often occur in large proteins and protein aggregates with repeated structural motifs such as amyloid fibers, β-solenoids, leucine rich repeats, and TIM barrels.[39-41] Parallel strands are also found in mixed parallel/antiparallel β-sheets in systems where the orientations of the strands are enforced by the overall protein fold.[38] Nonetheless, proteins that are large enough to form parallel β-sheets are also difficult to isotope label. In multi-peptide systems, like amyloid fibers, one cannot fully control the arrangement of isotope labels to probe all the necessary coupling constants. As a result, the coupling constants of parallel β-sheet are probably the least well characterized of all the common secondary structure elements, even as they are becoming increasingly important with the rising interest in protein aggregation diseases and biocompatible peptide gels.[42]

In this paper, we focus on experimentally validating and testing theoretical models for the couplings and environmental effects of parallel β-sheets. To overcome the above problems associated with parallel β-sheets, we utilize a synthetic peptide macrocycle shown in Fig. 4.1A composed of two β-strands. The strands are linked at the N-termini by a dimer of glycine and succinic acid and at the C-termini by a dimer of D-Proline and 1,2-diamino-1,1-dimethylethane.[36] These non-natural turns force the β-strands into parallel alignment that is well-characterized by NMR (See Appendix 2). We isotope label pairs of residues to sample interactions between residues that we believe are the most important matrix elements in the exciton Hamiltonian. The insights gained from these simulations and experiments provide new

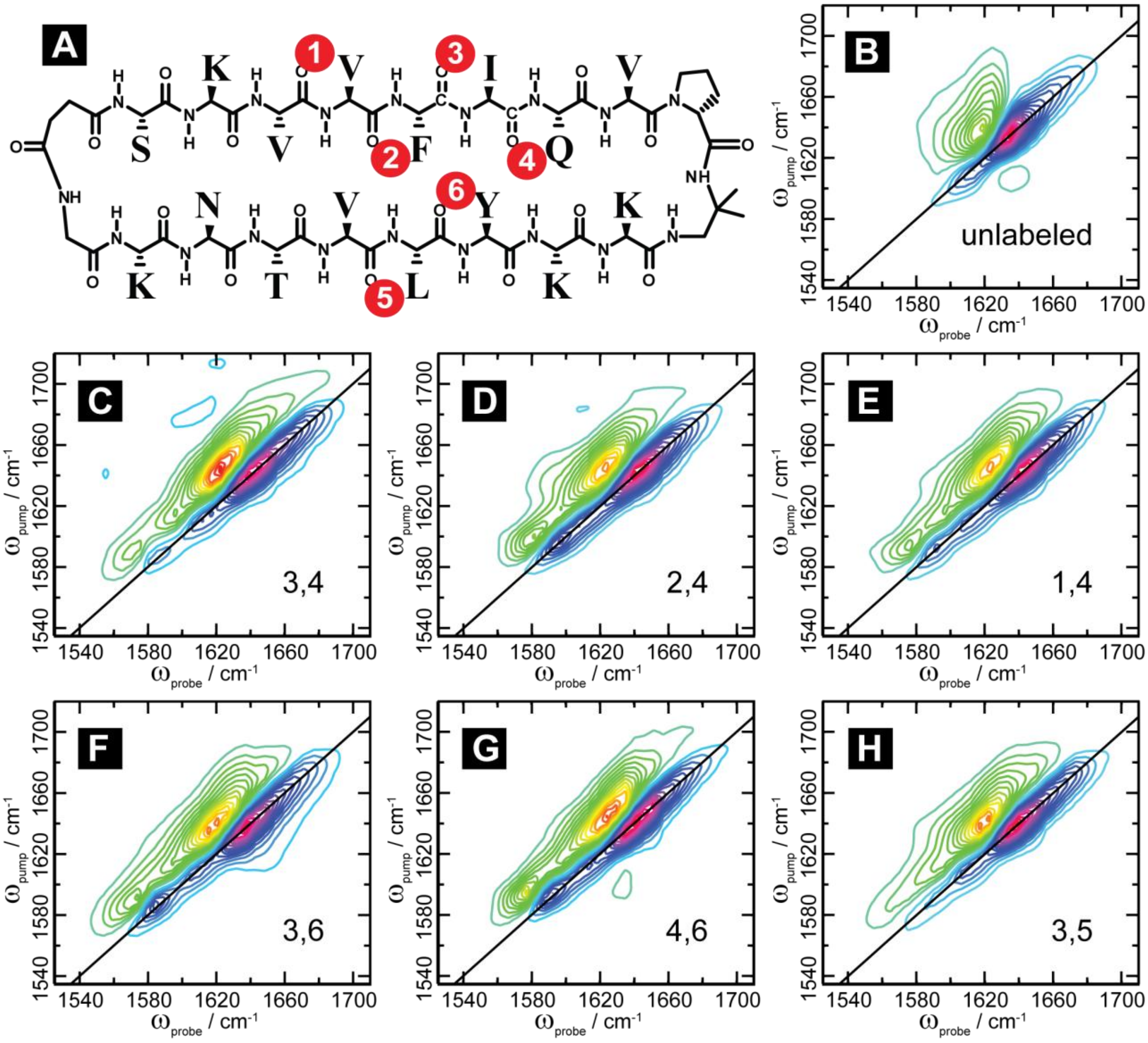


Figure 4.1 - Macrocycle structure with numbering scheme for isotope labeled pairs (a). 2D IR data for the unlabeled macrocycle (b) and for the doubly-labeled macrocycle (c-h).

insight into how to interpret the infrared spectra of parallel β-sheet proteins and the design of isotope labeling strategies for testing structural models, such as for amyloid structures.

4.3 Experimental Section

4.3.1 Peptide Synthesis and Isotope Labeling

1-^{13}C amino acids (99% 13C, Cambridge Isotope) were isotope labeled by acid catalyzed ^{18}O exchange with ^{18}O water (95% 18O, Isotec) under stream of Argon gas. Isotope labeling efficiency was 84-94% as assessed by LCMS-ESI. Macrocyclic peptides were synthesized using standard microwave assisted Fmoc-based solid phase peptide synthesis techniques with side-chain attachment to the resin to facilitate backbone cyclization as previously described. More details, including purification, characterization, and NOE NMR measurements, are provided in the SI. To prepare the samples for 2D IR experiments, residual TFA from HPLC purification was removed and the peptide simultaneously deuterated by performing three additions of ~1 mL 0.1 DCl in D2O per mg of peptide with subsequent lyophilization. The peptides were then dissolved in deuterated phosphate buffer (pH 3.8) at about 10 mM concentration.

4.3.2 Spectroscopic Methods

The 2D IR setup used was previously described.[43] In brief, 800 nm, 60 fs pulses were generated by a Ti:Sapphire laser, which were converted to six micron light by difference frequency mixing optical parametric amplifier. The mid-IR light was split such that a third of this light was sent through a Ge-AOM to shape the pulse envelope to a double pump pulse. The pump was overlapped spatially and temporally with the remaining mid-IR light (probe) on the sample to generate rephasing and non-rephasing signal heterodyned with the probe beam. The signal and probe were dispersed in the monochromator and detected on one-row array of 64 HCT elements.

The pump pulse pair was incrementally varied 0 to 2544 fs in 24 fs steps at the rate of the laser shots (1000 Hz). Lifetime measurements were performed by scanning a translational stage along the pump path.

4.3.3 Molecular Dynamics (MD) Simulations

The initial structure of cyclic peptide was generated according to NMR data. The GROMOS96 53a6 force field[44-46] was used to model the peptide, and the SPC model[47] was used for water molecules. Parameters for the C-terminal and N-terminal linker were assigned according to the atom types in the force field. The lysine side chains were protonated to give a net charge of +4 for the peptide. Chloride ions were added to neutralize the overall system charge. The system was solvated with 7713 water molecules. All bonds were constrained using the linear constraint solver method.[48] The system was simulated in the NPT ensemble at 298K and 1 bar using the Nose-Hoover thermostat[49,50] and the Parinello-Rahman barostat,[51] respectively. Long-range electrostatic interactions were treated with a particle-mesh Ewald sum.[52,53] Equilibrium MD simulations were performed with an integration time step of 2 fs for 100 ns using the GROMACS molecular simulation package.[54-56]

To ensure sufficient sampling of peptide conformations, we per-formed replica exchange MD (REMD). [57-59] 42 replicas were setup at temperatures chosen from 298K to 380K following the procedure of Rathore et al.[60] Exchanges between neighboring replicas were attempted every 2 ps. 100 ns simulation was performed for each replica, resulting a total simulation time of 4.2 microseconds.

4.3.4 Lineshape and Coupling Calculations

The $^{13}C{=}^{18}O$ isotope labels lower the amide I frequency so that the two labeled residues are spectrally isolated from all the other amide modes. Accordingly, they are modeled theoretically as isolated chromophores. For each pair, the amide I Hamiltonian is a 2 × 2 matrix, whose diagonal elements are the local frequencies of the two chromophores and the off-diagonal element is the coupling between them. The local frequencies were calculated using a backbone frequency map,[13] corrected by the nearest-neighbor frequency shift.[61] The $^{13}C{=}^{18}O$ isotope shift was taken into account by shifting the local frequencies by -61.5 cm^{-1}, which is derived from experiments discussed below. Couplings between adjacent carbonyls were treated using a nearest-neighbor coupling map.[61] Couplings between all other carbonyls groups were calculated using a transition dipole coupling (TDC) model developed by Torii and Tasumi.[62] From the Hamiltonian, 2DIR absorptive spectra of the zzzz polarization were calculated,[63,64] with T_1=600 fs and T_2=0.5*T_1, which matches experimental conditions. The polarization anisotropy effects were accounted as described previously.[65]

To calculate the 2DIR spectra, MD simulations were performed using GROMACS with the same parameters described in the MD Simulations section. To better sample the configurations of the macrocycle, a total of 10 simulations were performed, each starting from configurations 10 ns apart in the 100 ns MD simulations. Each simulation was run for 5 ns and the trajectories were saved every 20 fs for spectra calculations. The spectra and spectral features reported below were averaged over the 10 simulations, and the error bars are presented as twice the standard deviation of the mean.

4.4 Results and Discussion

4.4.1 2D IR spectra of isotopically labeled macrocycles reveal large frequency shifts indicative of structure.

Shown in Fig. 4.1A is the structure of the β-sheet macrocycle used in this study. Seven sets of macrocycles were synthesized. Six macrocycles each had two $^{13}C^{18}O$ isotope labels, located at residues chosen from 2 of the 6 locations marked in Fig. 4.1A. One macrocycle was unlabeled. The combinations of isotope labels were selected to probe the inter-strand couplings, intra-strand couplings and the frequency shifts associated with carbonyl groups hydrogen bonded in the β-sheet structure versus those pointed outwards and solvated by water. Shown in Fig. 4.1B-H are the 2D IR spectra for each of the 7 macrocycles. The unlabeled macrocycle (Fig. 4.1B) has the strongest absorption at ~1640 cm^{-1} and stretches to 1680 cm^{-1}, as measured by the diagonal frequency. These features are caused by the β-sheet modes of the unlabeled amino acids. The weaker peak pair at 1612 cm^{-1} is due to the amide group of the macrocycle linker and the proline backbone amide I mode.[66] 2D IR spectra of the isotope labeled macrocycles exhibit an additional absorption near 1590 cm^{-1}, which is due to the pairs of $^{13}C{=}^{18}O$ labeled residues. It is this absorption band that is the main focus of our study.

We are mainly interested in four quantities. The first quantity is the vibrational frequencies of the isotope labeled features, which provide information about the coupling between the isotope labels and their individual amide I frequencies. The second quantity is the frequency of the unlabeled amide feature, because its frequency provides information about the extent that the vibrational excitons are delocalization across the parallel β-sheet. The third quantity is the 2D lineshapes of the isotope labels, which is a measure of the homogeneous and inhomogeneous vibrational dynamics, and thus provides information on the distribution of frequencies caused by the protein and solvent structural distribution. Finally, we measure the vibrational lifetimes of the isotope labels, to ascertain the proportion of the lineshape governed by vibrational relaxation and to correlate hydration to lifetime, as has been observed in other systems.[67]

It is immediately apparent from the data that the isotope labels are sensitive to the macrocycle structure and environment. Plotted in Fig. 4.2A are slices through the diagonal (ω_{probe}=ω_{pump}) of the 2D IR spectra. Simple inspection reveals that the isotope label frequency spans 10 cm^{-1} and has a linewidth that ranges from very sharp to quite broad. The peak maximum of the unlabeled features near 1640 cm^{-1} also depends on the location of the isotope labels, indicating that the delocalization of the excitons across the β-sheets is disrupted by the labels. To quantitatively extract the frequencies and linewidths of these features, we fit these slices to three Gaussians: one for the label, one for the proline absorption, and one for the unlabeled residues. The frequency of the proline Gaussian was constrained to +/- 2 cm^{-1} at 1610 cm^{-1}, since its frequency should not vary from sample to sample. Because of the large frequency separation, the fit to the unlabeled amide absorption band has little influence on the isotope labeled band. A single Gaussian was used to fit the isotope labeled region because the individual isotope labels are not resolved. We did not expect to resolve the individual residues, because previous studies have found that the typical frequency dependence of an isotope labeled amide I band is about 10 cm^{-1}.[4] Nonetheless, the peak maximum and linewidth, which will correspond to the summation of the two labels, is sensitive to structure as our data shows and has also been true in isotope labeled α-helices.[3,23,27] The fits are shown in the Fig. A.2.10. The extracted

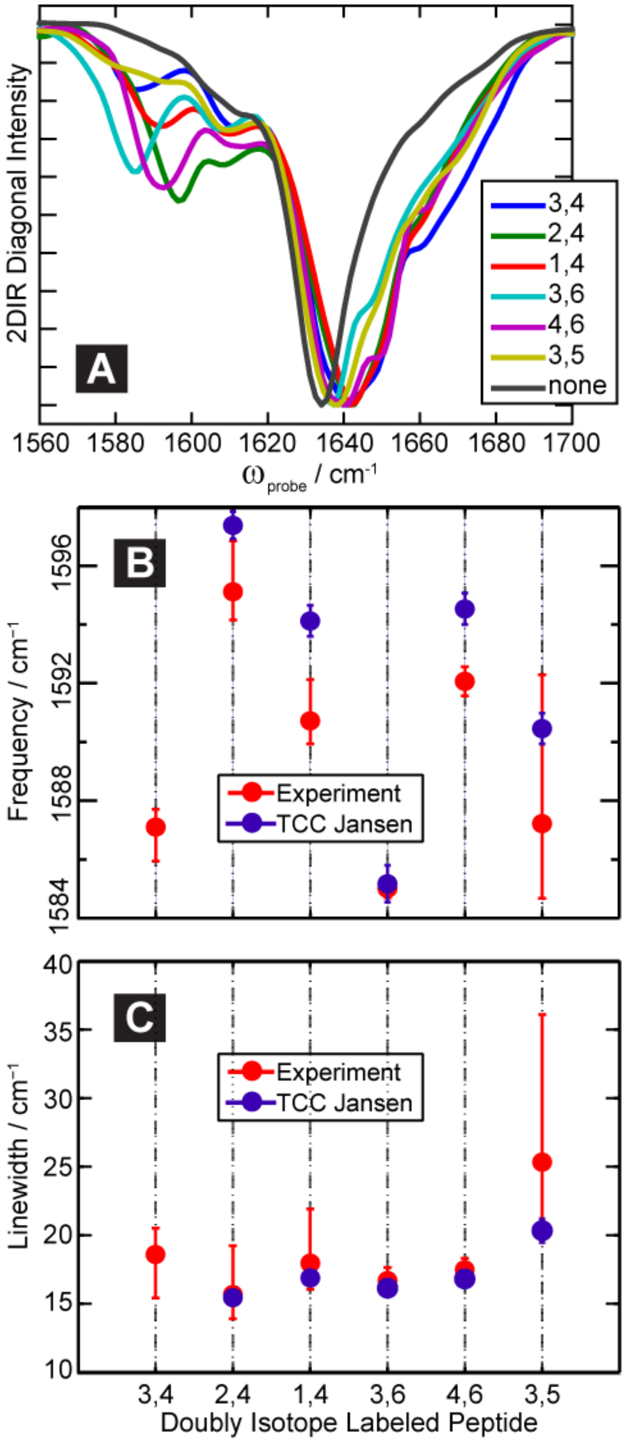


Figure 4.2 - Overlapped slices along the diagonal of experimental 2DIR spectra.

frequencies and lineshapes are plotted in Fig. 4.2B-C (red). These fits provide three of the desired experimental quantities named above.

From this data, we note a few observations. First, regarding the isotope labeled region of the spectra, the sample with the lowest frequency absorption is the (3,6) labeled peptide. These labels lie on different strands and are in-register. In the literature, it is commonly thought that the coupling between these two residues is the most important, because transition dipole coupling (which is one of the oldest and most often used coupling models) predicts this coupling to be large and negative,[68] which is consistent with our data. Second, the spectrum of the (3,5) labeled peptide, in which both labels face outward, has a much broader linewidth than any of the others samples. This observation is consistent with environmental disorder being more important for residues on the outside than those that point inward and are less solvent exposed. Third, the $^{12}C^{16}O$ feature exhibits the largest blue shift for all three peptides in which the two labels are located on the same strand. Macrocycles in which the labels are located on opposite strands shift only half as much. This observation reveals that the most important couplings in parallel β-sheets are intra-strand, not inter-strand, couplings. As we discuss below, these empirical observations are confirmed by our simulations.

4.4.2 Structural Distribution of the Macrocycle

Infrared spectroscopy samples molecular conformations on the timescale of a few picoseconds. Thus, it provides a snapshot of the structural distribution on a timescale in which only the fast solvent rearrangements or hydrogen bonding fluctuations occur.[69] As a result, the infrared spectrum can be calculated from a molecular dynamics simulation of only 1 or 2 nanoseconds, which is an exciting feature of infrared spectroscopy because nanosecond simulations are easily accessible with modern computers. However, since infrared spectroscopy is an ensemble measurement, it measures all structural conformations present in the sample. Thus, to properly simulate an FTIR or 2D IR spectrum, one needs to sample and obtain the relative weightings for the conformational distribution.

The structure of the parallel macrocycle is well characterized by NMR See Appenix 2). We took two approaches to characterize the distribution of structures. First, we ran a 100 ns trajectory. Second, we employed replica exchange dynamics. In both cases, the distribution of structures is similar to the distribution generated from using the experimental NMR constraints. As a result, we are confident that we have fully sampled the peptide structural distribution.

Shown in Fig. 4.3 is an overlay of 10 NMR derived structures. Notice that the macrocycle becomes structurally disordered at the two ends, but is much more stable in the middle where the isotope labels reside. In the middle, all 10 structures retain their inter-strand hydrogen bonding. Deviations from the average structure come mostly from out-of-plane rotations. At no point in any of the simulations is a large-scale disruption or puckering of the macrocycle observed. Thus, frequency fluctuations in the 2D IR spectra will come from the out-of-plane rotations, which will alter the coupling constants, hydrogen bond shifts associated with these rotations, and the effects of solvation (which are not shown in this figure).

4.4.3 Modeling Parameters and Simulated Infrared Spectra

Using the distribution of macrocycle structures discussed above, we have simulated the 2D IR spectra of the isotope labels. We aim to obtain insights into three different sets of parameters that contribute to these simulations. These parameters are the local mode frequency shifts caused by environmental electrostatics and hydrogen bonding, the dependence of the local

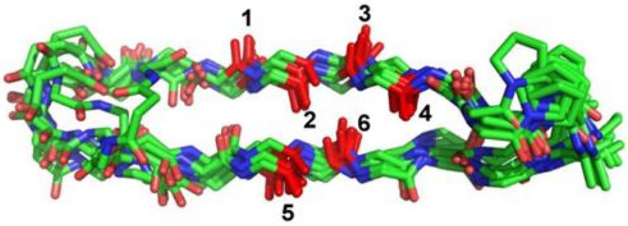


Figure 4.3 - NMR-NOE structure of the macrocycle overlaid with numbering scheme for isotope labeling indicated.

mode frequency on the dihedral angles, and the model used to calculate through-space coupling. We briefly review these parameters below.

The excitonic Hamiltonian is a matrix whose diagonal elements represent the local mode frequencies of the individual amide I vibrational modes and off-diagonal elements contain the couplings between these vibrational modes. There are two factors that contribute to the local mode frequency. The first is the dependence of the local mode frequency on its environment. In general, electrostatic forces from the environment red shift the vibrational frequency from the gas-phase value, with larger shifts occurring for more polar environments. However, the precise frequency shift is set by the instantaneous structure of the microenvironment. Moreover, hydrogen bonding also contributes, which generally causes an additional red shift. To simulate these two factors, several methods have been developed. We utilize a correlation has been established between the C and N atoms of the amide I mode with the electrostatic field of the environment (including the protein, except for the nearest amide atoms (C, O, N, H) and the two nearest α-carbons). The correlation is given by:

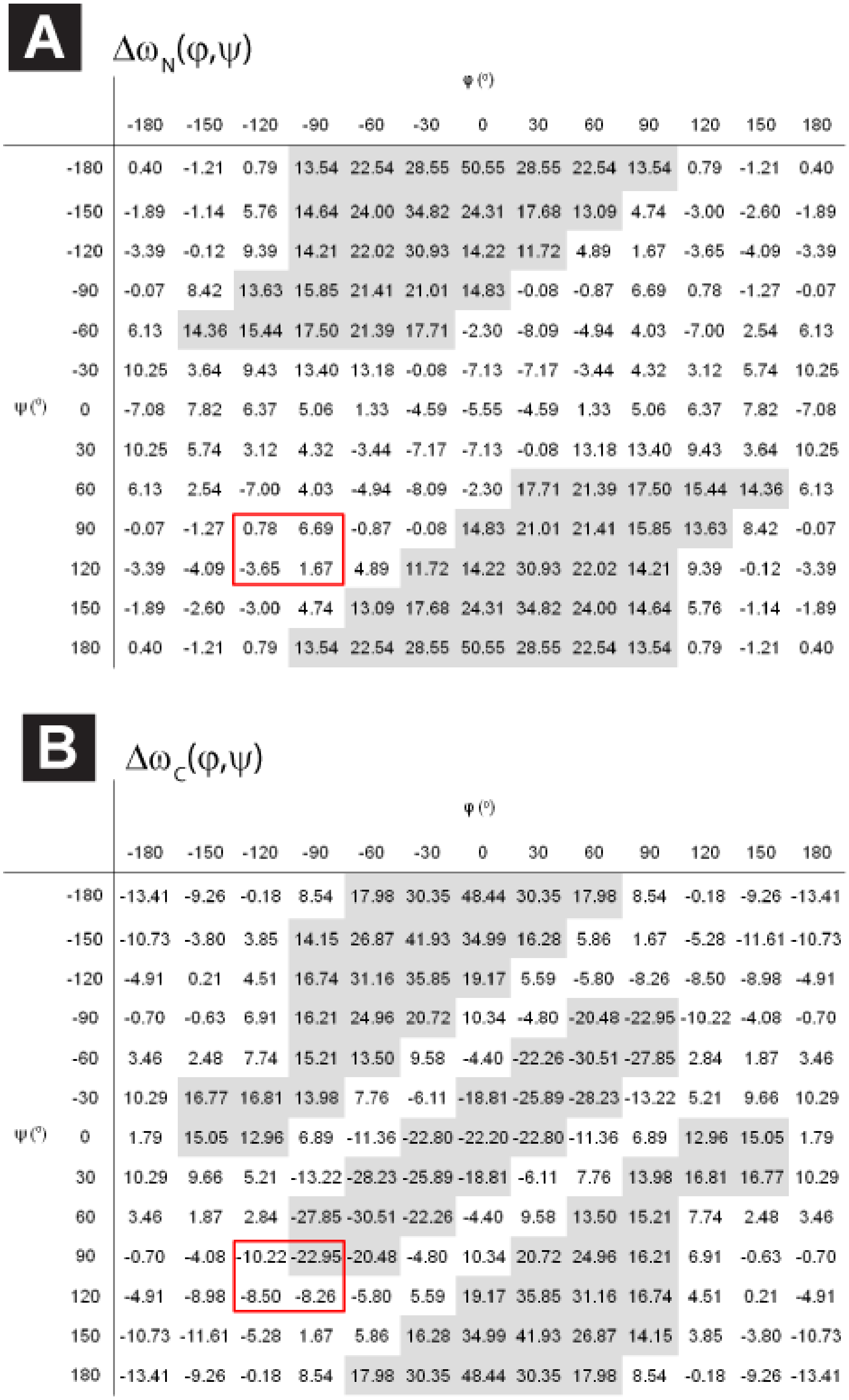
$$\omega_i = 1684 + 7729E_{Ci} - 3576E_{Ni}, \qquad \text{Eq. (1)}$$

Where E_{Ci} and E_{Ni} are the projections of the electrostatic field onto the C and N atoms in the C=O bond direction. This correlation has been obtained by empirically fitting the infrared spectra for the model compound, N-methylactamide, in a variety of solvents.[13] This model has been shown to reproduce the IR spectra of model peptides with various secondary structures, including the rat and human amylin peptides, and the isotope-labeled 2DIR spectra of human amylin fibril well.[4]

The second factor that contributes to the diagonal frequencies is the character of the local mode as a function of the dihedral angles.[70,71] For a particular amide group, the effect of the adjacent amide groups on its local frequency is quantum mechanical in nature and cannot be described by simple electrostatics. We have utilized the nearest neighbor frequency shift maps developed from gas phase ab initio calculations to account for such effects.[62] Shown in Fig. 4.4 are the frequency shifts for these effects determined from gas phase ab initio calculations. Fig. 4.4A is the map for the frequency shift as a function of the dihedral angles to the N-terminal side of the amide I mode being calculated and Fig. 4.4B is the map for the C-terminal side. The two values are added, which in turn are added to the environmental frequency shift calculated from Eq. 1. The total sum is the final value used for the diagonal element in the exciton Hamiltonian for that particular amide I mode.

The third set of parameters in modeling the Hamiltonian is the off-diagonal elements for the couplings. Models for the vibrational couplings have been studied for decades. These include transition dipole coupling, transition charge coupling, and transition dipole density coupling models.[7,21,61,62,72,73] In general, these three coupling models represent three different levels of approximation to the coupling caused by electrostatics. Transition dipole coupling was the first coupling model proposed to explain the dependence of infrared spectra on peptide conformation.[7] At close distances, it is not accurate, and in some cases electrostatics may breaks down, such as for nearest neighbors. In these situations, one can use ab initio calculations to compute the coupling constants.[62] For our simulations, presented below, we have used an ab initio derived map for nearest neighbor couplings.

The final piece of necessary information is the inherent frequency shift caused by the $^{13}C{=}^{18}O$ isotope label itself. Surprisingly, this quantity is not well established. A back-of-the-envelope calculation, assuming that the amide I mode solely consists of C and O motions, puts the frequency shift at -77 cm^{-1}. Electronic structure calculations on a model compound predict -



A $\Delta\omega_N(\varphi,\psi)$

φ (°)

		-180	-150	-120	-90	-60	-30	0	30	60	90	120	150	180
	-180	0.40	-1.21	0.79	13.54	22.54	28.55	50.55	28.55	22.54	13.54	0.79	-1.21	0.40
	-150	-1.89	-1.14	5.76	14.64	24.00	34.82	24.31	17.68	13.09	4.74	-3.00	-2.60	-1.89
	-120	-3.39	-0.12	9.39	14.21	22.02	30.93	14.22	11.72	4.89	1.67	-3.65	-4.09	-3.39
	-90	-0.07	8.42	13.63	15.85	21.41	21.01	14.83	-0.08	-0.87	6.69	0.78	-1.27	-0.07
	-60	6.13	14.36	15.44	17.50	21.39	17.71	-2.30	-8.09	-4.94	4.03	-7.00	2.54	6.13
	-30	10.25	3.64	9.43	13.40	13.18	-0.08	-7.13	-7.17	-3.44	4.32	3.12	5.74	10.25
ψ (°)	0	-7.08	7.82	6.37	5.06	1.33	-4.59	-5.55	-4.59	1.33	5.06	6.37	7.82	-7.08
	30	10.25	5.74	3.12	4.32	-3.44	-7.17	-7.13	-0.08	13.18	13.40	9.43	3.64	10.25
	60	6.13	2.54	-7.00	4.03	-4.94	-8.09	-2.30	17.71	21.39	17.50	15.44	14.36	6.13
	90	-0.07	-1.27	0.78	6.69	-0.87	-0.08	14.83	21.01	21.41	15.85	13.63	8.42	-0.07
	120	-3.39	-4.09	-3.65	1.67	4.89	11.72	14.22	30.93	22.02	14.21	9.39	-0.12	-3.39
	150	-1.89	-2.60	-3.00	4.74	13.09	17.68	24.31	34.82	24.00	14.64	5.76	-1.14	-1.89
	180	0.40	-1.21	0.79	13.54	22.54	28.55	50.55	28.55	22.54	13.54	0.79	-1.21	0.40

B $\Delta\omega_C(\varphi,\psi)$

φ (°)

		-180	-150	-120	-90	-60	-30	0	30	60	90	120	150	180
	-180	-13.41	-9.26	-0.18	8.54	17.98	30.35	48.44	30.35	17.98	8.54	-0.18	-9.26	-13.41
	-150	-10.73	-3.80	3.85	14.15	26.87	41.93	34.99	16.28	5.86	1.67	-5.28	-11.61	-10.73
	-120	-4.91	0.21	4.51	16.74	31.16	35.85	19.17	5.59	-5.80	-8.26	-8.50	-8.98	-4.91
	-90	-0.70	-0.63	6.91	16.21	24.96	20.72	10.34	-4.80	-20.48	-22.95	-10.22	-4.08	-0.70
	-60	3.46	2.48	7.74	15.21	13.50	9.58	-4.40	-22.26	-30.51	-27.85	2.84	1.87	3.46
	-30	10.29	16.77	16.81	13.98	7.76	-6.11	-18.81	-25.89	-28.23	-13.22	5.21	9.66	10.29
ψ (°)	0	1.79	15.05	12.96	6.89	-11.36	-22.80	-22.20	-22.80	-11.36	6.89	12.96	15.05	1.79
	30	10.29	9.66	5.21	-13.22	-28.23	-25.89	-18.81	-6.11	7.76	13.98	16.81	16.77	10.29
	60	3.46	1.87	2.84	-27.85	-30.51	-22.26	-4.40	9.58	13.50	15.21	7.74	2.48	3.46
	90	-0.70	-4.08	-10.22	-22.95	-20.48	-4.80	10.34	20.72	24.96	16.21	6.91	-0.63	-0.70
	120	-4.91	-8.98	-8.50	-8.26	-5.80	5.59	19.17	35.85	31.16	16.74	4.51	0.21	-4.91
	150	-10.73	-11.61	-5.28	1.67	5.86	16.28	34.99	41.93	26.87	14.15	3.85	-3.80	-10.73
	180	-13.41	-9.26	-0.18	8.54	17.98	30.35	48.44	30.35	17.98	8.54	-0.18	-9.26	-13.41

Figure 4.4 - Frequency shift dependency on dihedral angle about the N atom (a) and C atom (b) in the peptide bond.

73 cm^{-1}. Experimental values reported in the literature range from -75 to -59.6 cm^{-1}.[4,30,74,75] This range of values is probably due to two factors: (1) the use of model compounds that do not accurately represent an amide I mode in a peptide and (2) the frequency shift is mistakenly referenced to an absorption band that is itself shifted due to excitonic coupling. We have estimated the $^{13}C^{18}O$ shift using the M2 polypeptide, which has been synthesized on separate occasions with a ^{13}C and $^{13}C^{18}O$ label at Ala29 and Ala30.[76] The ^{13}C frequency of Ala29 was measured to be 1619 cm^{-1} while the $^{13}C^{18}O$ is 1597.5 cm^{-1}. Thus, the frequency shift due to ^{18}O labeling is 21.5 cm^{-1}. The frequency shift caused by ^{13}C coupling alone is 40 cm^{-1}, which is the most consistent value reported in the literature, and especially well-established by recent uniform ^{13}C protein expression studies.[77] Thus, the total $^{13}C^{18}O$ shift is -61.5 cm^{-1}, which agrees with a similar analysis of Ala30. This number is in good agreement with the -59.6 cm^{-1} shift measured for a model compound that closely approximates an amide I mode. To our knowledge, the M2 polypeptide is the only system in which both a ^{13}C and a $^{13}C^{18}O$ frequency have been reported for the same residue on the same peptide under identical conditions.

Shown in Fig. 4.5 are the diagonal slices through the computed 2D IR spectra for each of the isotope label pairs. These spectra are calculated using the molecular dynamics simulations and the three parameters discussed above along with mixed quantum/classical lineshape theory. Details of the simulation methods are given in the Appendix 2 and elsewhere.[13] For each pair of labels, spectra are shown for each individual isotope label, the sum of the individual isotope labels (by zeroing the off-diagonal matrix elements), and the spectrum resulted from full coupling. In these spectra, couplings other than nearest neighbor were calculated using transition dipole coupling. The corresponding coupling strengths are given in Table 4.1.

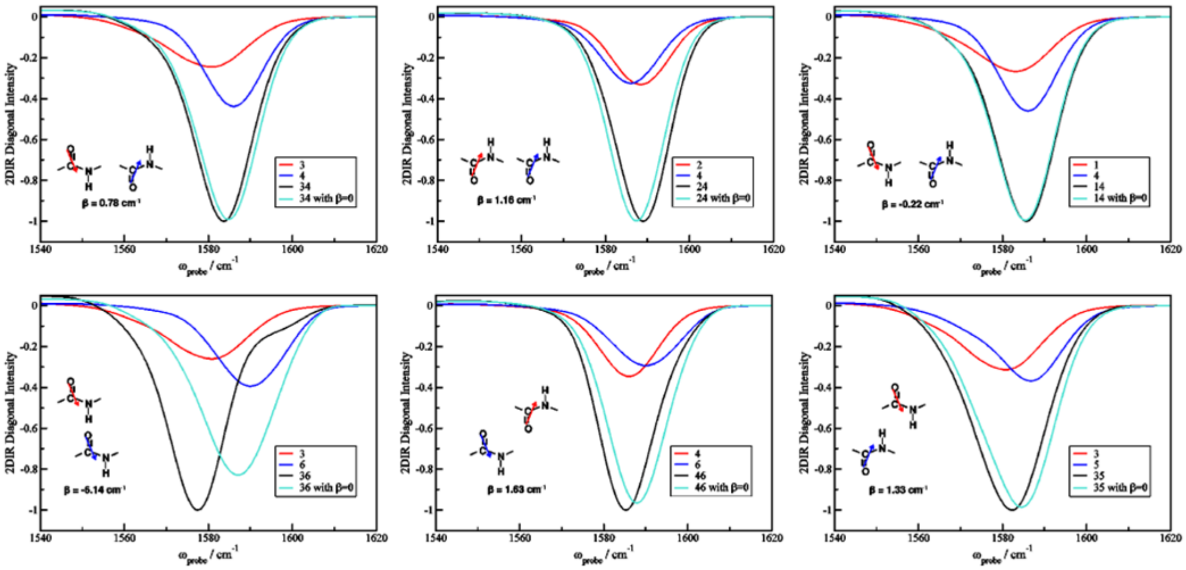


Figure 4.5 - Calculated 2DIR diagonal spectra for each isotope label without coupling (red and blue), the sum of the pair’s intensity without coupling (cyan), and the sum with coupling (black).

Table 4.1. Coupling calculated for each isotope label pair using TDC and TCC coupling models.

Isotope Label	**TDC Torii 92**	**TDC Torii 98**	**TCC Hamm**	**TCC Jansen**
2,4	2.18 ± 0.49	1.16 ± 0.24	1.52 ± 0.38	1.08 ± 0.55
1,4	-0.36 ± 0.15	-0.22 ± 0.09	-0.36 ± 0.13	-0.10 ± 0.15
3,6	-10.80 ± 1.38	-6.14 ± 0.73	-8.82 ± 1.12	-7.03 ± 0.82
4,6	-0.55 ± 3.68	1.63 ± 1.58	1.96 ± 1.59	1.10 ± 2.03
3,5	2.00 ± 0.24	1.33 ± 0.22	2.05 ± 0.45	1.56 ± 0.33

From these simulations, one ascertains the following. First, amide I modes whose carbonyl group point away from the sheets have lower frequencies and broader linewidths. The lower frequencies are a result of the electrostatic fields of the water being larger than the peptide backbone. The larger bandwidth is because the water is more disordered than the backbone. Second, the summation of the individual spectra closely resembles the coupled mode spectra, except for the (3,6) pair. The similarity is a consequence of the small couplings strengths (<2 cm^{-1}) and the large differences in local mode frequencies for most pairs. Third, none of the pairs have a large enough coupling strength to significantly alter the frequency distribution. Not even the (3,6) pair spans a frequency range larger than the individual oscillators. Fourth, the major effect of the coupling is to redistribute the oscillator strength. The most obvious redistribution is for the (3,6) pair, which causes the spectrum to maximize at the lowest frequency even though the lower frequency oscillator has a much weaker intensity than its high frequency counterpart. All 6 pairs studied here exhibit a redistribution of oscillator strength either to lower or higher frequencies, due to the sign of the coupling constant and the relative angles of the labeled carbonyls. These observations provide a basis for us to intuitively understand the infrared spectra of β-sheets.

4.4.4 Testing the Coupling Models

With experimental and simulated results in hand, we compare the two to test the accuracy of the structural/spectra relationship and pinpoint places for which the models might be improved. Shown in Fig. 4.6 are the experimental values of the maximum frequency and linewidth for the isotope labeled region as extracted from the fits to the spectra described above. In a similar manner, we have plotted the peak frequency and linewidth taken from the

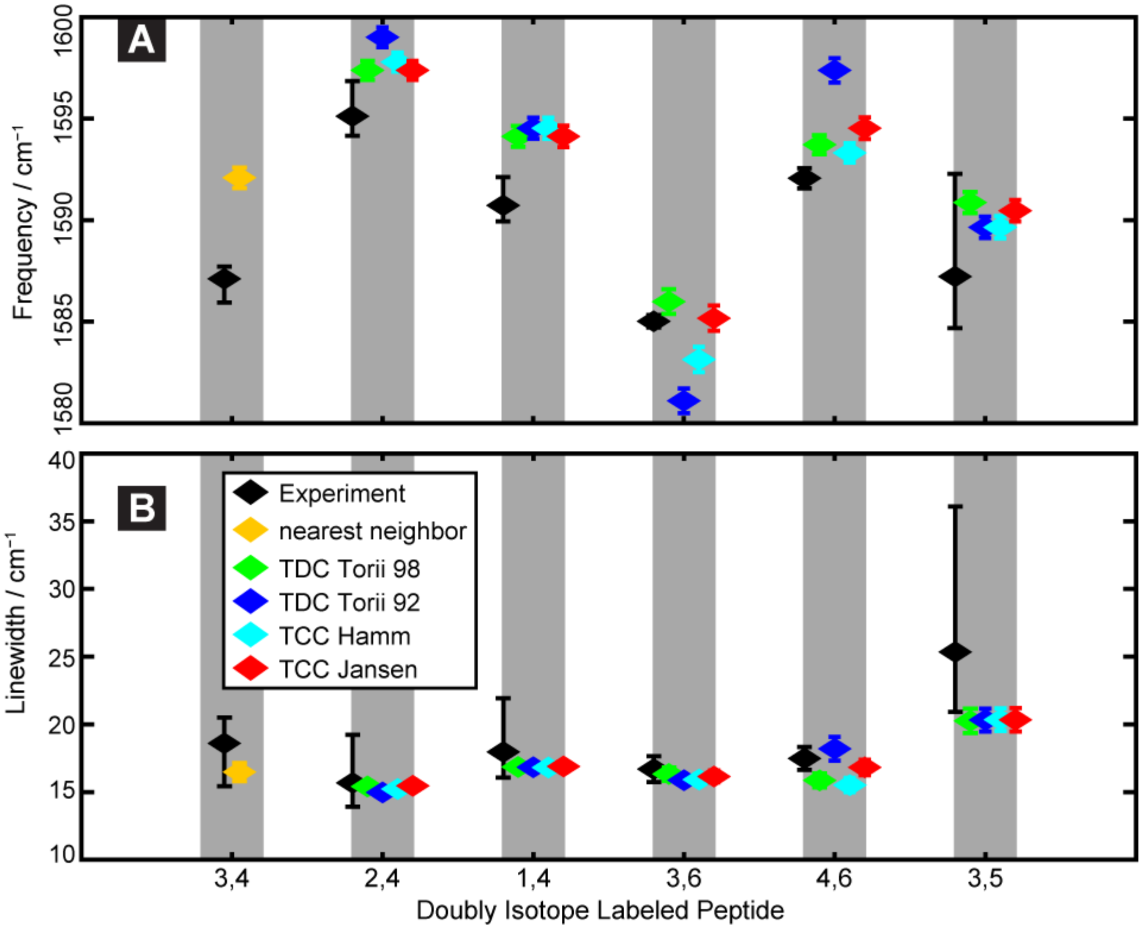


Figure 4.6 - Comparison of experimental and calculated 2D IR (a) frequency and (b) linewidths.

simulations. The simulated spectra have been calculated using two different transition dipole coupling (TDC) models and two transition charge coupling (TCC) models.[21,61,62,72] TDC, first proposed by Krimm and coworkers, is by far the simplest model that describes the interactions between non-adjacent amide groups.[7] It assumes that couplings between these chromophores are electrostatic in nature and can be represented by the interaction of the transition dipoles. The two TDC models differ in the location, orientation and magnitude of the transition dipole relative to the C, O and N atoms that comprise the amide I mode.[62,72] The model listed as TDC Torii92, refers to the parameters developed by Torrii and Tasumi in 1992.[72] The TDC Torii98 model was developed by Torii in 1998 through ab initio molecular orbital (MO) calculations.[62] TCC models have been proposed in order to improve upon TDC by including the higher-order multipoles. The first TCC model was developed by Hamm and coworkers through DFT calculations, and is termed as TCC Hamm in the following.[21] In a similar manner, Jansen and coworkers developed a TCC model later, which we denote as TCC Jansen.[61] All four of these models have been utilized extensively in the literature.[16,78-85] As discussed above, the nearest neighbor couplings are calculated from an ab initio derived nearest-neighbor coupling map.[62] Since residue pair (3,4) are nearest neighbors, we cannot use the transition charge residue for this pair, but only the TDC model.

The simulated linewidths reproduce the experiment very well. They predict the correct linewidth to within a few wavenumbers for all measured pairs. This comparison indicates that the correlation between frequency and electrostatic field (Eq. 1) is quite accurate. The largest discrepancy is for the (3,5) pair in which both residues face outwards. As a result, this pair has the largest linewidth, the weakest intensity and the experimental spectrum is the most difficult to fit. The coupling model does not have a large influence on the vibrational linewidth, because the couplings are not large enough to significantly alter the frequency distribution, as discussed above. Thus, we conclude that the correlation between structure and infrared linewidths is accurate enough that one can confidently assign amide I bands that have >20 cm^{-1} linewidth to residues that are solvent exposed. Narrower linewidths indicate that the amide I backbone is partially solvent protected. Earlier work on the ovispirin polypeptide, which was partially exposed on a membrane bilayer, supports this conclusion.[23]

A comparison between the experimental and simulated frequencies is shown in Fig. 4.6A. We observe that all coupling models produce the correct frequency trends; the relative frequency trends are correctly predicted in all cases. However, the simulated frequencies are ~2 cm^{-1} higher than observed experimentally, except for the (3,6) pair. Thus, as the model stands, one can quite accurately simulate the 2D IR spectra from a molecular dynamics simulation and thereby compare structural and environmental predictions with experiment. The simulations are accurate enough that simulated structures can be tested against experiment so long as the structures have frequency differences larger than 2 cm^{-1}. Structural comparisons using multiple pairs of labels will be even more accurate, since relative frequency comparisons can be made. Structural comparisons based on linewidths are also more accurate, since the correlation between experiment and theory for 2D linewidths is near quantitative (Fig. 4.6B).

4.4.5 Potential Improvements to the Structure/Spectra Modeling

The comparison of experiment to simulations above points to a likely place in which the models may be improved. As outlined above, there are four quantities that go into determining the measured frequency: the electrostatic field, the coupling strength, the diagonal frequency dependence on the dihedral angles, and the $^{13}C^{18}O$ isotopic labels themselves. It is difficult to test these quantities individually, but because we have six sets of isolated residues in a structurally well-defined system, we can make a reasonable hypothesis for the origin of the two discrepancies of the simulated frequencies discussed above. First, it is unlikely that the discrepancies are caused by a faulty molecular structure, because both the 100 ns trajectory and replica exchange simulations agree with the NMR derived structure. Second, it is unlikely that the coupling models are at fault, because all four of the models (the two transition dipole and the two transition charge) predict the same frequencies to within a few wavenumbers (except for the (3,6) pair, which we discuss below). Third, the diagonal frequencies derived from the electrostatic field appear to be quite accurate, as discussed above. The remaining quantities are the diagonal frequency dependence on the dihedral angles and the shift caused by isotopic labeling.

Shown in Fig. 4.4 are color coded maps for the dihedral dependence of the diagonal frequency. On top of these maps are plotted the range of dihedral angles for the isotope labeled residues in this study. All six of the labeled residues lie in a narrow range of dihedral angles because they all residue in the well-ordered region of the parallel β-sheet. A perfect β-sheet has dihedral angles of phi=-119° and psi=113°. In our macrocycle, the dihedral angles are, on average, phi=-110.0+/-22.5° and psi=126.8+/-22.9° from NMR structures, and phi=-101.0+/-17.4° and psi=113.6+/-18.2° from MD data. For the C-terminal map, this range of dihedral angles lies in a portion of the map that is very steep; dropping from -8.5 to -22.95 cm^{-1} over 30 deg. change in dihedral angles. In this range, the diagonal frequency is predicted to vary 0.5 cm^{-1} per deg. change in phi or psi angle, assuming a linear interpolation between points. This steep slope occurs because at psi=-90 and phi=90, the amide I bond is deformed from what would occur in a small molecule, for example. Of course, this region is not usually sampled in standard protein conformations. Nonetheless, because the β-sheet dihedral angles fall on this steep slope, it is influencing the predicted shifts. The N-terminal map has a much smaller frequency dependence on the dihedral angles. Therefore, one reason for the 2 cm^{-1} discrepancy in the simulated frequencies as compared to the experiment (Fig. 4.6A) may be due linear interpolating. Linear interpolation results in a 4 cm^{-1} frequency shift occurring over only a 8 deg. change in dihedral angle. Thus, we think that the C-terminal map needs more calculated points in this range because it is not a linear function. We suspect that the drop in frequency occurs over a small range of dihedral angles. A similar inspection of the maps with regards to the typical dihedral angles seen for α-helices, point to the N-terminal map as potentially problematic. This hypothesis can be tested by simply calculating more points to make more finely gridded maps.

It would also be useful to have more than the M2 sample from which the $^{13}C^{18}O$ frequency shift is deduced.[76] This frequency shift is necessary for a quantitative comparison of experiment and simulations. We believe it is best obtained by comparison of polypeptides with both ^{13}C and $^{13}C^{18}O$ labels under identical conditions.

The last point we make is in regards to the (3,6) pair frequency. This interstrand coupling is the largest off-diagonal element in the Hamiltonian, as has been previously discussed. Thus, this pair is the most sensitive to the coupling model. From the 4 models tested, it appears that the TDC Torii98 coupling model best reproduces the experimental frequency. It is comparable to the other models for all other pairs. The parameters in TDC Torii98 were optimized so that the coupling constants between the second nearest amide groups reproduce the values obtained by ab initio MO calculations on glycine tripeptide. Although the TDC model is oversimplified, such an optimization procedure might effectively take into account the effect due to higher-order multipoles.

4.4.6 Better Understanding Protein β-sheet, Amyloid Fibers and Peptide Assemblies

There exist many infrared and 2D IR experiments on β-sheet containing proteins and many simulation studies.[82,86-90] What we learn in this study is the importance of the diagonal frequencies on the interpretation of the spectra. Our macrocycle consists of equal amounts of solvent exposed and protected amide I bands. We learned that the frequency shift caused by the solvent is one of the most important factors in determining the Hamiltonian. Thus, in β-sheet containing proteins, accurately predicting solvent frequency shifts for solvent exposed residues is critical. Fortunately, it appears that the empirical model used here is quite accurate – accurate frequencies were obtained even in the presence of various side chains, solvent exposure, and nearby charges.

In amyloid and peptide assemblies in which most or all of the peptides experience similar amounts of solvent exposure, accurate couplings are important, since the diagonal disorder is minimized and the off-diagonal elements of the Hamiltonian will take on more significant. In these cases, the couplings will redistribute the transition dipole intensities according to the packing. In previous work, we have published isotope labeling strategies for probing amyloid and peptide assembly structures.[4,87,91] In this work, we find that the interstrand in-register coupling is smaller than previously predicted. According to the frequency shifts of the unlabeled amide I modes (Fig. 4.2A and Table 4.2), there is significant exciton delocalization along the strands. We hypothesize that this intrastrand coupling may be the reason that not all isotope labeled residues in the amyloid fibers formed from the peptide amylin have the same intensities.

Table 4.2. Frequency, linewidth, and lifetime of isotope labeled amide I peak from 2D IR spectra.

Isotope Labeled	**Unlabeled Frequency /**	**$^{13}C^{18}O$ Frequency /**	**$^{13}C^{18}O$ Linewidth**	**$^{13}C^{18}O$ Lifetime / fs**
3,4	1640.8	1587.1	18.6	1075
2,4	1642.5	1595.1	15.7	493
1,4	1643.6	1590.7	18.0	385
3,6	1637.0	1585.0	16.7	528
4,6	1639.1	1592.1	17.5	413
3,5	1637.7	1587.2	25.3	365
Unlabeled	1634.2	-	-	-

Regardless, when interpreting the infrared spectra of order protein arrays, one should consider significant delocalization along the strands in additional to across them.

4.5 Conclusion

Precisely understanding the origins of parallel β-sheets infrared spectra is very important, since infrared spectroscopy is playing an increasingly larger role in probing the structures of amyloid fibers and peptide assemblies. There exist many measured and simulated spectra of anti-parallel β-sheet containing proteins and protein assemblies.[38,82,88] We believe that this report provides the most comprehensive study into the origins of the couplings in parallel β-sheets. The detailed comparison between experiment and theory strengthens our confidence in the ability to quantitatively calculate solvent dependent frequency shifts and linewidths, but also points to a deficiency in modeling. The deficiency is not a problem of concept or method, but only in the course grained maps that currently exist and the uncertainty in the frequency shift caused by isotope labeling. These problems are straightforwardly remedied. We do not attempt to make finer maps here, because, while straightforward, it is computationally time consuming, as is the synthesis of additional peptides. Nonetheless, it is clear that near-quantitative agreement can be obtained between simulation and experiment. This study establishes that one can distinguish between structures or environments that differ by only a few wavenumbers. The results point to a better understanding and quantification of the coupling models utilized over the past 50 years to interpret the infrared spectra of β-sheet containing proteins. This improved understand will lead to improved structural and dynamical characterization of amyloid forming polypeptides.

4.6 References

(1) Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. *Analyst (Cambridge, U. K.)* **2012**, *137*, 1793.

(2) Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30*, 95.

(3) Manor, J.; Mukherjee, P.; Lin, Y. S.; Leonov, H.; Skinner, J. L.; Zanni, M. T.; Arkin, I. T. *Structure* **2009**, *17*, 247.

(4) Wang, L.; Middleton, C. T.; Singh, S.; Reddy, A. S.; Woys, A. M.; Strasfeld, D. B.; Marek, P.; Raleigh, D. P.; de, P. J. J.; Zanni, M. T.; Skinner, J. L. *J. Am. Chem. Soc.* **2011**, *133*, 16062.

(5) Arora, A.; Tamm, L. K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 540.

(6) Miyazawa, T. *J. Chem. Phys.* **1960**, *32*, 1647.

(7) Moore, W. H.; Krimm, S. *Proc Natl Acad Sci U S A* **1975**, *72*, 4933.

(8) Choi, J.-H.; Cho, M. *Chem. Phys.* **2009**, *361*, 168.

(9) Maekawa, H.; Ge, N.-H. *J. Phys. Chem. B* **2010**, *114*, 1434.

(10) Bloem, R.; Dijkstra, A. G.; Jansen, T. l. C.; Knoester, J. *J Chem Phys* **2008**, *129*, 055101.

(11) Hayashi, T.; Mukamel, S. *J. Mol. Liq.* **2008**, *141*, 149.

(12) Hayashi, T.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. A* **2005**, *109*, 9747.

(13) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2011**, *115*, 3713.

(14) Paschek, D.; Puehse, M.; Perez-Goicochea, A.; Gnanakaran, S.; Garcia, A. E.; Winter, R.; Geiger, A. *ChemPhysChem* **2008**, *9*, 2742.

(15) Viswanathan, R.; Dannenberg, J. J. *J. Phys. Chem. B* **2008**, *112*, 5199.

(16) Bagchi, S.; Falvo, C.; Mukamel, S.; Hochstrasser, R. M. *J. Phys. Chem. B* **2009**, *113*, 11260.

(17) Choi, J.-H.; Cho, M. *Chem. Phys.* **2009**, *361*, 168.

(18) Schweitzer-Stenner, R.; Measey, T. J. *Spectroscopy (Amsterdam, Neth.)* **2010**, *24*, 25.

(19) Dijkstra, A. G.; Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2011**, *115*, 5392.

(20) la Cour Jansen, T.; Knoester, J. *J Chem Phys* **2006**, *124*, 044502.

(21) Hamm, P.; Woutersen, S. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 985.

(22) Hayashi, T.; Mukamel, S. *J. Mol. Liq.* **2008**, *141*, 149.

(23) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de, P. J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832.

(24) Wang, J.; Zhuang, W.; Mukamel, S.; Hochstrasser, R. *The Journal of Physical Chemistry B* **2007**, *112*, 5930.

(25) Barth, A. *Prog Biophys Mol Biol* **2000**, *74*, 141.

(26) Torres, J.; Adams, P. D.; Arkin, I. T. *J. Mol. Biol.* **2000**, *300*, 677.

(27) Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T. *Proc Natl Acad Sci U S A* **2006**, *103*, 3528.

(28) Middleton, C. T.; Marek, P.; Cao, P.; Chiu, C.-c.; Singh, S.; Woys, A. M.; de, P. J. J.; Raleigh, D. P.; Zanni, M. T. *Nat. Chem.,* Ahead of Print.

(29) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R. A. G. D.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346.

(30) Fang, C.; Wang, J.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B.; Decatur, S. M.; Hochstrasser, R. M. *Chem. Phys. Lett.* **2003**, *382*, 586.

(31) Fang, C.; Wang, J.; Kim, Y. S.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B., III; Decatur, S. M.; Hochstrasser, R. M. *J. Phys. Chem. B* **2004**, *108*, 10415.

(32) Lakhani, A.; Roy, A.; De, P. M.; Nakaema, M.; Formaggio, F.; Toniolo, C.; Keiderling, T. A. *J Phys Chem B* **2011**, *115*, 6252.

(33) Huang, R.; Setnicka, V.; Etienne, M. A.; Kim, J.; Kubelka, J.; Hammer, R. P.; Keiderling, T. A. *J. Am. Chem. Soc.* **2007**, *129*, 13592.

(34) Huang, R.; Wu, L.; McElheny, D.; Bour, P.; Roy, A.; Keiderling, T. A. *J. Phys. Chem. B* **2009**, *113*, 5661.

(35) Hahn, S.; Ham, S.; Cho, M. *The Journal of Physical Chemistry B* **2005**, *109*, 11789.

(36) Freire, F.; Almeida, A. M.; Fisk, J. D.; Steinkruger, J. D.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2011**, *50*, 8735.

(37) Fisk, J. D.; Powell, D. R.; Gellman, S. H. *J. Am. Chem. Soc.* **2000**, *122*, 5443.

(38) Tsutsumi, M.; Otaki, J. M. *Journal of Chemical Information and Modeling* **2011**, *51*, 1457.

(39) Tycko, R.; Savtchenko, R.; Ostapchenko, V. G.; Makarava, N.; Baskakov, I. V. *Biochemistry* **2010**, *49*, 9488.

(40) Kizaki, H.; Hata, Y.; Watanabe, K.; Katsube, Y.; Suzuki, Y. *J. Biochem.* **1993**, *113*, 646.

(41) Enkhbayar, P.; Kamiya, M.; Osaki, M.; Matsumoto, T.; Matsushima, N. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 394.

(42) Yan, C.; Mackay, M. E.; Czymmek, K.; Nagarkar, R. P.; Schneider, J. P.; Pochan, D. J. *Langmuir* **2012**, *28*, 6076.

(43) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* **2010**, *52*, 12.

(44) Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG an der ETH Zürich, Zürich, 1996.

(45) Scott, W. R. P.; Huenenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krueger, P.; van, G. W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.

(46) Oostenbrink, C.; Villa, A.; Mark, A. E.; van, G. W. F. *J. Comput. Chem.* **2004**, *25*, 1656.

(47) Berendsen, H. J. C.; Postma, J. P. M.; Van, G. W. F.; Hermans, J. *Jerusalem Symp. Quantum Chem. Biochem.* **1981**, *14*, 331.

(48) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.

(49) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

(50) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695.

(51) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

(52) Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089.

(53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of Chemical Physics* **1995**, *103*, 8577.

(54) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43.

(55) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J Comput Chem* **2005**, *26*, 1701.

(56) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *Journal of Chemical Theory and Computation* **2008**, *4*, 435.

(57) Yan, Q.; de Pablo, J. J. *The Journal of Chemical Physics* **1999**, *111*, 9509.

(58) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(59) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Peptide Science* **2001**, *60*, 96.

(60) Rathore, N.; Chopra, M.; de Pablo, J. J. *The Journal of Chemical Physics* **2005**, *122*, 024111.

(61) Jansen, T. l. C.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. *The Journal of Chemical Physics* **2006**, *125*, 044312.

(62) Torii, H.; Tasumi, M. *Journal of Raman Spectroscopy* **1998**, *29*, 81.

(63) Jansen, T. l. C.; Knoester, J. *J Phys Chem B* **2006**, *110*, 22910.

(64) Jansen, T. l. C.; Knoester, J. *Acc. Chem. Res.* **2009**, *42*, 1405.

(65) Jansen, T. l. C.; Auer, B. M.; Yang, M.; Skinner, J. L. *The Journal of Chemical Physics* **2010**, *132*, 224503.

(66) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. *The Journal of Chemical Physics* **2011**, *135*, 234507.

(67) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc Natl Acad Sci U S A* **2009**, *106*, 17751.

(68) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *The Journal of Chemical Physics* **2005**, *123*, 084905.

(69) Hamm, P.; Helbing, J.; Bredenbeck, J. *Annu. Rev. Phys. Chem.* **2008**, *59*, 291.

(70) Wang, J. *J. Phys. Chem. B* **2008**, *112*, 4790.

(71) Choi, J.-H.; Ham, S.; Cho, M. *The Journal of Physical Chemistry B* **2003**, *107*, 9132.

(72) Torii, H.; Tasumi, M. *The Journal of Chemical Physics* **1992**, *96*, 3379.

(73) Krueger, B. P.; Scholes, G. D.; Fleming, G. R. *The Journal of Physical Chemistry B* **1998**, *102*, 5378.

(74) Ham, S.; Cha, S.; Choi, J.-H.; Cho, M. *The Journal of Chemical Physics* **2003**, *119*, 1451.

(75) Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T. *Biopolymers* **2001**, *59*, 396.

(76) Kukol, A.; Adams, P. D.; Rice, L. M.; Brunger, A. T.; Arkin, T. I. *J Mol Biol* **1999**, *286*, 951.

(77) Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 3329.

(78) Remorino, A.; Hochstrasser, R. M. *Acc. Chem. Res.*, Ahead of Print.

(79) Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14237.

(80) Maekawa, H.; Formaggio, F.; Toniolo, C.; Ge, N.-H. *J. Am. Chem. Soc.* **2008**, *130*, 6556.

(81) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2004**, *108*, 16965.

(82) Demirdoeven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.

(83) Karjalainen, E.-L.; Barth, A. *J. Phys. Chem. B* **2012**, *116*, 4448.

(84) Karjalainen, E.-L.; Ravi, H. K.; Barth, A. *J. Phys. Chem. B* **2011**, *115*, 749.

(85) Volkov, V. V.; Chelli, R.; Muniz-Miranda, F.; Righini, R. *J. Phys. Chem. B* **2011**, *115*, 5294.

(86) Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3329.

(87) Strasfeld, D. B.; Ling, Y. L.; Gupta, R.; Raleigh, D. P.; Zanni, M. T. *J. Phys. Chem. B* **2009**, *113*, 15679.

(88) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *J. Chem. Phys.* **2005**, *123*, 084905/1.

(89) DeFlores, L. P.; Ganim, Z.; Nicodemus, R. A.; Tokmakoff, A. *J. Am. Chem. Soc.* **2009**, *131*, 3385.

(90) Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 612.

(91) Marek, P.; Woys, A. M.; Sutton, K.; Zanni, M. T.; Raleigh, D. P. *Org. Lett.* **2010**, *12*, 4848.

CHAPTER 5

Synthesis of Cysteine-Selective Cyclopentadienyl Rhenium Carbonyl Labels for FTIR and 2D IR Probes of Protein Structure.[*]

5.1 Abstract

Vibrational modes are generally sensitive to hydrogen bonding, solvation, and electrostatics, which make them good environmental probes. In this paper, we report the synthesis and 2D IR spectra for a series of non-natural infrared probes that attach to proteins via cysteine amino acids. The compounds are tricarbonyl(η^5-cyclopentadienyl)rhenium(I) to which an alkyl chain of either one, three or four carbons is terminated with methanethiosulfonate. Under gentle reaction conditions, ring-substituted $CpRe(CO)_3$ compounds covalently bond with cysteine forming a disulfide bond that is stable to the oxygen and water-rich environment of the protein in analogy to the leaving groups used for nitroxide spin labels in electron pair resonance spectroscopy. To demonstrate the method, we report the 2D IR spectra for this series of labels attached to cysteine mutated α-synuclein, both in aqueous buffer and bound to lipid bilayers. The cyclopentadienyl rhenium tricarbonyl compounds have an extremely strong extinction coefficient and absorb in a background free window of the infrared spectrum that lacks natural protein and water absorption. As a result, it can be used to study proteins at much lower

---

[*] This chapter will be submitted to *Organometallics* in April 2012. It was prepared in collaboration with D. Skoff, S. D. Moran, M. T. Zanni.

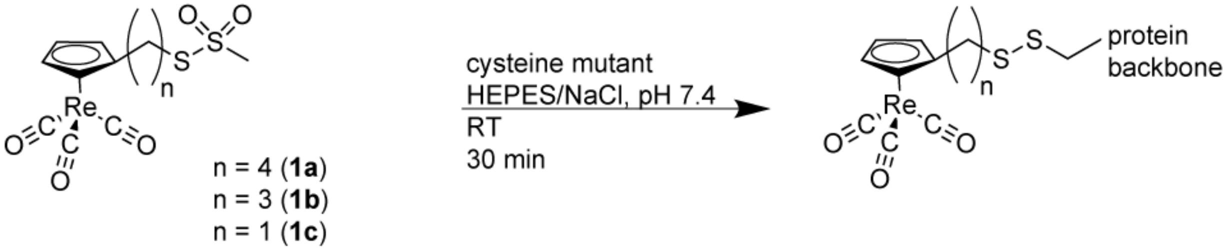
concentrations than other natural and non-natural vibrational probes. These probes will be useful for probing protein structural and environmental changes for a wide range of soluble, aggregating and membrane bound proteins.

5.2 Introduction

Natural and non-natural vibrational chromophores are a good means of probing the electrostatic and hydrogen bonding environment of polypeptides and proteins. Natural probes include $^{13}C{=}^{18}O$ backbone labels to resolve individual amide I modes, C-D isotope substitutions such as in side-chains, and cysteine mutants whose S-H frequency is unlike any other natural amino acid absorption.[1-3] Non-natural probes include nitriles, nitrophenyl, and thiocyanates, whose triple bond character puts them in a window of the infrared spectrum that does not overlap with most protein vibrational modes or with water absorbance.[4-7] These natural and non-natural probes are small enough that they are generally minimally or non-perturbative to the protein structure and so can often be incorporated near the protein active site.[8] However, their incorporation into the system of interest is not always straightforward. For polypeptides, many labels can be incorporated via solid phase synthesis, but for proteins that are too large to be synthesized, other methods are required. For example, non-natural amino acids containing nitrile or azido substituents must be incorporated via semisynthesis or genetic methods.[7,9-13] On the other hand, thiocyano derivatives of cysteine may be produced post-translationally.[6] However, most of these probes are weak absorbers and thus require high protein concentrations on the order of 1-10 mM,[6,7,9-13] which are difficult to achieve in many systems.

To overcome the limitations caused by protein concentration and to ease the effort required to site-specifically label proteins, we have synthesized a non-natural probe that has a very strong infrared extinction coefficient and is straightforwardly attached to the thiol in the cysteine side-chain (Scheme 5.1). Compounds **1a-c** utilize the strong, isolated IR absorption of $CpRe(CO)_3$ with the Cp ring substituted with $–(CH_2)_nSSO_2CH_3$, where n is 1, 3, or 4. The methanethiosulfonate reacts with the cysteine side-chain to form a disulfide bond, resulting in a covalent attachment that is stable to the oxygen and water-rich environment of the protein. As a result, our label can be incorporated into proteins using standard site-specific mutagenesis protocols to generate a single cysteine residue. Other metal-carbonyl probes have been previously synthesized using other residue-specific functional groups, such as *N*-hydroxysuccinimide esters, imidoesters, pyrylium ions, and Fischer-type metallocarbene complexes that attach to primary amines like those on lysine residues.[14] However, those compounds are not generally useful as site-specific probes because most soluble proteins contain several lysine residues as well as an amine at the N-terminus. Another functional group employed for protein labeling is the maleimide group, which, at neutral pH and room temperature, is cysteine selective.[15,16] In fact, $CpRe(CO)_3$ has been attached to proteins via a maleimide linker.[15] However, the maleimide group is bulky. Our linker is an analog of a commonly used molecule for site-specific spin labeling in EPR spectroscopy.[17,18] Its methanethiosulfonate functional group is cysteine-specific at neutral pH and room temperature.[17,18] It is preferable to maleimide, because it results in a flexible, straight-chain linker. By using the same $–(CH_2)_nSSO_2CH_3$ linker and functional group as produce near quantitative reactivity toward cysteine residues under mild conditions.

Scheme 5.1.


O
O
S
S
n
Re
O≡C
C≡O
C
O
n = 4 (**1a**)
n = 3 (**1b**)
n = 1 (**1c**)
cysteine mutant
HEPES/NaCl, pH 7.4
RT
30 min
S
S
protein
backbone
n
Re
O≡C
C≡O
C
O

In this paper, we demonstrate our approach by attaching [η^5-$(C_5H_4)(CH_2)_4SSO_2CH_3$]Re(CO)$_3$ to a single cysteine mutant of α-synuclein, a synaptic protein that is natively disordered in solution but adopts an α-helical conformation when associated with a phospholipid bilayer. We collect 2D IR spectra of the protein in aqueous buffer and bound to lipid vesicles to illustrate some aspects of the information content of the labels. Rhenium carbonyl vibrational modes have an extinction coefficient of ~4,100 $M^{-1}cm^{-1}$. For comparison, the nitrile group, which is the most commonly used non-natural vibrational probe, has an extinction coefficient of 22-240 $M^{-1}cm^{-1}$.[19] Signal strengths for FTIR spectra scale linearly with the extinction coefficient, while for 2D IR spectra, they scale quadratically. Thus, one can obtain comparable signal-to-noise for the $Re(CO)_3$ compound as compared to a nitrile at ~40 times lower concentrations for FTIR and ~1600 times lower concentration for 2D IR spectra. In the data reported here, the α-synuclein concentration is <1 mM, which is much lower than the concentrations typically used for nitriles.[11] The signal-to-noise of our measurements demonstrate that protein concentrations <100 μM can be studied with this label. Of course, the metal carbonyl labels are larger and structurally more perturbative, but as we show below, will be good probes for surface protein sites in analogy to placements used for EPR spin labels.

5.3 Experimental Section

5.3.1 General Procedures

$CpRe(CO)_3$ was prepared as published[39] and subsequently reacted to give [η^5-$(C_5H_4)CH_2CH_2CH_2CH_2I$]Re(CO)$_3$, [$\eta_5$-$(C_5H_4)CH_2CH_2CH_2I$]Re(CO)$_3$, and [$\eta^5$-$(C^5H^4)CH^2OH$]Re(CO)$_3$ as described by literature methods.[21, 24, 40] Glassware and needles were dried in an oven prior to use. All reactions were performed under nitrogen. CH_2Cl_2 and ACN were purified over alumina.[41] DMF vacuum distilled from CaH_2. Acetone was used as received. All other reagents were used as received. TLC was performed with Silica Gel 60 F254 plates. 40-63 μm SiliaFlash® P60 was used for flash column chromatography.

Silica gel columns were prepared with Silicycle 40-63 Å silica (230-400 mesh) according to the method described by Still, except that the packing was by the slurry method.[42] Solvents were chosen based on TLC.[43]

The ^1H and ^{13}C NMR spectra for $[\eta^5\text{-}(C_5H_4)CH_2I]Re(CO)_3$ were collected on a Bruker AC+ 300 spectrometer in the UW-Madison Chemistry Department. All other ^1H and ^{13}C spectra were acquired with an Avance-500 spectrometer at NMRFAM. The HSQC spectra were used for determining peak assignments and collected on Varian Inova-600 and Inova-500 spectrometers in the UW-Madison Chemistry Department. All spectra were referenced with TMS at 0 ppm. EI mass spectrometry was performed with a Waters (Micromass) AutoSpec® in the UW-Madison Chemistry Department. MALDI-TOF spectra were collected on an Applied Biosystems/MDS SCIEX 4800 MALDI TOF/TOF. FTIR spectra were acquired with a Thermo Scientific Nicolet iS10 FT-IR Spectrometer.

5.3.2 Synthesis of $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$

To $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2I]Re(CO)_3$ (185 mg, 0.357 mmol) was added $NaSSO_2CH_3$ (192 mg, 1.43 mmol) in four molar excess and DMF (8 mL). The reaction was left to stir overnight. After aqueous workup, 174 mg (0.346 mmol) of product $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$ was retrieved in 97% yield. IR (DMSO): ν_{CO} 1915.17 cm^{-1} as, 2014.17 cm^{-1} s. IR (1:1 CCl_4:CH_2Cl_2): ν_{CO} 1926.36 cm^{-1} as, 2022.36 cm^{-1} s.

^{1}H NMR: (499.8 MHz, $CDCl_3$) δ 5.195 (Cp, m, 4H), 3.262 ($-CH_3$, s, 3H), 3.116 ($-CH_2S-$, t, J = 7.3 Hz, 2H), 2.399 ($CpCH_2-$, m, 2H), 1.775 ($C\mathit{H}_2CH_2S-$, p, J = 7.6 Hz, 2H), 1.572 ($CpCH_2C\mathit{H}_2-$, m, 2H). ^{13}C NMR: (125.7 MHz, $CDCl_3$) δ 194.7 (*C*O), 110.6 (*ipso*-Cp), 84.0 (Cp), 83.2 (Cp), 50.9 ($-\mathit{C}H_3$), 36.1 ($-\mathit{C}H_2S-$), 30.7 ($CpCH_2\mathit{C}H_2-$), 29.4 ($-\mathit{C}H_2CH_2S-$), 27.7 ($Cp\mathit{C}H_2-$). EI mass $[M+Na]^+$ meas. 522.9766 m/z (calc. 522.9783 m/z).

5.3.3 Synthesis of $[\eta^5-(C_5H_4)CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$

To $[\eta^5-(C_5H_4)CH_2CH_2CH_2I]Re(CO)_3$ (69.0 mg, 0.137 mmol) was added $NaSSO_2CH_3$ (101 mg, 0.753 mmol) in four molar excess and DMF (8 mL). The reaction was left to stir overnight. After aqueous workup, 24.7 mg (0.0492 mmol) of product $[\eta^5-(C_5H_4)CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$ was retrieved in 36% yield. IR (DMSO): ν_{CO} 1917.16 cm^{-1} as, 2015.57 cm^{-1} s. IR (1:1 CCl_4:CH_2Cl_2): ν_{CO} 1931.00 cm^{-1} as, 2024.48 cm^{-1} s. ^{1}H NMR: (499.8 MHz, $CDCl_3$) δ 5.211 (Cp, m, 4H), 3.272 ($-CH_3$, s, 3H), 3.141 ($-CH_2S-$, t, J = 7.2 Hz, 2H), 2.508 ($CpCH_2-$, m, J(H,H) = 7.8 Hz, 2H), 1.775 ($CH_2C\mathit{H}_2-CH_2-$, p, J = 7.5 Hz, 2H). ^{13}C NMR: (125.7 MHz, $CDCl_3$) δ 194.4 (*C*O), 109.1 (*ipso*-Cp), 84.1 (Cp), 83.4 (Cp), 51.0 ($-\mathit{C}H_3$), 35.7 ($-\mathit{C}H_2S-$), 31.7 ($CpCH_2\mathit{C}H_2-$), 26.998 ($Cp\mathit{C}H_2-$). EI mass $[M+Na]^+$ meas. 508.9605 m/z (calc. 508.9627 m/z).

5.3.4 Synthesis of $[\eta^5-(C_5H_4)CH_2I]Re(CO)_3$

A solution of $[\eta^5-(C_5H_4)CH_2OH]Re(CO)_3$ (396 mg, 1.13 mmol) and NaI (163 mg, 1.21 mmol) in ACN (15 mL) was stirred under a stream of nitrogen as TMS-Cl (137 µL, 1.08 mmol) was added dropwise. The resulting solution was stirred at room temperature for 20 minutes. Ether (15 mL) was added to the reaction mixture to extract the product, which was successively washed with water (15 mL), thiosulfate (10%, 15 mL), and brine (15 mL). The ether extracts were dried with $MgSO_4$. Celite filtration and removal of the solvent afforded 455 mg (0.957 mmol, 84% yield) of $[\eta^5\text{-}(C_5H_4)CH_2I]Re(CO)_3$. ^1H NMR: (300.1 MHz, $CDCl_3$) δ 5.513 (Cp, t, J = 2.2 Hz, 2H), 5.289 (Cp, t, J = 2.2 Hz, 2H), 4.1444 (-C*H*$_2$-, s, 2H). ^{13}C NMR: (75.4 MHz, $CDCl_3$) δ 193.7 (*C*O), 104.0 (*ipso*-Cp), 85.6 (Cp), 84.4 (Cp), 29.9 (-*C*H$_2$S-).

5.3.5 Synthesis of $[\eta^5\text{-}(C_5H_4)CH_2SSO_2CH_3]Re(CO)_3$

$[\eta^5\text{-}(C_5H_4)CH_2I]Re(CO)_3$ (454.9 mg, 0.957) and $NaSSO_2CH_3$ (642 mg, 4.79 mmol) were dissolved in DMF (8 mL) and stirred overnight. An aqueous workup afforded $[\eta^5\text{-}(C_5H_4)CH_2SSO_2CH_3]Re(CO)_3$ (79.0 mg, 0.172 mmol) in 18% yield. IR (DMSO): ν_{CO} 1914.88 cm^{-1} as, 2014.04 cm^{-1} s. IR (1:1 CCl_4:CH_2Cl_2): ν_{CO} 1930.72 cm^{-1} as, 2024.06 cm^{-1} s. ^1H NMR: (499.8 MHz, $CDCl_3$) δ 5.483 (Cp, t, J = 2.2 Hz, 2H), 5.262 (Cp, t, J = 2.2 Hz, 2H), 4.064 (CpC*H*$_2$S-, s, 2H), 3.228 (-CH$_3$, s, 3H). ^{13}C NMR: (125.7 MHz, $CDCl_3$) δ 193.3 (*C*O), 101.6 (*ipso*-Cp), 85.7 (Cp), 84.7 (Cp), 51.5 (-*C*H$_3$), 33.9 (-*C*H$_2$S-). EI mass $[M+Na]^+$ meas. 480.9325 m/z (calc. 480.9314 m/z).

5.3.6 Expression and purification of recombinant V71C α–synuclein mutant

A codon-optimized gene for wild-type human α–synuclein was purchased from Genscript (Piscataway, NJ) and subcloned into a modified pET32a vector (Novagen) containing a Factor Xa site for His$_6$-tag removal, using Bam HI and Eco RI restriction sites. The V71C mutant was generated by linearization of the plasmid via PCR with 5'-phosphorylated mutagenic primers followed by blunt-ended ligation using T4 ligase under standard reaction conditions. DNA sequencing was used to confirm the presence of the V71C mutant. Protein expression and purification were performed as described previously,[44] using a modified cleavage buffer for the Factor Xa reaction (20 mM Tris, 50 mM NaCl, 0.1 M $CaCl_2$, pH 7.5). Prior to reverse phase chromatography purification the protein was incubated in 0.1 mM DL-dithiotheitol at 95^{o}C for 5 minutes to reduce any disulfide bonds. MALDI mass $[M+H]^+$ meas. 14,463.0 m/z (calc. 14,465.1 m/z).

5.3.7 Labeling α–synuclein protein with $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$

V71C (300 µg, 20.1 nmol) was dissolved in 10 mM HEPES, 100mM NaCl buffer (pH 7.4, 1 mL). Preformed protein aggregates were removed by microcentrifugation at 14,000 g for 60 minutes. 148 µg (294 nmol) of Re-label, $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$, was added in acetone (3.4 µL, ~10 eq.). The reaction was shaken for 1 h. Labeled protein was purified by size-exclusion chromatography on an HPLC by flowing the reaction solution over a Tricorn 10/600 column packed with Sephadex G-75 gel filtration medium with 10 mM NH_4CO_3 buffer (pH 7). MALDI mass $[M+H]^+$ meas. 14,884.0 m/z (calc. 14,887.1 m/z).

5.3.8 2D IR spectroscopy of V71C mutant α–synuclein labeled with $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$

For the solution sample, a 65 µg (4.3 nmol) aliquot of V71C was dissolved in 20 mM deuterated phosphate buffer (pH 7.4, 5 µL) buffer, which produced a 24 mOD absorbance at 1924 cm^{-1} and 16 mOD at 2019 cm^{-1} in the FTIR spectrum (56 micron cell path-length). The 3:1 POPC/POPS lipid vesicles were prepared by evaporating under high vacuum, 27 µL POPC and 12 µL POPS (in chloroform, 25 mg/mL). This film was suspended in 20 mM deuterated phosphate buffer (pH 7.4, 5 µL) and sonicated for 30 minutes. This lipid solution was added to the protein sample and mixed.

Samples were placed between CaF_2 windows, and mid-IR pump and probe beams were overlapped on the sample for the collection of 2D IR spectra. The 800 nm output of a 1 kHz Ti:sapphire KMLabs Wyvern regenerative amplifier (1.3 mJ, 45 fs transform limited pulses) is used to pump a home-built optical parametric amplifier (OPA). Here, 800 nm pulses are downconverted into signal and idler pulses (180 μJ) in a type II BBO crystal ($\theta = 28.0°$, 2 mm thick) in two stages with collinear alignment. Mid-IR pulses with FWHM bandwidths of approximately 185 cm^{-1} at 5 μm and pulse energies of >5 μJ are then generated by difference frequency mixing of the signal and idler pulses in a type II $AgGaS_2$ crystal ($\theta = 45.4°$, 1 mm thick). The Mid-IR is then split into pump (97%) and probe (3%) beams, and the pump beam is sent through a pulse shaper to create pulse pairs as described previously.

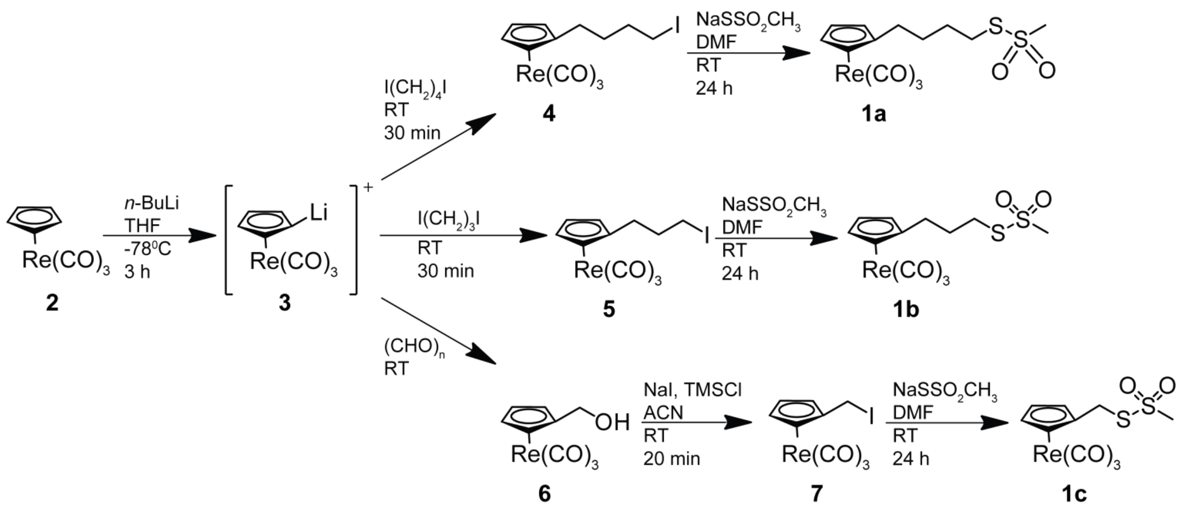
5.4 Results and Discussion

5.4.1 Substitution of $CpRe(CO)_3$ to Produce Iodide Intermediates

The synthesis of the labels is shown in Scheme 5.2. The synthesis of all labels, **1a-c**, began by activating the Cp ring in **2** with n-BuLi in THF at -78^{o}C. The activation occurred specifically at the ring and the ring was substituted singly so long as **2** and *n*-BuLi were present in equal molar portions. This method is reported frequently for **2** with yields varying up to 89% and usually >45% depending on the following reagent.[20-26]

5 was previously synthesized using *n*-BuLi under these conditions and then by subsequent addition of diiodopropane in small excess.[21] A similar method was followed for **4**, except that in this case, diiodobutane is used for alkylation. For both **4** and **5**, the reaction solutions warmed to room temperature and were monitored with TLC. After 30 minutes, the

Scheme 5.2.


Re(CO)3
2
n-BuLi
THF
-78⁰C
3 h
Li
Re(CO)3
3
I(CH2)4I
RT
30 min
I
Re(CO)3
4
NaSSO2CH3
DMF
RT
24 h
S
S
O
O
Re(CO)3
1a
I(CH2)3I
RT
30 min
I
Re(CO)3
5
NaSSO2CH3
DMF
RT
24 h
O
O
S
S
Re(CO)3
1b
(CHO)n
RT
OH
Re(CO)3
6
NaI, TMSCl
ACN
RT
20 min
I
Re(CO)3
7
NaSSO2CH3
DMF
RT
24 h
O
O
S
S
Re(CO)3
1c

products were purified with gradient flash chromatography with hexanes/DCM to give **4** and **5**. The ^{1}H and ^{13}C NMR peak shifts and multiplicities agreed with the literature report for **4**.[21] ^{1}H and ^{13}C NMR spectra for **5**, not previously reported, agreed with the expected result based on NMR theory.

For the one carbon-linker label, **1c**, the Li-activated Cp-ring compound (**3**) was alkylated with the addition of paraformaldehyde similar to the method by Herberhold and Biersack.[24] The reaction was brought to room temperature. The disubstituted methylhydroxylated product was observed in small quantities in the NMR spectra of the unpurified product and easily removed purification over an ether flash column. ^{1}H and ^{13}C NMR peaks of the purified product, **6**, matched literature values.[24] The hydroxyl group in **6** was replaced with iodide by way of a TMS intermediate after addition of NaI and TMSCl in ACN, which resulted in **7**, as similarly performed by Olah et al.[27] See Appendix 2 for NMR and EI mass spectra.

5.4.2 Reaction of $[\eta^{5}\text{-}(C_5H_4)(CH_2)_nI]Re(CO)_3$ with Sodium Methanethiosulfonate

The iodide in **4**, **5**, and **7** is replaced by adding five molar excess of sodium methanethiosulfonate and stirring in DMF overnight. The product was purified and DMF removed by diluting in water and extracting into ethyl acetate, which was subsequently washed with water and dried to give **1a-c**.

Evidence for **1a-c** appears in ^{1}H and ^{13}C NMR spectra as the methyl peak appears at ~3.25 ppm and ~51.1 ppm, respectively. Assignment of the methyl peak is supported by HSQC experiments (see Appendix 2). The proper mass for compounds **1a-c** is found in EI spectra. For the IR data, we were unable to sufficiently dissolve **1a-c** in water; however, IR spectra was collected in DMSO and 1:1 CH_2Cl_2/CCl_4, in which a vibrational solvatochromic shift of 11-16 cm^{-1} was observed between the two solvents for the carbonyl modes. The IR spectra are presented in the Appendix 2.

5.4.3 Expression and Characterization of α-Synuclein (71C) Mutant

The V71C mutant of human α-synuclein was generated by site-directed mutagenesis of the engineered gene for the wild type protein sequence, contained in a modified pET32(a) expression vector. DNA sequencing confirmed the presence of the mutants. Expression of His-tagged thioredoxin fusions of the mutants, followed by cleavage with Factor Xa and reversed-phase HPLC, resulted in the isolation of the proteins with >95% purity as measured by HPLC. The V71C mutant was obtained in yields of 4-5 mg per liter of culture. MALDI MS confirmed the identities of the a-synuclein mutants, with observed masses of 14,463.0 Da for the V71C mutant (14,465.1 Da expected).

5.4.4 Reaction of [η^5-(C_5H_4)(CH_2)$_4$$SSO_2Me$]Re(CO)$_3$ with α-Synuclein Mutant (V71C)

Prior to HPLC purification of the mutant, DL-dithiothreitol (DTT) reduced the thiol groups to maximize labeling efficiency. Label **1a** was dissolved in acetone, and ~5 µL of this stock was added to ~1 mL of the protein in HEPES buffer (pH 7.4). Vials were mixed gently for 30 minutes at room temperature, after which the unreacted label was separated from the protein with size-exclusion chromatography. The observed mass of the labeled protein was 14,884.0 Da (14,887.1 Da expected). Given the assumption that the neutrally charged label does not affect ionization, labeling efficiency is ascertained as quantitative from MALDI spectra. All labels (**1a-c**) were attached to cysteine mutants of ubiquitin using the same methodology.[28]

5.4.5 2D IR Characterization of CpRe(CO)$_3$-Labeled α-Synuclein in Buffer With and Without Vesicles

2D IR spectra collected for V71C α-synuclein labeled with **1a** in buffer and with lipid vesicles are shown in Fig. 5.1. In solution, α-synuclein is intrinsically disordered. It takes an α-helical conformation in the presence of lipid vesicles, which is confirmed by a narrowing and red-shift of the amide I absorption band (not shown). Absorption along the diagonal (where $\omega_{pump}=\omega_{probe}$, shown in Fig. 5.1, inset) results from signals that only involve transitions between the vibrational ground and the first excited state, v=0→1 (labeled a and c). v=1→2 transitions appear off-diagonal (labeled b and d) with opposite sign. The shift along the probe axis gives the diagonal anharmonic shift. Peaks a and b are due to two carbonyl asymmetric stretch normal modes, which are doubly degenerate, while peaks c and d are due to the symmetric stretch normal mode. Cross peaks, due to coupling between the normal modes, are found off-diagonal at pump and probe frequencies roughly corresponding to those of the diagonal peaks. They also appear as out-of-phase pairs of peaks (peaks e and f, for example) separated by the off-diagonal anharmonic shift, which is related to the strength of coupling between the normal modes.

The 2D lineshapes of these spectra are a measure of the homogeneous and inhomogeneous broadening caused by the electrostatic heterogeneity of the environment around the metal carbonyl and the dynamical timescales upon which the environment fluctuate. Peaks that are elongated along the diagonal (and thus have a tilt in the node between the v=0>1

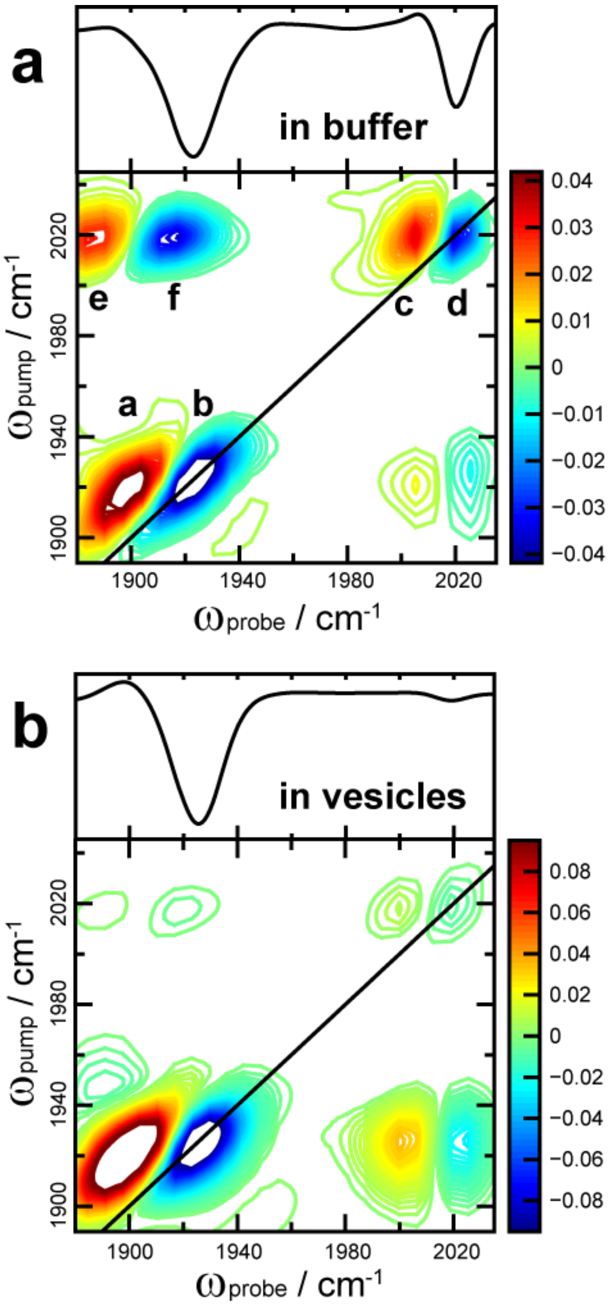


Figure 5.1. 2D IR spectra of α-synuclein labeled with $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$ in buffer (a) and vesicles in buffer (b).

and v=1>2 transitions), have an inhomogeneous contribution to the lineshape. The 2D lineshapes of the $CpRe(CO)_3$ label at V71C in buffer are elongated, indicating that in solution the $CpRe(CO)_3$ label is experiencing a strong electrostatic environment that is slowly evolving on the few picoseconds timescale. Previous FTIR spectra of the vibrational modes of metal carbonyl compounds exhibit broader lineshapes in polar than non-polar solvents, consistent to what we observed here. From FTIR spectra, one cannot determine the origin of the broadening. From the 2D IR lineshapes, we learn that inhomogeneous broadening mechanisms are the culprit.

To better evaluate the influence of environment on the symmetric stretch mode lineshape, high resolution 2D IR spectra were collected for samples in solution and on lipid bilayers (Fig 5.2a-b). The peak frequencies, diagonal linewidths, and node slopes of this mode are reported in Table 5.1. The 2D lineshapes are remarkably similar in the two spectra, indicating that the $CpRe(CO)_3$ label encounters a similar environment when α-synuclein is bound to the lipid vesicles. The vesicle-bound protein spectrum is slightly more inhomogeneously broadened, as evidenced by the wider diagonal linewidth, indicating that the label is experiencing a broader range of electrostatic microenvironments when bound to the lipids. According to the wheel diagram in Fig. 5.2c, V71C resides at the membrane interface, and previous experiments put the depth at 3.9 Å,[17] which lies at the bilayer/buffer interfacial region. Thus, the data is consistent with V71C retaining a similar level of hydration in the buffer and lipids. However, we do not know the propensity for the $CpRe(CO)_3$ to solvate in water versus the membrane. The carbonyl groups in $CpRe(CO)_3$ may extend away from V71C according to the length of the flexible linker and the hydrophobicity of the probe. Future studies will characterize the solvation propensity of these probes, as had been done for EPR

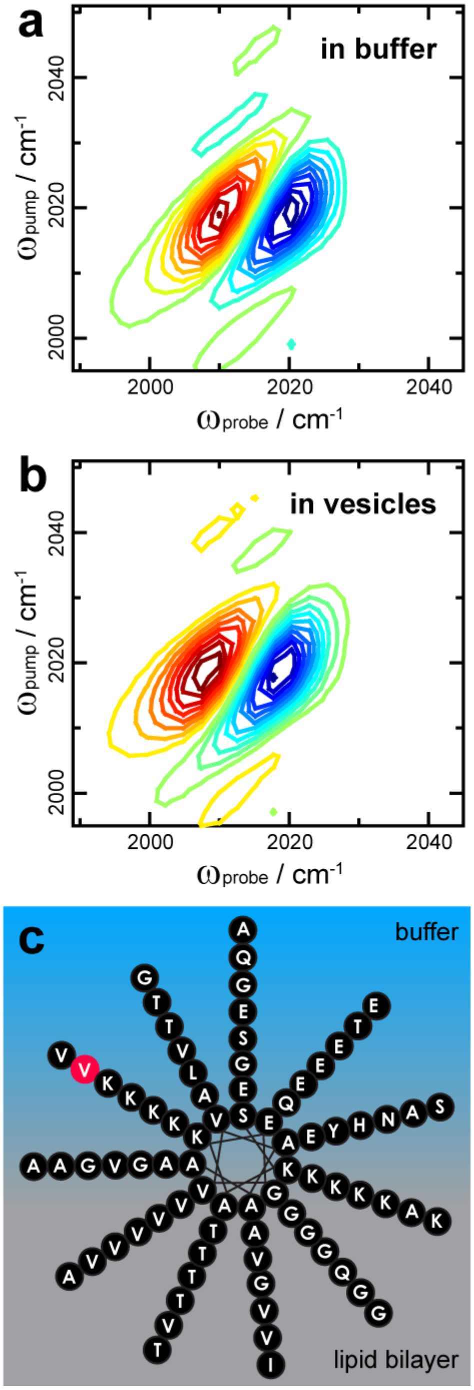


Figure 5.2. 2D IR spectra with 150 g/mm grating for V71C in buffer (a) and in buffer with lipids (b). The helical wheel diagram (c) for –synuclein shows the relative orientation of the protein to the lipid bilayer with the labeled residue (V71) highlighted.

Table 5.1. 2D IR peak frequency, linewidth, and node slope from spectra collected with 150 g/mm monochromator dispersion.

Sample	Frequency (cm^{-1})	Linewidth (cm^{-1})	Node Slope
V71C Buffer	2020.7	8.0	3.9
V71C Lipids	2019.2	9.9	3.8

labels.[29] One distinct difference EPR and FTIR or 2D IR spectroscopies, is that infrared spectra can be calculated from molecular dynamics simulations of only a few microseconds.[30] Thus, understanding the microscopic origins of the 2D IR spectra reported here can be accomplished by comparison of simulations with experiment under a variety of conditions.[3, 31-34]

Vesicle samples give rise to large amounts of background scatter and yet high signal-to-noise spectra are obtained, illustrating the utility of these samples to be used on inhomogeneous systems. Although not reported here, 2D IR spectroscopy can be implemented in a time-resolved fashion to probe dynamics on picoseconds to milliseconds to seconds timescale.[9, 35-38] We believe that these compounds provide a new method for protein structure, dynamics, and environment on systems such as membrane bound and aggregating proteins.

5.5 Conclusion

We have synthesized a series of $CpRe(CO)_3$ metal-carbonyl compounds that can be covalently bonded to proteins through cysteine-selective disulfide bonds using protocols similar to those that are used in EPR spin labeling. Cysteine mutations are a well-established method for probing protein structure site-specifically.[17] Application to V71C mutated α-synuclein demonstrates that high signal-to-noise spectra can be obtained even at low sample concentrations and in membrane environments that cause background scatter. . The 2D IR spectra of these compounds contain information on environmental electrostatics and dynamics, which alter the 2D intensities and lineshapes. If similar to other metal-carbonyl compounds, their lifetimes may also be reflective of hydration.[45] In future studies, we will characterize the hydrophobicity of this label and its effect on protein structure by varying the mutation site on the protein and the length of the linker. The ability to choose a variety of flexible linkers will foster these studies, since long linkers will allow the probe to reside in environments favorable to its hydrophobicity while short linkers will force it to lie close to the protein surface. The probes reported here are larger than other non-natural vibrational probes, but comparable in size to EPR spin probes. Like EPR probes, they will be useful in characterizing protein structure or monitoring dynamics through residues on the outsides of proteins. Due to their high extinction coefficients, samples at <100 μM concentrations can be studied, which is much lower than possible for existing non-natural vibrational probes and thus applicable to a much broader range of proteins. Although not reported here, 2D IR spectroscopy can be implemented in a time-resolved fashion to probe dynamics on picoseconds to milliseconds to seconds timescale.[9, 35-37] We believe that these compounds provide a new method for protein structure, dynamics, and environment that will be especially useful for studying membrane bound and aggregating proteins.

5.6 References

(1) Rubtsov, I. V. *Acc. Chem. Res.* **2009**, *42*, 1385.

(2) Schade, M.; Moretto, A.; Crisma, M.; Toniolo, C.; Hamm, P. *J. Phys. Chem. B* **2009**, *113*, 13393.

(3) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de, P. J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832.

(4) Schkolnik, G.; Utesch, T.; Salewski, J.; Tenger, K.; Millo, D.; Kranich, A.; Zebger, I.; Schulz, C.; Zimanyi, L.; Rakhely, G.; Mroginski, M. A.; Hildebrandt, P. *Chem. Commun.* **2012**, *48*, 70.

(5) Smith, E. E.; Linderman, B. Y.; Luskin, A. C.; Brewer, S. H. *J. Phys. Chem. B* **2011**, *115*, 2380.

(6) Alfieri, K. N.; Vienneau, A. R.; Londergan, C. H. *Biochemistry* **2011**, *50*, 11097.

(7) Bagchi, S.; Boxer, S. G.; Fayer, M. D. *J. Phys. Chem. B* **2012**, *116*, 4034.

(8) Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D. *J. Phys. Chem. B* **2011**, *115*, 11294.

(9) Waegele, M. M.; Culik, R. M.; Gai, F. *J. Phys. Chem. Lett.* **2011**, *2*, 2598.

(10) Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D. *The Journal of Physical Chemistry B* **2011**, *115*, 11294.

(11) Waegele, M. M.; Culik, R. M.; Gai, F. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2598.

(12) Choi, J.-H.; Raleigh, D.; Cho, M. *J Phys Chem Lett* **2011**, *2*, 2158.

(13) Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F.; Hochstrasser, R. M. *Journal of Physical Chemistry Letters* **2010**, *1*, 3311.

(14) Salmain, M.; Jaouen, G. *Comptes Rendus Chimie* **2003**, *6*, 249.

(15) Haquette, P.; Salmain, M.; Svedlung, K.; Martel, A.; Rudolf, B.; Zakrzewski, J.; Cordier, S.; Roisnel, T.; Fosse, C.; Jaouen, G. *ChemBioChem* **2007**, *8*, 224.

(16) Rudolf, B.; Palusiak, M.; Zakrzewski, J.; Salmain, M.; Jaouen, G. *Bioconjugate Chem.* **2005**, *16*, 1218.

(17) Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19666.

(18) Berliner, L. J.; Grunwald, J.; Hankovszky, H. O.; Hideg, K. *Anal. Biochem.* **1982**, *119*, 450.

(19) Jo, H.; Culik, R. M.; Korendovych, I. V.; DeGrado, W. F.; Gai, F. *Biochemistry* **2010**, *49*, 10354.

(20) Chong, D.; Laws, D. R.; Nafady, A.; Costa, P. J.; Rheingold, A. L.; Calhorda, M. J.; Geiger, W. E. *J. Am. Chem. Soc.* **2008**, *130*, 2692.

(21) Mavunkal, I. J.; Moss, J. R.; Bacsa, J. *J. Organomet. Chem.* **2000**, *593-594*, 361.

(22) Casey, C. P.; Czerwinski, C. J.; Fusie, K. A.; Hayashi, R. K. *J. Am. Chem. Soc.* **1997**, *119*, 3971.

(23) Casey, C. P.; Czerwinski, C. J.; Hayashi, R. K. *J. Am. Chem. Soc.* **1995**, *117*, 4189.

(24) Herberhold, M.; Biersack, M. *J. Organomet. Chem.* **1995**, *503*, 277.

(25) Herberhold, M.; Biersack, M. *J. Organomet. Chem.* **1990**, *381*, 379.

(26) Lynch, T. J.; Dominguez, R.; Helvenston, M. C. *Organometallics* **1988**, *7*, 2566.

(27) Olah, G. A.; Narang, S. C.; Gupta, B. G. B.; Malhotra, R. *The Journal of Organic Chemistry* **1979**, *44*, 1247.

(28) Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *J. Am. Chem. Soc.* **in preparation**.

(29) Volkov, A.; Dockter, C.; Bund, T.; Paulsen, H.; Jeschke, G. *Biophys. J.* **2009**, *96*, 1124.

(30) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2011**, *115*, 3713.

(31) Wang, L.; Middleton, C. T.; Singh, S.; Reddy, A. S.; Woys, A. M.; Strasfeld, D. B.; Marek, P.; Raleigh, D. P.; de, P. J. J.; Zanni, M. T.; Skinner, J. L. *J. Am. Chem. Soc.* **2011**, *133*, 16062.

(32) Dijkstra, A. G.; Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2011**, *115*, 5392.

(33) Zhuang, W.; Abramavicius, D.; Voronine, D. V.; Mukamel, S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14233.

(34) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *J. Chem. Phys.* **2005**, *123*, 084905/1.

(35) Smith, A. W.; Tokmakoff, A. *Angew. Chem., Int. Ed.* **2007**, *46*, 7984.

(36) Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T. *Proc Natl Acad Sci U S A* **2006**, *103*, 3528.

(37) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. *Nature (London, U. K.)* **2006**, *444*, 469.

(38) Hamm, P.; Helbing, J.; Bredenbeck, J. *Annu. Rev. Phys. Chem.* **2008**, *59*, 291.

(39) Agbossou, F.; Ramsden, J. A.; Huang, Y. H.; Arif, A. M.; Gladysz, J. A. *Organometallics* **1992**, *11*, 693.

(40) Yang, G. K.; Bergman, R. G. *Organometallics* **1985**, *4*, 129.

(41) Pangborn, A. B.; Giardello, M. A.; Grubbs, R. H.; Rosen, R. K.; Timmers, F. J. *Organometallics* **1996**, *15*, 1518.

(42) Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923.

(43) Fair, J. D.; Kormos, C. M. *J Chromatogr A* **2008**, *1211*, 49.

(44) Westerlund, K.; Moran, S. D.; Privett, H. K.; Hay, S.; Jarvet, J.; Gibney, B. R.; Tommos, C. *Protein Engineering Design & Selection* **2008**, *21*, 645.

(45) King, J. T.; Ross, M. R.; Kubarych, K. J. *J. Phys. Chem. B* **2012**, *116*, 3754.

CHAPTER 6

A Non-Natural Infrared Probe of Protein Electrostatics and Solvation[*]

6.1 Abstract

A non-natural infrared probe of protein hydration and surface electrostatics is reported. The probe consists of a metal-tricarbonyl compound modified with a linker and a leaving group for facile attachment to cysteine residues so that it can be attached to proteins site-specifically via mutagenesis. In solvent, unattached to a protein, the symmetric carbonyl stretch modes of the label has a frequency of 2014.0 to 2014.0 cm^{-1} and a lifetime of 5.6 to 19 ps in DMSO and CH_2Cl_2/CCl_4 (1:1), respectively, as measured by 2D IR spectroscopy. The frequency and lifetime are strongly correlated in these liquids. On ubiquitin in aqueous buffer, the 2D IR spectra of the label exhibit a similar span of frequencies and lifetimes. The correlation between frequency and lifetime holds for linkers consisting of 3 or 4 CH_2 groups, but breaks down for linkers with only a single CH_2 in which the label is constrained to reside near the surface of the protein. We conclude that the lifetime is a measure of solvation whereas the frequency reflects the protein electrostatic field via a Stark shift. These labels have extremely strong absorptions and their incorporation onto protein mimics that of EPR spin labels. We expect that they will be useful for studying the structures and dynamics of large proteins via site-specific hydration and

---

[*] This chapter will be submitted to *J. Am. Chem. Soc.* in April 2012. It was prepared in collaboration with S. Mukherjee and M. T. Zanni.

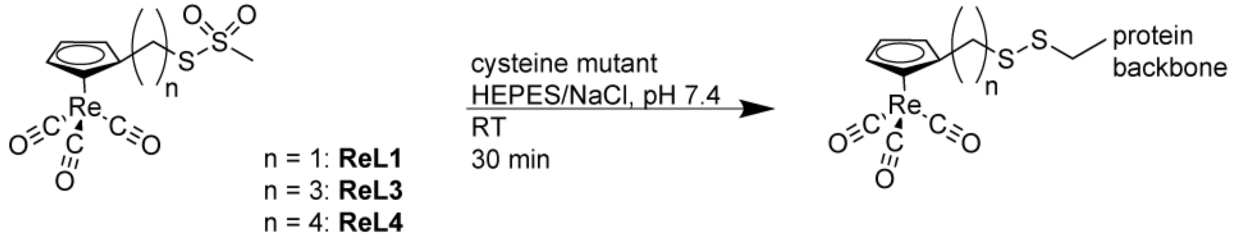
electrostatic fields, such as occurs in membrane bound, aggregated, and kinetically evolving proteins.

6.2 Introduction

The vibrational modes of most molecules are sensitive to their environment. Frequencies scale with the local electrostatic field while the linewidths depend on the disorder of the surrounding environment.[1-3] As such, they are good reporters of protein solvation, electrostatics, and hydrogen bonding. For proteins, the challenge is to resolve a single vibrational mode from a congested infrared spectrum. One method is to use isotope labeling, such as a backbone $^{13}C^{18}O$ carbonyl label[4] or a sidechain with a CD substitution.[5] Isotope labels are ideal reporters because they do not perturb the protein structure or dynamics. They can be easily used to study polypeptides that can be synthesized, but incorporating isotope labels site-specifically into proteins is challenging, although methods do exist. Another approach is to use non-natural vibrational probes, such as nitriles,[6] nitrophenyl,[7] and thiocyanates,[8,9] whose triple bond character puts them in a window of the infrared spectrum that does not overlap with most protein vibrational modes or with water absorbance. They can be straightforwardly incorporated into proteins via mutagenesis.[7-10] They are small enough that they are minimally perturbative to the protein structure and thus can reside inside the protein, even near the active site.[10] Their drawback is that they are weak absorbers.[11] Thus, they typically require mM concentrations, which limit their use to highly stable proteins.[8,10-14] Their weak oscillator strength makes especially difficult to use in conjunction with 2D IR spectroscopy, because 2D IR signals scale as the square of the absorption signal.[15] In this study, we report on a new non-natural probe that we have synthesized that is designed to be used at very low protein concentrations and can be incorporated at arbitrary positions in a protein using cysteine or methionine mutagenesis. This probe can be utilized to study protein structure and dynamics, because as we report in this paper, its frequency and lifetime are sensitive to local electrostatic fields and hydration.

The family of probes that we have synthesized is shown in Scheme 6.1. They utilize the carbonyl stretch modes of $CpRe(CO)_3$, which lie in a region of the infrared spectrum that is well resolved from native protein vibrational transitions and water absorbance.[16] The Cp ring is substituted with $–(CH_2)_nSSO_2CH_3$, where n is 1, 3, or 4 (Scheme 6.1, hereafter referred to as ReL1, ReL3, ReL4, or generally as ReLx). Its synthesis is reported elsewhere.[17] The methanethiosulfonate is a leaving group commonly used to attach nitroxide spin labels to proteins via cysteine in EPR experiments.[18,19] It forms a disulfide bond, resulting in a covalent attachment that is stable to the oxygen and water-rich environment of a protein. As a result, our compound can be incorporated into proteins using site-specific mutagenesis with the same attachment protocols. Metal carbonyl compounds have been studied extensively with infrared spectroscopy, transient infrared spectroscopy and 2D IR spectroscopy due to their interesting catalytic properties.[20] Their carbonyl stretch modes have very large oscillator strengths, which means that the corresponding proteins to which they are attached can be studied at very low concentrations. From the data presented below, we estimate protein concentration <100 μM can be routinely measured. Most metal-carbonyl molecules are insoluble in water, and so the full extent to which their frequency, lineshapes or lifetimes are sensitive to environment has not yet been well-studied. However, inhomogeneous lineshapes have been measured in the 2D IR spectra of $Ni_2(PPh_3)_2(CO)_2$ in THF–hexane (1:3)[21] and the lifetime of a semi-water soluble metal

Scheme 6.1


O
S
O
S
n
Re
O≡C
C≡O
C
O
n = 1: **ReL1**
n = 3: **ReL3**
n = 4: **ReL4**
cysteine mutant
HEPES/NaCl, pH 7.4
RT
30 min
S
S
n
protein
backbone
Re
O≡C
C≡O
C
O

carbonyl has recently been shown to be quite sensitive to solvent polarity.[22] We report here that these molecules are indeed good reports of protein environment.

6.3 Experimental Section

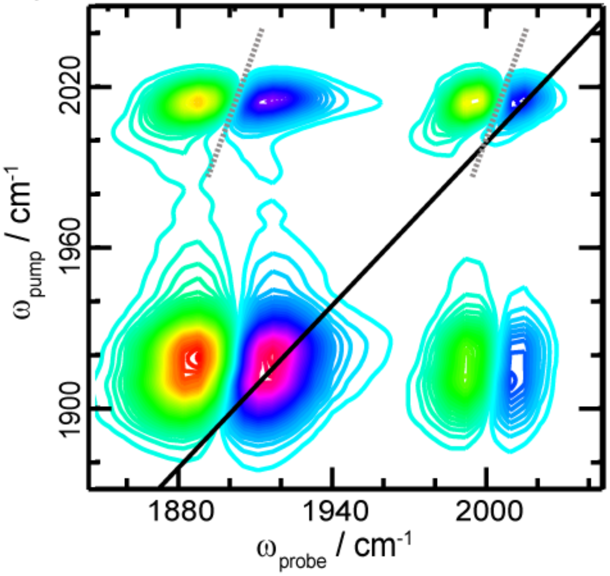
Synthesis of the labels is described in detail elsewhere.[17] In brief, $Re_2(CO)_{10}$ is oxidized with cyclopentadiene to give $CpRe(CO)_3$.[16] A hydrogen atom on the cyclopentadienyl ligand is substituted by activation with n-BuLi in THF at -78^{o}C. For ReL1-label, hydroxymethylation occurs with paraformaldehyde. After purification, the hydroxyl group is replaced with an iodide via a TMS intermediate to give $[\eta^5\text{-}(C_5H_4)CH_2I]Re(CO)_3$. For the ReL3- and ReL4-labels, the route to the iodide is direct from the lithium-activated Cp-ring with diiodopropane or diiodobutane to $[\eta^5\text{-}(C_5H_4)(CH_2)_3I]Re(CO)_3$ or $[\eta^5\text{-}(C_5H_4)(CH_2)_4I]Re(CO)_3$, respectively.[23] For any of these iodide products, $[\eta^5\text{-}(C_5H_4)(CH_2)_nI]Re(CO)_3$ where n=1,3 or 4, compounds, the iodide is replaced using $NaSSO_2CH_3$ in DMF to give $[\eta^5\text{-}(C_5H_4)(CH_2)_nSSO_2CH_3]Re(CO)_3$ where n=1,3 or 4. After aqueous workup the labels are ready to react with any reduced thiols on a protein.

The ubiquitin mutants K6C and K63C were expressed by Ellen Valkevich in Eric Streiter's lab in the UW-Madison Chemistry Department. Five-fold excess of each label was reacted with 500 μg aliquot of each protein in 1 mL of 10 mM HEPES, 100mM NaCl buffer (pH 7.4) buffer. The solution was stirred for 1 hour, after which time unreacted label was separated from the protein with a 60cm Sephadex 75 size-exclusion column and 10 mM NH_4CO_3 buffer (pH 7) eluting buffer. Elutants containing the labeled protein were lyophilized and characterized by MALDI (See SI). For FTIR and 2DIR measurements, the protein was resuspended in 20 mM deuterated-phosphate buffer at 500 μM concentrations and placed between two calcium fluoride windows separated with a 56 micron Teflon spacer.

Mid IR pulse (1960/1970 cm^{-1} center frequency, 185 cm^{-1} bandwidth) were generated from a Ti Sapphire femtosecond laser system pumping an optical parametric amplifier and difference frequency mixing the output. The mid IR light was split into pump (1μJ) and probe (0.1μJ) pulses and spatially and temporally overlapped in the sample. A mid IR pulse shaper described elsewhere[24] was used to shape the pump pulses to generate a pulse pair with time delays scanned up to 4 ps in steps of 40 fs. The waiting time was scanned by a calibrated motorized translational stage in 200-1000 fs steps up to 50 ps. The generated signal was then detected by heterodyning with the probe and frequency resolving with a spectrometer on a linear one dimensional MCT array with 64 pixels. The frequency resolution along the pump and probe axis are 8 cm^{-1} and 3 cm^{-1} respectively. The pump and probe polarizations were set parallel.

6.4 Results and Discussion

The 2D IR spectrum of ReL1 in DMSO and attached to K63C ubiquitin is shown in Fig. 6.1. The diagonal and cross peaks in the spectrum are elongated along the diagonal, thereby creating nodal slopes (dashed lines) that are tilted. The slope of the node between the overtone and fundamental diagonal peak is a measure of the inhomogeneity of the vibrational mode,[25] whereas the slope of the cross peak node also depends on the correlation in their vibrational frequency fluctuations.[26] In non-polar solvents, the nodal slopes are more vertical, indicating that the modes are more homogeneous (Table 6.1). Thus, the carbonyl stretch modes are sensitive to solvent polarity and dynamics, similar to the amide I carbonyl stretches of protein backbones that are often used to study protein structure. As has been done for the amide I a.) DMSO



b.) K63C

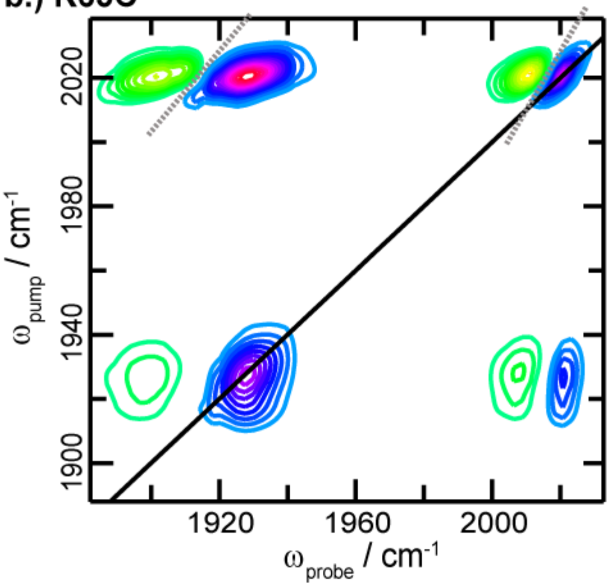


Figure 6.1 - 2DIR spectra of the ReL1 in DMSO (a) and attached to ubiquitin mutant, K63C (b).

Table 6.1. Symmetric mode frequency, lifetime, node slope, and cross peak node slope ReLx in in solvent or covalently linked to the protein.

Solvent / Protein	**Label**	**Frequency, ω_{sym} (cm^{-1})**	**Lifetime, ω_{sym} (ps)**	**Node Slope, ω_{sym}**	**Node Slope – Cross Peak, $\omega_{pump=sym}\omega_{pump}$**
DMSO	ReL1	2014.0	6.1	2.5	2.4
	ReL3	2015.6	10	3.4	2.3
	ReL4	2014.2	8.4	3.8	2.3
CH_2Cl_2/CCl_4	ReL1	2024.1	16	1.9	1.9
	ReL3	2024.5	19	2.0	1.7
	ReL4	2022.4	14	5.6	4.0
K6C	ReL1	2022.2	8.5	1.6	1.2
	ReL3	2019.4	12	1.6	1.5
	ReL4	2015.4	4.7	1.6	1.0
K63C	ReL1	2020.8	5.8	1.9	1.3
	ReL3	2017.0	5.7	1.7	1.2
	ReL4	2020.4	14	1.9	1.2

modes, it should be possible to develop a correlation between frequency and solvent electrostatic field,[27-30] which would enable a quantitative measure of solvent polarity. In this report we do not dwell on these features, but instead focus the frequencies and lifetimes instead.

Shown in Fig. 6.2a is the peak associated with the symmetric stretch mode as illustrated by a slice through the diagonal ($\omega_{pump}=\omega_{probe}$) of 2D IR spectra for ReLx in DMSO and 1:1 CH_2Cl_2/CCl_4. There is a 12 cm^{-1} frequency difference between these two solvents, which is as large as the FWHM of the peaks themselves. The 1:1 CH_2Cl_2/CCl_4 mixture is used here because the compounds with the linker is not soluble in CCl_4 alone. The other compounds show the same solvent frequency dependence (Table S1. Spectra reported in the SI). Within the same solvent, the frequency varies ~2 cm^{-1}, which we attribute differences in the degree to which the leaving group interacts with the carbonyl ligandsReL4. Thus, the symmetric stretch is expected to have a frequency of about 2015 cm^{-1} in polar environments and 2024 cm^{-1} in non-polar environments.

The lifetime of the ReLx symmetric stretch correlates with its frequency. In the 1:1 CH_2Cl_2/CCl_4 mixture, the ReL1 lifetime is 16 ps, which is similar to other metal carbonyl compounds in non-polar solvents.[31] In DMSO, the ReL1 lifetime is 6.1 ps. To our knowledge, the lifetime of very few metal carbonyl compounds are known in polar solvents. All the compounds exhibited lifetimes <10 ps in DMSO and > 14 ps in 1:1 CH_2Cl_2/CCl_4 (Table 6.1 and S1). The lifetime can be measured from both the diagonal absorption and the off-diagonal overtone. In what follows, we discuss the overtone because there is no overlap with sample background scatter, when present, and thus is generally more accurate especially in heterogeneous samples like those involving membranes.). The antisymmetric stretch mode

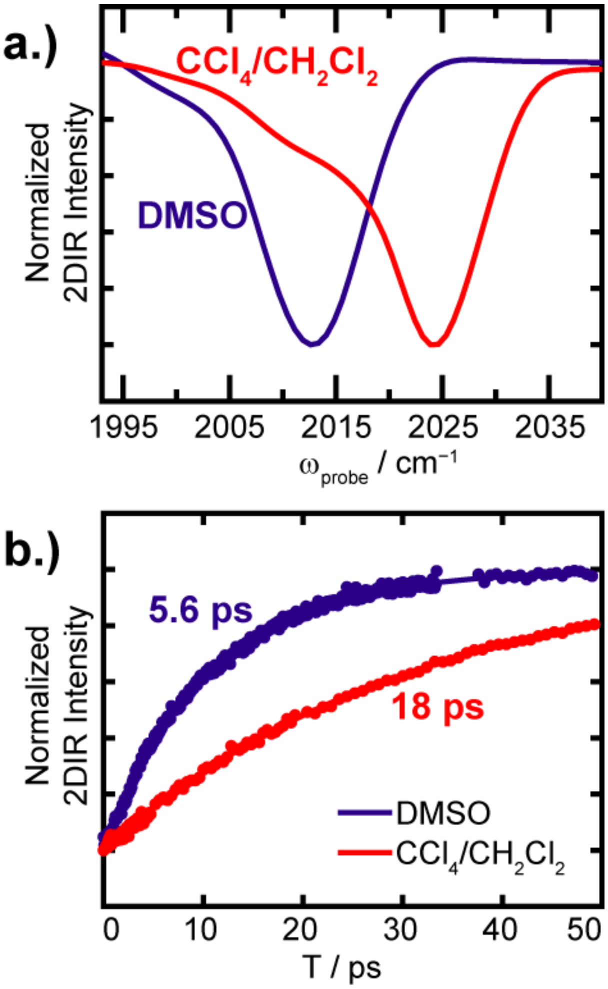


Figure 6.2 - 2DIR intensity through the 2D IR diagonal ($\omega_{pump}=\omega_{probe}$) and peak intensities over waiting times for samples of each label in either DMSO or 1:1 CH_2Cl_2/CCl_4.

lifetime and frequency depends on solvent, but since it is a degenerate mode, we do not pursue it more here. Thus, there is a very large dynamic range in both the lifetime and the frequency of the symmetric stretch mode. The strong correlation between lifetime and frequency is evident in Fig. 6.3a, which plots both in the same graph.

Having identified an empirical relationship between frequency and lifetimes, we attached them to two different locations on ubiquitin. We attached each of the labels to each of the two sites on ubiquitin, resulting in a total of six samples. Representative 2D IR spectra are shown in Fig. S1. From these measurements, we extracted the symmetric stretch frequency and the lifetime, which are plotted in Fig. 6.3. The six frequencies span from 2015.4 to 2022.2 cm^{-1}. On average, they lie close to the frequency observed for ReLx in DMSO than in the CH_2Cl_2/CCl_4 mixture (Fig. 6.3). The lifetimes span from 4.7 to 14, which is also more similar to the lifetimes of ReLx in DMSO than CH_2Cl_2/CCl_4. These averages are consistent with the protein being solvated in water. What is more interesting is that the correlation between frequency and lifetime is retained for linkers ReL3 and ReL4, but not for ReL1. For ReL3 and ReL4, whether the label is attached to K6C or to K63C, a high frequency is accompanied by a long lifetime, as was observed for the model compounds (Fig. 6.3). However, for ReL1, a high frequency is accompanied by a short lifetime.

We interpret this data with regards to the hydration of the probe and the electrostatic field of the protein. Vibrational frequencies correlate with the surrounding field strength due to the Stark effect.[32,33] The Stark effect is utilized in small vibrational probes, like nitriles, to measure the internal and external electric fields of proteins. The same principle holds true for more complicated vibrations like the amide modes of protein backbones. In contrast, vibrational lifetimes are often dictated by time-dependent fluctuations in the frequency more so than the

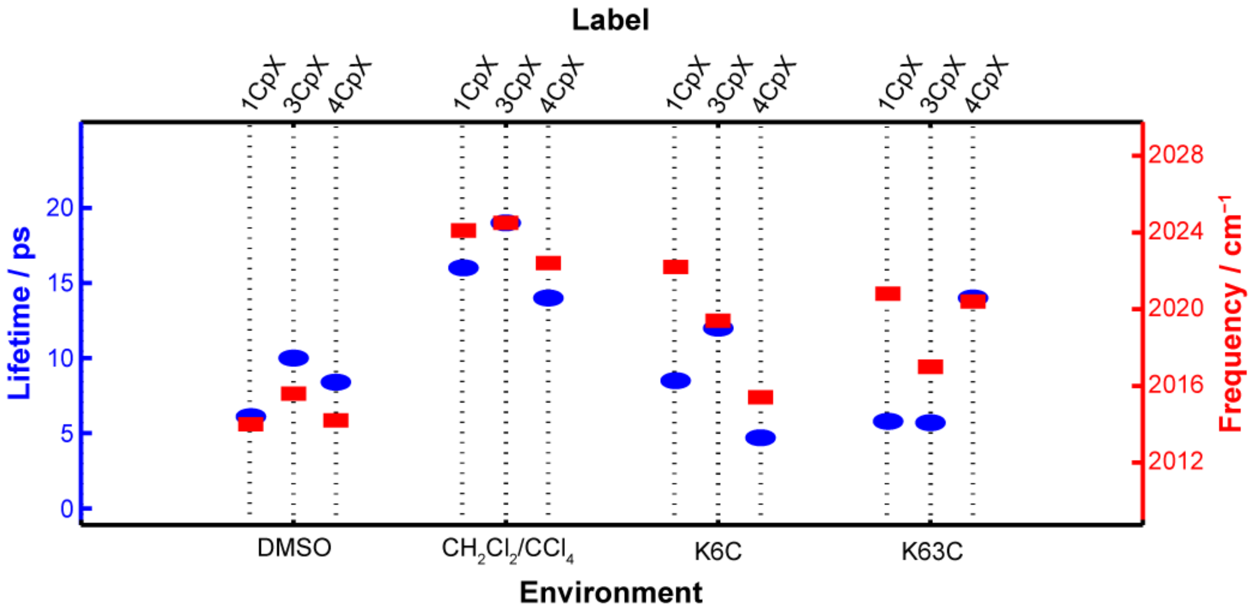


Figure 6.3 - Lifetime and frequency plotted for the labels just in solvent and covalently linked to two mutants of ubiquitin.

frequencies themselves. The link between fluctuations and lifetimes occurs because vibrational energy of high frequency modes like those studied here, flows through covalent linkages to populate lower vibrational energy states, but still resides in the molecule. Inter-molecular energy transfer does not typically occur until vibrational energy is redistributed throughout the internal modes of the probe molecule. Energy redistribution in the molecule requires a time-dependent breakdown of the normal modes, which is provided by solvent collisions. Thus, in general, vibrational frequencies reflect the average electrostatic field of their surroundings while the lifetime is a measure of the rate and magnitude of environment induced frequency fluctuations.[34]

In water and DMSO, the above description of frequencies and lifetimes holds for CpLx. DMSO has, on average, stronger electrostatic fields and larger frequency fluctuations then does the non-polar solvent mixture CH_2Cl_2/CCl_4 (the magnitude of the frequency fluctuations is observed in the broader FWHM of the symmetric stretch in DMSO than CH_2Cl_2/CCl_4 in Fig. 6.2). Thus, the two are correlated in simple solvents that provide a homogeneous environment around the label. From the protein data, it appears that ReL4 at K6C and ReL3 at K63C sample a homogeneous environment that is similar to water, since both the lifetime and frequency have values close to the label in liquid water. Likewise, the frequency and lifetime of ReL3 at K6C and ReL4 at K63C are correlated, which we interpret to mean that these two labels are also sampling a homogeneous environment, although the environment is less polar than water since both the frequency and lifetime assume intermediate values as in the water and CH_2Cl_2/CCl_4 model studies. In contrast to the ReL3 and ReL4 linkers, the ReL1 linkers do not follow the same trend. At both positions on ubiquitin, the lifetime is short and the frequency is high, which suggests that the ReL1 label is experiencing rather weak electrostatic fields but very dynamic field fluctuations, which is unlike what is observed in liquids. From these observations, we conclude that the labels held by ReL1 linkers are so short that their frequencies are dominated by the electrostatics of the protein surface, but that the solvent to which they are exposed is still sufficient to cause fast vibrational lifetimes. In contrast, the ReL3 and ReL4 labels are long enough that they can more broadly sample the homogeneous solvent environment.

Shown in Fig. 6.4 is a map of the ubiquitin surface hydrophobicity in each of the two regions to which our labels are attached. Drawn on this map is one possible conformation that each label might adopt, with a circle drawn to illustrate the extent to which the label could sample the protein surface if the linker group is fully extended. K6C lies in a pocket while K63C is located on the edge of the protein. These conformations are not energy minimized and ignore steric overlaps. Thus, the region that the label samples is most likely smaller, but they permit us to evaluate the outer limit to the possible conformations. For example, the metal-carbonyl headgroup cannot extend further than 7.1 Å from the protein surface when attached at K6C and no further than 8.8 Å at position K63C. Thus, it is likely that labels with the ReL1 linker will feel the protein electrostatic field, as deduced from the experiments above. The ReL3 and ReL4 linkers allow the headgroup to sample a much wider range of the protein surface. Presumably, the label will preferentially cluster near hydrophobic regions of the protein (shown in red), since the label is hydrophobic. The carbonyl groups on metal-carbonyl compounds do not form hydrogen bonds. Thus, the carbonyl groups could face towards the protein to be protected from the solvent. However, there may be stabilizing forces between the cyclopentadienyl ring and rings of side chains, like phenylalanine, that prefer the carbonyl groups to face towards the solvent. Thus, samples in which frequency and lifetime do not correlate may be an indication of preferential surface orientation.

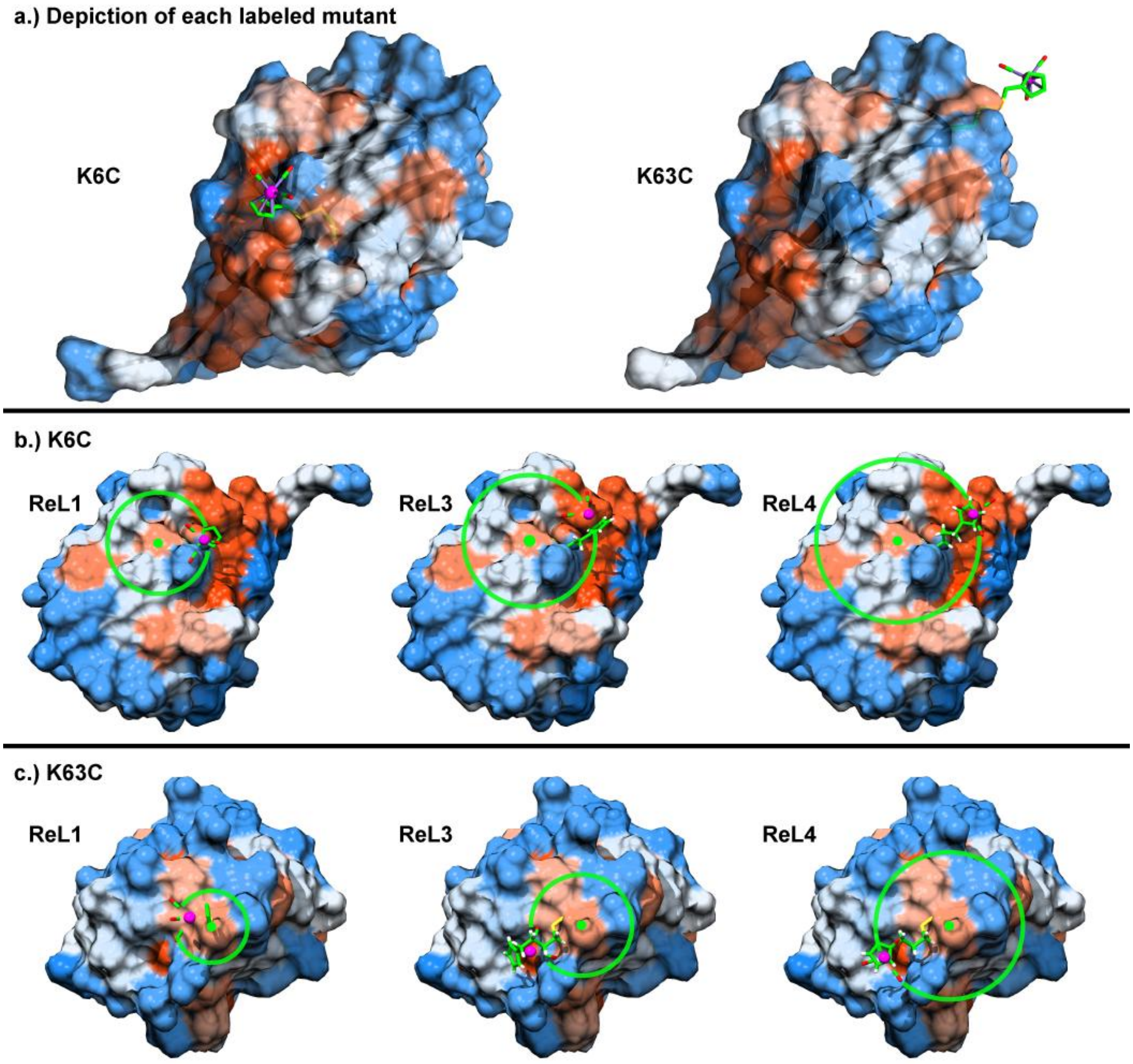


Figure 6.4 - Side view of the two labeled positions of ubiquitin, K6C and K63C with ReL1 label attached (a). Drawings of ReL1, ReL3, and ReL4 labels attached to K6C (b) and K63C (c) with circles drawn about the cysteine β-carbon.

6.5 Conclusion

We have deduced an empirical relationship between lifetime and frequency as related to solvation and electrostatics for this novel set of infrared probes. As illustrated in these experiments, these probes reflect the microscopic structure and environment of the protein surface. Additional insights into the conformational propensities of these labels will come from additional site mutations as well as molecular dynamics simulations. A useful feature of infrared spectroscopy is that correlations between electrostatic fields and 2D IR spectra can be obtained, which can also be used to calculate lifetimes. The theory will have to account for the normal mode behavior of these compounds, since all three carbonyl stretches contribute to the observables. Nonetheless, it will be possible to quantitatively link the experimental observables to the atomic level features of the protein and its solvent. Steps along these lines are already underway for other metal carbonyl groups. The utility of the labels presented here, is that they can be applied to site-specifically to a wide range of proteins at very low concentrations. We anticipate that they will be especially useful in studies of membrane bound proteins and kinetically evolving proteins such as occurs during amyloid fiber formation.

Before exploiting the advantages of 2DIR spectroscopy to study aggregation and membrane proteins, we needed to carefully examine what the 2DIR data tells us exactly about the environment. To test these labels, we first studied their vibrational dynamics in solvents of ranging polarity, for which we selected DMSO and 1:1 CH_2Cl_2/CCl_4. Then we choose a well-studied and easy-to-handle model protein for characterization of the labels attached to a protein. For this protein, we chose two ubiquitin mutants, K6C and K63C. Two these two mutants, our three labels were attached and used for 2DIR waiting time experiments.

6.6 References

(1) Kubo, R. In *Adv. Chem. Phys.*; John Wiley & Sons, Inc.: 2007, p 101.

(2) Berne, B. J.; Pecora, R. *Dynamic Light Scattering: with Applications to Chemistry, Biology, and Physics*; Wiley-Interscience, 1975.

(3) Hamm, P.; Zanni, M. *Concepts and Methods of 2d Infrared Spectroscopy*; Cambridge University Press, 2011.

(4) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de, P. J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832.

(5) Naraharisetty, S. R. G.; Kurochkin, D. V.; Rubtsov, I. V. *Chem. Phys. Lett.* **2007**, *437*, 262.

(6) Marek, P.; Mukherjee, S.; Zanni, M. T.; Raleigh, D. P. *J. Mol. Biol.* **2010**, *400*, 878.

(7) Smith, E. E.; Linderman, B. Y.; Luskin, A. C.; Brewer, S. H. *J. Phys. Chem. B* **2011**, *115*, 2380.

(8) Alfieri, K. N.; Vienneau, A. R.; Londergan, C. H. *Biochemistry* **2011**, *50*, 11097.

(9) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 13356.

(10) Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D. *J. Phys. Chem. B* **2011**, *115*, 11294.

(11) Waegele, M. M.; Culik, R. M.; Gai, F. *J. Phys. Chem. Lett.* **2011**, *2*, 2598.

(12) Bagchi, S.; Boxer, S. G.; Fayer, M. D. *J. Phys. Chem. B* **2012**, *116*, 4034.

(13) Choi, J.-H.; Raleigh, D.; Cho, M. *J. Phys. Chem. Lett.* **2011**, *2*, 2158.

(14) Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F.; Hochstrasser, R. M. *Journal of Physical Chemistry Letters* **2010**, *1*, 3311.

(15) Kim, Y. S.; Hochstrasser, R. M. *J Phys Chem B* **2009**, *113*, 8231.

(16) Agbossou, F.; Ramsden, J. A.; Huang, Y. H.; Arif, A. M.; Gladysz, J. A. *Organometallics* **1992**, *11*, 693.

(17) Woys, A. M.; Skoff, D.; Moran, S. D.; Zanni, M. T. *Organometallics* **in preparation**.

(18) Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R. *Proc Natl Acad Sci U S A* **2008**, *105*, 19666.

(19) Berliner, L. J.; Grunwald, J.; Hankovszky, H. O.; Hideg, K. *Anal. Biochem.* **1982**, *119*, 450.

(20) Baiz, C. R.; McRobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. *Acc. Chem. Res.* **2009**, *42*, 1395.

(21) Shim, S.-H.; Zanni, M. T. *Physical Chemistry Chemical Physics* **2009**, *11*, 748.

(22) King, J. T.; Ross, M. R.; Kubarych, K. J. *J. Phys. Chem. B* **2012**, *116*, 3754.

(23) Mavunkal, I. J.; Moss, J. R.; Bacsa, J. *J. Organomet. Chem.* **2000**, *593-594*, 361.

(24) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* **2010**, *52*, 12.

(25) Kwak, K.; Rosenfeld, D. E.; Fayer, M. D. *J. Chem. Phys.* **2008**, *128*, 204505/1.

(26) Hayashi, T.; Mukamel, S. *J. Chem. Phys.* **2006**, *125*, 194510/1.

(27) Hayashi, T.; Mukamel, S. *J. Mol. Liq.* **2008**, *141*, 149.

(28) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2009**, *113*, 592.

(29) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2011**, *115*, 3713.

(30) Lee, H.; Lee, G.; Jeon, J.; Cho, M. *J. Phys. Chem. A* **2012**, *116*, 347.

(31) Heilweil, E. J.; Cavanagh, R. R.; Stephenson, J. C. *Chem. Phys. Lett.* **1987**, *134*, 181.

(32) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science (Washington, DC, U. S.)* **2006**, *313*, 200.

(33) Choi, J.-H.; Oh, K.-I.; Cho, M. *J. Chem. Phys.* **2008**, *129*, 174512/1.

(34) Eaves, J. D.; Tokmakoff, A.; Geissler, P. L. *J. Phys. Chem. A* **2005**, *109*, 9424.

APPENDIX 1

Supporting Information for

CHAPTER 3

2D IR Lineshapes Probe Ovispirin Peptide Conformation and Depth in Lipid Bilayers *

A1.1 Peptide Synthesis

Ovispirin has the sequence H_2N-KNLRR IIRKI IHIIK KYG-COOH. Samples were prepared by standard FMOC solid phase peptide synthesis. Eleven samples were each $^{13}C{=}^{18}O$ labeled at the backbone carbonyl group for the native amino acids at residues 1, 3, 6, 7, 9, 10, 11, 13, 14, 15, 16. At positions 4, 5, 8 and 12 a $^{13}C{=}^{18}O$ labeled glycine mutation was performed instead. FMOC amino acids were prepared by acid catalyzed exchange of the carboxylic acid oxygen with free amino acids already containing ^{13}C at the carbonyl carbon (Cambridge Isotope Laboratories) similar to the method described by Torres et al.,[1] and then protecting the α-amine group with FMOC.[2] For lysine, protection of the side chain is necessary, which was accomplished by protecting the α-amine with copper, protecting the epsilon amine with Boc, and removing the copper protection.[3] The Fmoc group was then added using the same protocol employed for the other amino acids. Peptides were synthesized with a Protein Systems

---

* This chapter was originally published in *J. Am. Chem. Soc.* **2010**, *132*, 2832-2838. It was prepared in collaboration with Y.-S. Lin, A. S. Reddy, W. Xiong, J. J. de Pablo, J. L. Skinner, and M. T. Zanni.

Symphony synthesizer. All peptides were purified with reverse phase HPLC on a Vydac C-18 column with gradient 0.1% TFA water/acetonitrile solvent system. Masses were confirmed with MALDI mass spectrometry. Residual trifluoroacetic acid was removed from peptide samples by performing three iterations of 0.1 M HCl dissolution and lyopholization of purified peptide. Labeling efficiency was >70% as determined from electrospray mass spectroscopy of protect amino acid labels.

A1.2 2D IR Spectroscopy

The peptides were mixed with lipid vesicles made from chloroform solutions of 1-palmitoyl-2oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) (Avanti Polar Lipids) that were combined in 3:1 molar ratio. This composition matches that used in NMR experiments.[4] Chloroform was removed with a Speedvac and lipids were reconstituted in 0.1 mM phosphate buffer containing peptide such that the peptide-to-lipid ratio was larger than 1:21. Ovispirin is in a surface associated state even at peptide-to-lipid concentrations as small as 1:11.[4] The optical density of the isotope labeled band was about 0.05 at 1595 cm^{-1} in these experiments. The experiments were performed in H_2O solvent so that arginine sidechain absorption does not interfere with the isotope labeled peaks. Excess bulk H_2O was removed with a nitrogen stream to reduce absorption at 1650 cm^{-1} by the water bend vibrational mode. The bilayers are still fully hydrated as monitored by the water stretching band at 3300 cm^{-1}.

2D IR spectra of each labeled peptide in lipid / H_2O buffer were collected with a pulse-shaping pump probe spectrometer in a collinear geometry as described previously. The pump was scanned from 0 to 2560 fs in 24 fs steps with the waiting time of 0 fs. A 75 g/mm grating dispersed the signal heterodyned with the pump onto a 64 channel array detector. The resulting 2D IR spectra have a resolution of 1.4 in the pump axis and 3.9 cm^{-1} in the probe access.

2D IR spectra were fit to extract the 2D lineshapes for the isotopically isolated amide I mode. To remove spectral congestion from the isotope peak region, the frequency range spanning from 1620 cm^{-1} to 1675 cm^{-1} were fit to the sum of six lineshape equations. Crosspeak contributions to the diagonal lineshapes are negligible and were ignored. The lineshape equation describes Bloch dynamics in the inhomogeneous limit. The rephasing and non-rephasing response functions are calculated as[5]

$$R_{123}(t_3,t_1) \sim e^{-2g(t_1)-2g(t_3)+g(t_1+t_3)}\left[2|\mu_{10}|^4 e^{-i\langle\omega_{10}\rangle(t_3-t_1)} e^{-\frac{t_1+t_3}{2T_{10}}} - |\mu_{10}|^2|\mu_{21}|^2 e^{-i(\langle\omega_{21}\rangle t_3-\langle\omega_{10}\rangle t_1)} e^{-\frac{t_1+t_3}{2T_{10}}} e^{-\frac{t_3}{2T}}\right] \tag{1}$$

$$R_{456}(t_3,t_1) \sim e^{-g(t_1+t_3)}\left[e^{-i\langle\omega_{10}\rangle(t_3+t_1)} e^{-\frac{t_1+t_3}{2T_{10}}} - e^{-i(\langle\omega_{21}\rangle t_3+\langle\omega_{10}\rangle t_1)} e^{-\frac{t_1+t_3}{2T_{10}}} e^{-\frac{t_3}{2T_{21}}}\right], \tag{2}$$

respectively and the fitting expression in the limit of δ-function pulses is

$$S(\omega_3,\omega_1) = \mathrm{Re}\left[\int_0^\infty dt_1 \int_0^\infty dt_3 e^{i\omega_3 t_3 - i\omega_1 t_1}\,\mathrm{Re}[R_{123}(t_3,t_1)] + \int_0^\infty dt_1 \int_0^\infty dt_3 e^{i\omega_3 t_3 + i\omega_1 t_1}\,\mathrm{Re}[R_{456}(t_3,t_1)]\right]. \tag{3}$$

Here g(t) is the lineshape function given below in Eq. 7, $<\omega_{10}>$ is the average transition frequency between states υ=1 and 0 and $<\omega_{10}>$ between states υ=2 and 1; T_{10} and T_{21} are the population relaxation times from state υ=1 to 0 and 2 to 1, respectively. The evolution period was extended to 2560 fs in 24 fs increments to mimic the pump conditions. The waiting time, t_2, was set to zero. The t_1 time was scanned to 4000 fs in 9 fs steps. An anharmonicity of 14 cm^{-1} was used, which is consistent with our fits and many other peptide studies.[6-8] Vibrational lifetimes were measured by collecting successive 2D spectra as the waiting time was varied for the I14 sample. T_{10} and T_{21} were determined by separately fitting a single exponential to the fundamental and overtone intensities as the function of the waiting time (t_2), which gave 852 fs and 683 fs, respectively, for the unlabeled amide I band. The ratio of $|\mu_{21}|^2/|\mu_{10}|^2$ was allowed to vary between 1.5 and 3.4 to account for attenuation of the excitation pulses by water absorption. Fits to all 2D IR spectra are shown in Figure A1.1. Linewidths are given in Table A1.1.

A1.3 MD Methods

Molecular dynamics simulations were performed on ovispirin with a POPC/POPG (3:1) lipid bilayer. The initial coordinates for the peptide were obtained from Protein Data Bank (PDB) entry 1HU5. The peptide was modeled using the GROMOS87 force field.[9] The ionization state of the amino acid side chains was decided based on their pKa values. The resulting peptide structure was found to have a net positive charge of +7. To make the system charge neutral, seven Cl^- ions were added to the simulation box. The peptide mutants were prepared from the corresponding wild type structures by reconstructing the appropriate side chains while maintaining their internal coordinates. The energy of the resulting structure was minimized and equilibrated for at least 10 ns.

The bilayer membrane considered in this work consisted of 96 POPC molecules and 32 POPG molecules. The lipid molecules were also modeled using the GROMOS87 force field.[9] The starting coordinates for the mixed bilayer was prepared starting from an equilibrated POPC bilayer with 128 lipids (64 per leaflet).[10] The equilibrated POPC bilayer coordinates were then scaled in x and y direction (bilayer normal is along z-axis) by 10%. Thirty two lipid molecules (16 in each leaflet) were then randomly selected and were replaced with POPG molecules.[11] To make the system charge neutral, 32 Na^+ counter ions were added. The energy of resulting

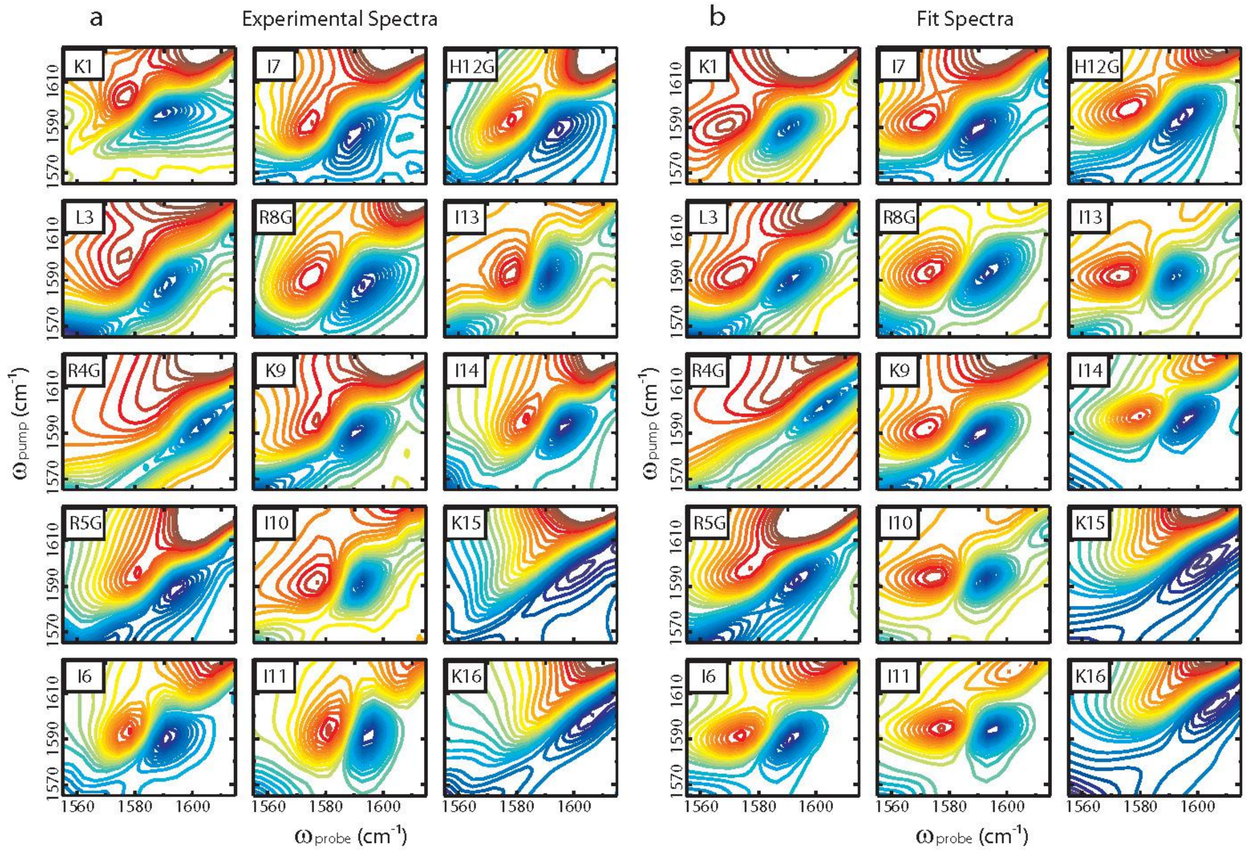


Figure A1.1. Data and fits for ovispirin peptide with $^{13}C{=}^{18}O$ labels at different backbone positions. (a) 2D IR spectra of each isotope labeled peptide in the region of the isotope label. (b) Calculated 2D IR spectra using parameters from fitting the experimental spectra.

Table A1.1. Experimental and simulated linewidths.

Residue	**Diagonal Linewidth, cm^{-1} (±3 cm^{-1})**	**Inhomogeneous Linewidth, cm^{-1}**	**Homogeneous Linewidth, cm^{-1}**	**Simulated Diagonal Linewidth, cm^{-1}**
K1	14.1	11.3	7.07	16.5
L3	16.8	15.0	6.59	14.7
R4G	19.0	17.5	5.04	13.7
R5G	16.3	15.0	6.75	20.0
I6	13.5	12.5	5.12	21.0
I7	15.9	13.8	9.48	15.6
R8G	17.5	16.3	7.71	20.2
K9	14.6	12.5	6.53	23.4
I10	11.2	7.5	5.71	22.0
I11	10.3	7.5	5.68	15.5
H12G	17.1	16.5	7.32	19.0
I13	11.1	7.5	7.53	17.4
I14	12.5	11.5	3.54	18.6
K15	24.3	24.4	9.98	20.5
K16	18.6	18.1	7.90	21.5

structure was minimized through a series of steps involving (a) energy minimization using steepest descent algorithm and (b) cyclic annealing between temperatures 298 and 450 K. These steps were followed by a 200 ns long equilibrium molecular dynamics (MD) simulation. The equilibrium MD simulations were performed at a temperature of 298 K and pressure of 1 bar using Berendsen coupling.[12] The average area per head group for equilibrated bilayer was found to be 0.68 nm^2.

The interaction of the ovispirin peptide and the lipid bilayer was studied using the "constraint method" as described in Trzesniak et al.[13] Briefly, in this method the distance between the center of mass of the peptide and the center of the bilayer is fixed and the value of force required to maintain this distance is recorded every 0.005 ps for a total of 5000 ps. The average of these recorded values, at different distances, was then integrated using Equation 4 to obtain the potential of mean force for the interaction of the peptide with the membrane. The potential of mean force is given by

$$\psi(\xi) = \int_{\xi_0}^{\xi} \langle f(\xi') \rangle d\xi' + k_b T \ln(\xi / \xi_0) \tag{4}$$

where ψ is the potential of mean force, ξ, denotes the distance between the peptide and the membrane, $\langle f(\xi') \rangle$, the average force at a particular value of separation, ξ_0, is the distance between peptide and membrane at zero P.M.F., k_b, the Boltzmann constant and T the temperature.

A1.4 Calculation of infrared spectra from MD trajectories

To calculate the 2D IR lineshape for each isolated amide I chromophore from the molecular dynamics (MD) simulations, the electric field components in the C=O direction on the

C and N atoms of each peptide bond, E_C and E_N, were computed with a cut-off radius of 20 Å every 2 fs. The contributions from atoms within three bonds were excluded, following the 1-4 exclusion principle adopted in the MD simulations.[14] An empirical equation is then used to calculate the frequencies from the electric field components:[15]

$$\omega_{10} = 1717 + 4213E_C + 2108E_N$$

(ω_{10} is in cm^{-1}; E_C and E_N are in atomic units). This frequency map has been applied to an isotope-labeled transmembrane peptide bundle CD3ζ and the calculated 2D IR diagonal widths are in good agreement with experiment.[15] Note that this map is for the deuterated amide I group, while the present experiments, in normal water, involve protonated amide I groups. We assume that this difference produces only a constant shift, which is not relevant for the linewidth calculations described herein. From the frequency trajectory obtained from a 2-ns long simulation, we calculated the frequency time-correlation function

$$C_{10}(t) = \langle \delta\omega_{10}(t)\delta\omega_{10}(0) \rangle, \tag{6}$$

where $\delta\omega_{10}(t)$ is the fluctuation of the instantaneous frequency from its average. The lineshape function $g(t)$ was calculated by its definition

$$g(t) = \int_0^t dt'\,(t-t')\,C_{10}(t'). \tag{7}$$

2D IR spectra were computed following Eqns. (1-3). The population relaxation times were set as T_{10} = 600 fs and T_{21} = 400 fs for all the calculations and an anharmonicity of Δ = 14 cm^{-1} was used. For each residue, 20 2-ns simulations were performed for wild type sequence. For the mutants, 10 2-ns simulations were performed. These simulations were used to calculate 2D-IR spectra, from which the diagonal widths were extracted. The diagonal widths averaged over those obtained from these 20 (wild type) or 10 (mutant) simulations are reported in Table S1.

A1.5 Lipid/Peptide Ratio Study

To ensure that a 21:1 lipid/peptide ratio was sufficient for the surface bound configuration and that ovispirin is binding as a monomer, three residues were measured at 21:1 and 42:1 lipid-to-peptide ratio. The residues were chosen to span the α-helix wheel diagram. Fig. A.1.2a shows the isotope labeled peaks for both the low and high lipid/peptide (L/P) ratio. Fits to the spectra are shown in Fig. A.1.2b. Peaks were characterized as above, and the diagonal linewidths are reported in Table A1.1. The diagonal linewidths are plotted in Fig. A.1.2c. Linewidths changes of less than 1.5 cm^{-1} were observed, which does not alter the observed trend in linewidths. Thus, we conclude that ovispirin is bound to the membrane as a monomer as expected.

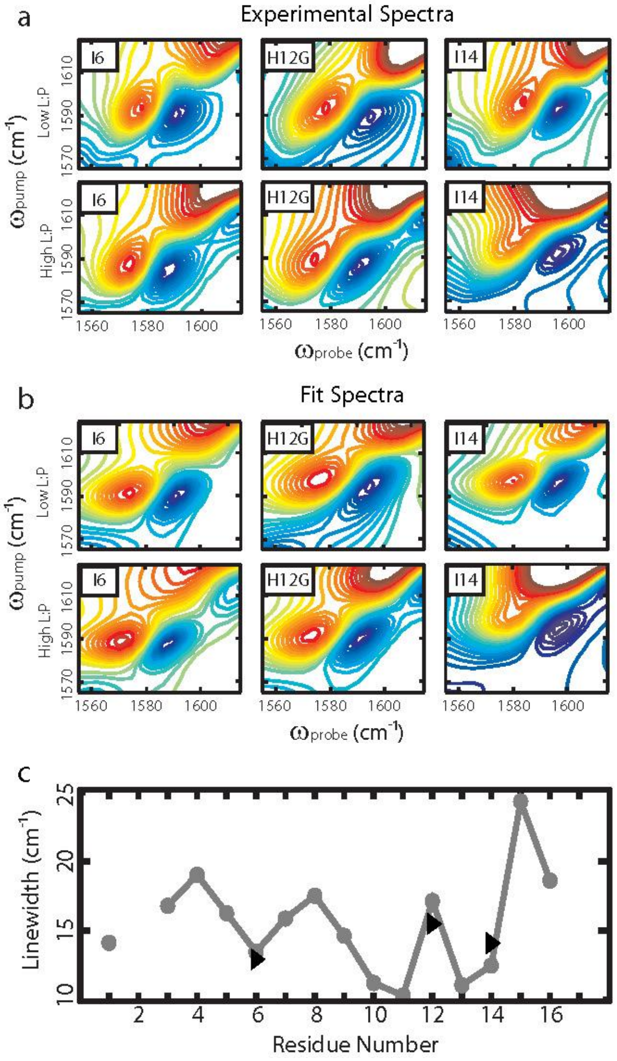


Figure A1.2. Comparison of labeled ovispirin data and fits for different L/P ratios. (a) Top panel is 2D IR isotope peaks reproduced from Figure A.1.1a at >21/1 L/P. Bottom panel is 2D spectra for the same isotope labels at >42/1 L/P. (b) 2D spectra calculated from parameters obtained from fits to spectra shown in (a). (c) 2D diagonal linewidth data reproduced for comparison from Fig. 3.2a with diagonal linewidths for samples with >42/1 L/P ratios included.

Table A1.2. 2D IR diagonal linewidth is invariant to increased L/P molar ratio.

L/P	Residue	Diagonal Linewidth, cm^{-1} ($\pm 3\ cm^{-1}$)
>21/1	I6	13.5
	H12G	17.1
	I14	12.5
>42/1	I6	12.9
	H12G	15.5
	I14	14.1

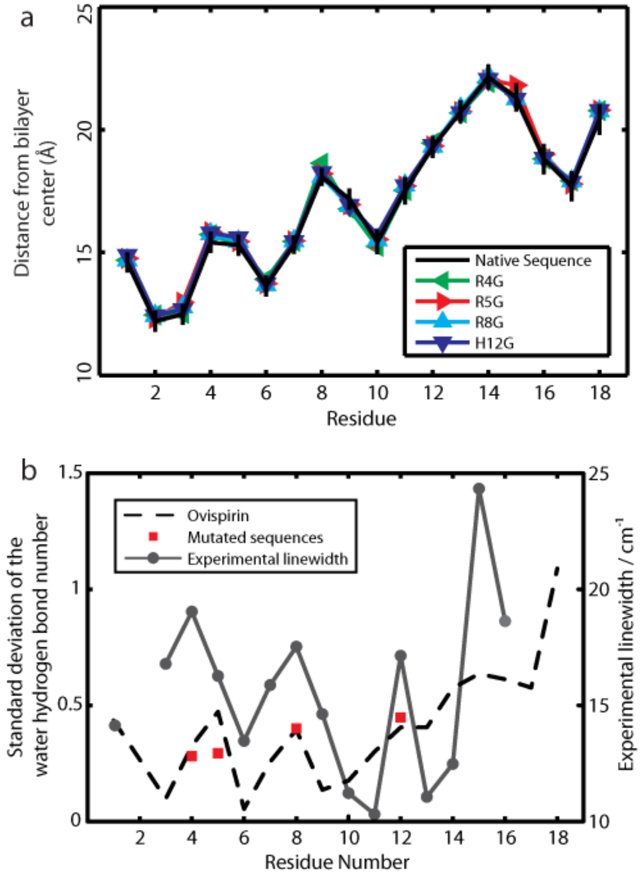


Figure A1.3. Comparison of MD simulations performed on native ovispirin and glycine mutants. (a) Distance from the center of the bilayer of backbone α-carbon atoms for each residue of native and mutant peptides. Error bars are the standard deviation in depth during the trajectory of the native sequence. (b) Standard deviation in number of water hydrogen bonds to backbone carbonyl groups for each residue (black dashed line) and at point of mutation from simulations of the four mutant peptides, R4G, R5G, R8G, and H12G (red squares). Linewidth data is reproduced from Fig. 3.2a (grey solid curve).

A1.6 References

(1) Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T. *Biopolymers* **2001**, *59*, 396-401.

(2) Cruz, L. J.; Beteta, N. G.; Ewenson, A.; Albericio, F. *Org. Process Res. Dev.* **2004**, *8*, 920-924.

(3) Wiejak, S.; Masiukiewicz, E.; Rzeszotarska, B. *Chem. Pharm. Bull. (Tokyo)* **1999**, *47*, 1489-1490.

(4) Yamaguchi, S.; Huster, D.; Waring, A.; Lehrer, R. I.; Kearney, W.; Tack, B. F.; Hong, M. *Biophys. J.* **2001**, *81*, 2203-2214.

(5) Mukamel, S. *Oxford Ser. Opt. Imaging Sci.* **1995**, *6*, 543 pp.

(6) Hamm, P.; Lim, M. H.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123-6138.

(7) Zanni, M. T.; Ge, N. H.; Kim, Y. S.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11265-11270.

(8) DeFlores, L. P.; Ganim, Z.; Ackley, S. F.; Chung, H. S.; Tokmakoff, A. *J. Phys. Chem. B* **2006**, *110*, 18973-18980.

(9) Van Gusteren, W. F. B., H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Groningen, The Netherlands, 1987.

(10) Tieleman, D.; Turner, R.; Vogel, H.; Weiner, J. *BBA-Biomembranes* **2008**, *1778*, 1697-1697.

(11) Zhao, W.; Rog, T.; Gurtovenko, A. A.; Vattulainen, I.; Karttunen, M. *Biophys. J.* **2007**, *92*, 1114-1124.

(12) Berendsen, H.; Postma, J.; Van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684.

(13) Trzesniak, D.; Kunz, A. P. E.; van Gunsteren, W. F. *Chemphyschem* **2007**, *8*, 162-169.

(14) Torii, H. *J. Phys. Chem. B* **2007**, *111*, 5434-5444.

(15) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2009**, *113*, 592-602.

APPENDIX 2

Supporting Information for

CHAPTER 4

Parallel β-sheet vibrational couplings revealed by 2D IR spectroscopy of an isotopically labeled macrocycle: Quantitative benchmark for the interpretation of amyloid and protein infrared spectra.*

A2.1 Peptide Synthesis

A2.1.1 Synthesis of Peptide 1 (Unlabeled)

Synthesis of **1** began with the preparation in solution of the fragment Alloc-Glu-Val-^{D}Pro-Dadme-Fmoc, as previously described.[1] The fragment was attached to Rink amide resin by the Glu side chain carboxyl group, and the amino acid residues corresponding to Lys$_{11}$ through Gly$_{19}$ were appended via standard Fmoc-based solid-phase peptide synthesis procedures. Fmoc-amino acids were activated as HOBt esters with HBTU and DIEA. All amino acids with reactive side chains were protected. The α-amino group of Glu$_{14}$ was then coupled with HO-Succinic-

---

* This chapter will be submitted to *J. Am. Chem. Soc*. in April 2012. It was prepared in collaboration with A. M. Almeida, L. Wang, C.-C. Chiu, M. McGovern, J. J. de Pablo, J. L. Skinner, S. H. Gellman, and M. T. Zanni.

1

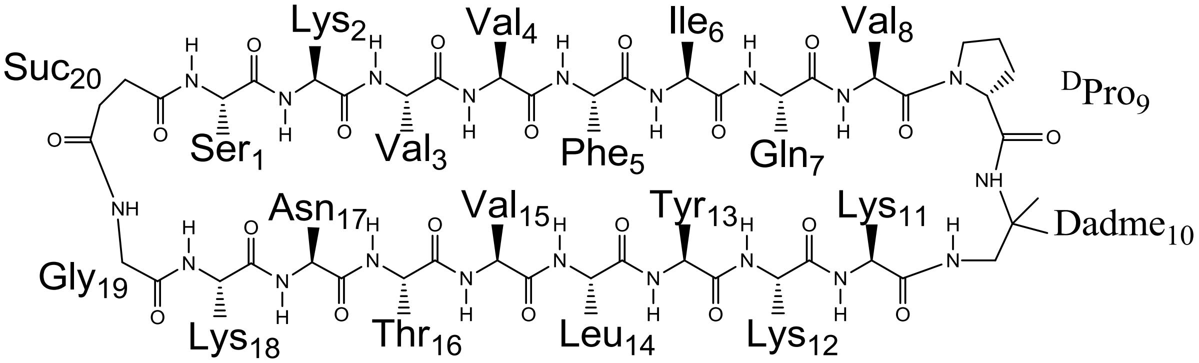


Figure A2.1. Cyclic Peptide: **1** (AMA_5_13).

OTMSE (prepared as previously described)[1]. The alloc protecting group was removed with a Pd^0 catalyst, and the amino acid residues corresponding to Ile$_6$ through Ser$_1$ were introduced via standard methods. The α-carboxyl group of Suc$_{20}$ was deprotected using tetrabutylammonium fluoride (TBAF). Coupling between the α-carboxyl group of Suc$_{20}$ and the α-amino group of Ser$_1$ was carried out using 1.1 eq. PyBOP, 2 eq. DIEA. cyclization was monitored via analytical HPLC. Upon completion of synthesis the resin was dried under vacuum. Resin cleavage and global deprotection were achieved by reacting the resin with 95 % TFA, 2.5 % H_2O, 2.5 % triisopropylsilane (v/v/v) for 3 hours, followed by peptide precipitation in cold ether. The precipitate was collected by centrifugation/decantation prior to HPLC purification.

A2.1.2 Synthesis of Peptide 2 (Dual $^{13}C{=}^{18}O$ carbonyl labeled)

Synthesis of **2** (Fig. A2.2) is shown as representative of synthesis of macrocyclic peptides incorporating isotope labeled amino acids (Fig. A2.3). Synthesis of **2** began with the preparation in solution of the fragment Alloc-Glu-Val-^{D}Pro-Dadme-Fmoc, as previously described.[1] The fragment was attached to Rink amide resin by the Glu side chain carboxyl group, and the amino acid residues corresponding to Lys$_{11}$ through Gly$_{19}$ were appended via standard Fmoc-based solid-phase peptide synthesis procedures. Fmoc-amino acids were activated as HOBt esters with HBTU and DIEA. All amino acids with reactive side chains were protected. The α-amino group of Glu$_{14}$ was then coupled with HO-Succinic-OTMSE (prepared as previously described)[1]. The alloc protecting group was removed with a Pd^0 catalyst, and then the α-amino group of Gln$_7$ was coupled with 1.1 eq. of Fmoc-Ile($^{13}C{=}^{18}O$)-^{18}OH (**3**) with 4 eq. HOBt and 8 eq. DIEA was coupled for an extended period of time (25 min, at 70°C w/ microwave irradiation). After removal of the Fmoc protecting group, the α-amino group of Ile$_8$ was coupled with 1.1 eq. of

2

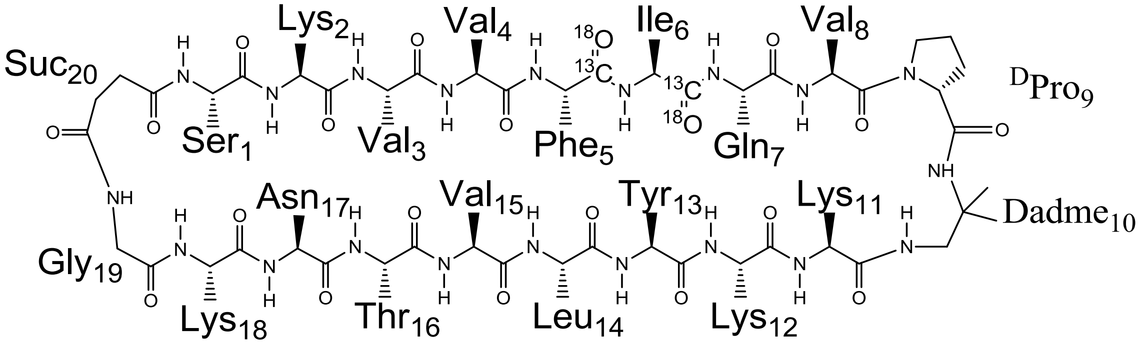
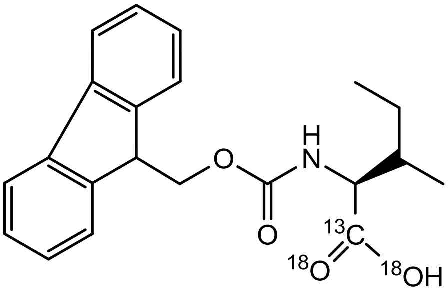


Figure A2.2. Cyclic Peptide: **2**.

3



4

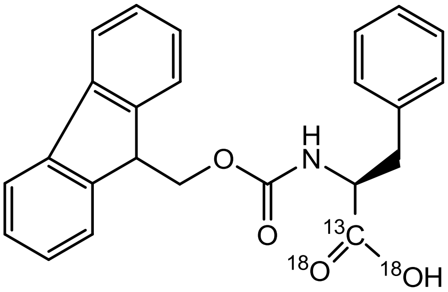


Figure A2.3. Isotope Labeled Amino Acids: **3** and **4**.

Fmoc-Phe(^{13}C=^{18}O)- ^{18}OH (**4**) with 4 eq. HOBt and 8 eq. DIEA was coupled for an extended period of time (25 min, at 70°C w/ microwave irradiation). After Fmoc removal the residues corresponding to Val_4 through Ser_1 were introduced via standard methods. The α-carboxyl group of Suc_{20} was deprotected using tetrabutylammonium fluoride (TBAF). Coupling between the α-carboxyl group of Suc_{20} and the α-amino group of Ser_1 was carried out using 1.1 eq. PyBOP, 2 eq. DIEA. Upon completion of synthesis the resin was dried under vacuum. Resin cleavage and global deprotection were achieved by reacting the resin with 95% TFA, 2.5% H_2O, 2.5% triisopropylsilane (v/v/v) for 3 hours, followed by peptide precipitation in cold ether. The precipitate was collected by centrifugation/decantation prior to HPLC purification.

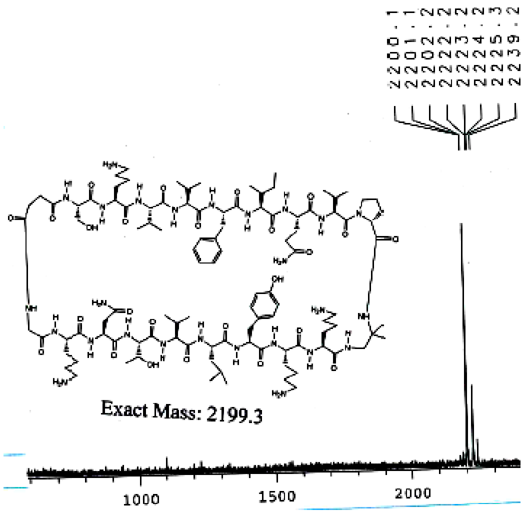
A2.2 Purification and Characterization of Peptides

All of the peptides were purified to homogeneity (≥ 95 %) on preparative scale Shimadzu HPLC using 15-35 % **B** solvent gradient over 40 minutes. Purity was assessed using an analytical scale Shimadzu HPLC using a 10-60 % **B** gradient over 50 minutes. Matrix assisted laser desorption ionization (MALDI) spectrometry was carried out using a Bruker REFLEX® (purchased with support of NSF Grant #9520868) **A** solvent in HPLC denotes H_2O:F_3CO_2H 100:0.01 v/v. **B** solvent denotes CH_3CN:F_3CO_2H 100:0.01 v/v. **Abu** denotes 2-Aminubutyric Acid.

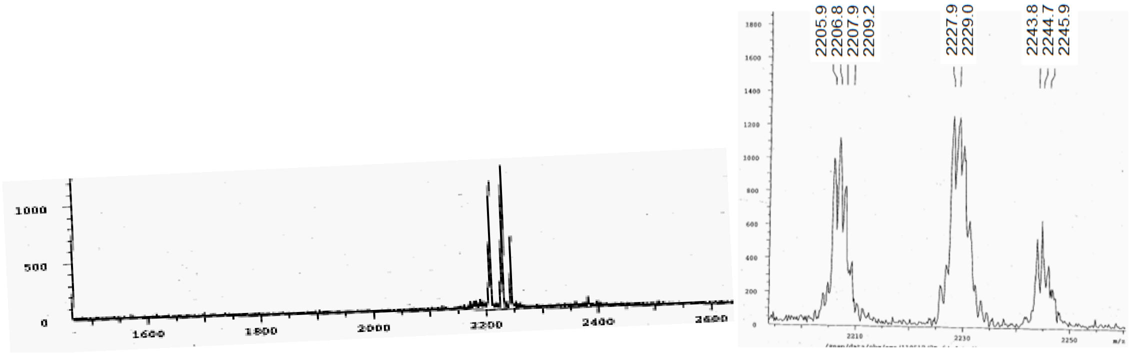
Table A2.1. Summary of HPLC & MALDI-TOF Characterization of Peptides.

Peptide	**% B Elution**	**Calculated Mass $[M+H]^+$**	**Observed $[M+H]^+$**	**Peptide**
1	21 %	2200.3	2200.1	**1**
2 (Ile$_5$-Phe$_6$) i→i+1	21 %	2206.3	2206.0	**2** (Ile$_5$-Phe$_6$) i→i+1
3 (Val$_4$-Phe$_6$) i→i+2	21 %	2206.3	2205.9	**3** (Val$_4$-Phe$_6$) i→i+2
4 (Val$_3$-Phe$_6$) i→i+3	21 %	2206.3	2206.4	**4** (Val$_3$-Phe$_6$) i→i+3
5 (Phe$_6$-Leu$_{14}$) i→j	21 %	2206.3	2206.2	**5** (Phe$_6$-Leu$_{14}$) i→j
6 (Ile$_5$-Leu$_{14}$) i→j+1 (in)	21 %	2206.3	2206.6	**6** (Ile$_5$-Leu$_{14}$) i→j+1 (in)
7 (Phe$_6$-Val$_{15}$) i→j+1 (out)	21 %	2206.3	2206.4	**7** (Phe$_6$-Val$_{15}$) i→j+1 (out)

Peptide 1



Peptide 3



Peptide 6

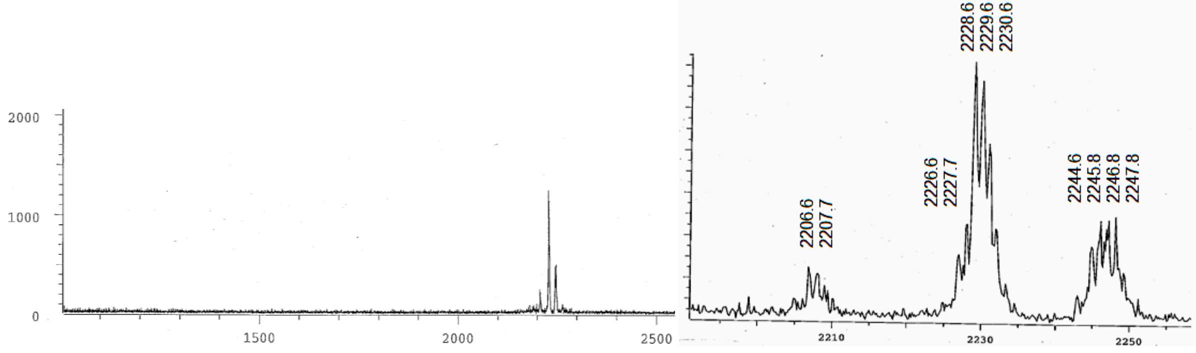
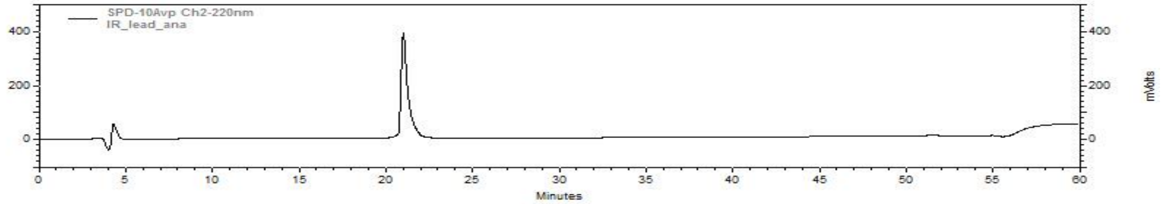
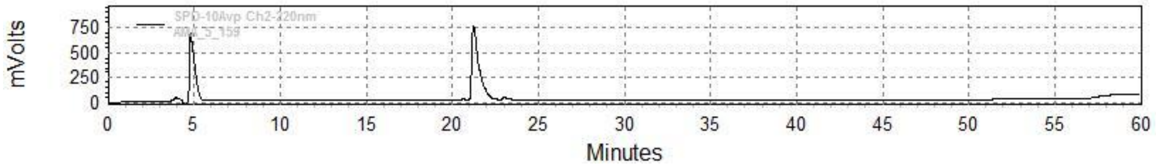


Figure A2.4. Sample MALDI spectra.

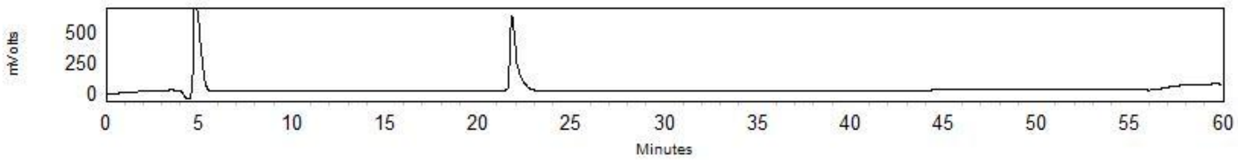
Analytical HPLC trace for ***1***



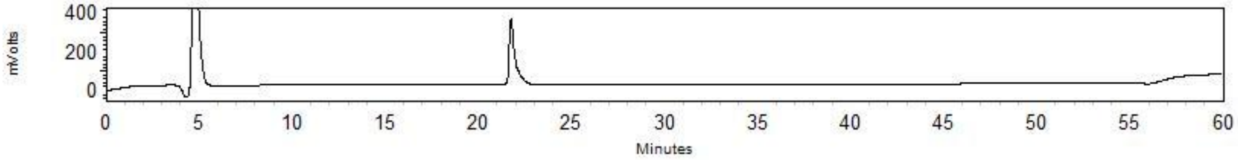
Analytical HPLC trace for ***2***



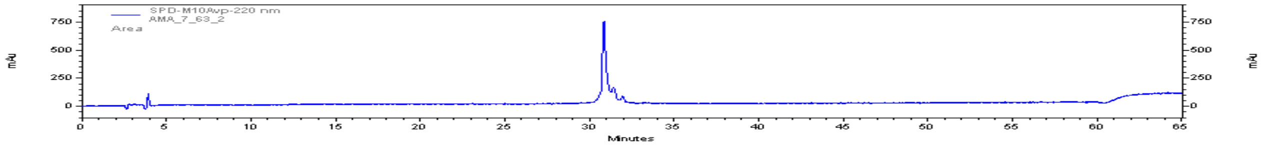
Analytical HPLC trace for ***3***



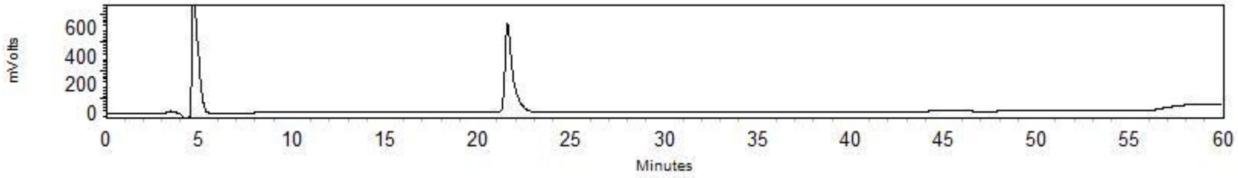
Analytical HPLC trace for ***4***



Analytical HPLC trace for ***5***



Analytical HPLC trace for ***6***



Analytical HPLC trace for ***7***

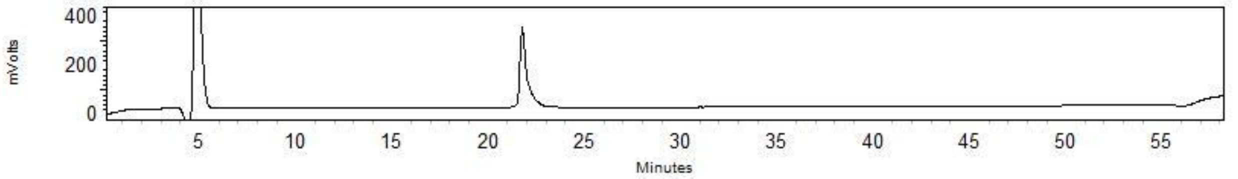


Figure A2.5. Analytical HPLC traces (10-60 % B over 50 min) unless stated otherwise.

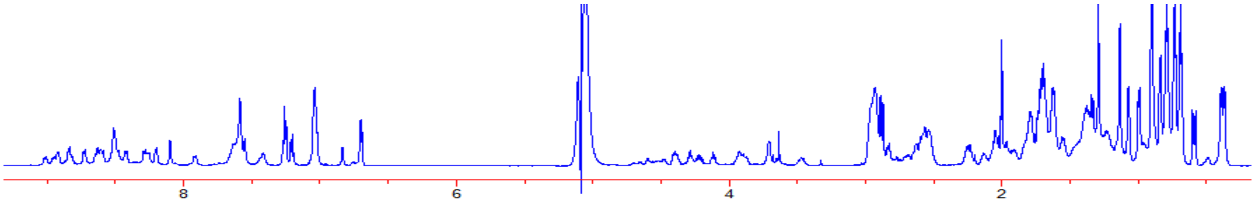


Figure A2.6. 1D-^{1}H-NMR of **1** (2.5 mM peptide in 9:1 H_2O:D_2O, pH 3.80, 100 mM sodium deuterioacetate buffer (buffer pH was not corrected for isotope effects)

A2.3 NMR Sample Preparation, Aggregation Analysis and Data Acquisition

NMR experiments were performed on a Varian INOVA 600 MHz spectrometer (Purchased with support of NIH Grant # 1 S10 RR13866-01) at 277 K using a 3 mm Varian $^1H/^{13}C/^{15}N$ probe with 3 axis PFG. The reported temperatures are presumed to be accurate to ≈ 1 K. Samples were prepared with total volumes of approximately 300 μL in 3 mm tubes.

A2.3.1 NMR Acquisition

NMR sample were prepared by dissolving lyophilized peptides in 9:1 H_2O:D_2O, pH 3.80, 100 mM sodium deuterioacetate buffer (buffer pH was not corrected for isotope effects). Peptide concentration was ~2.5 mM. Samples were stable in solution for weeks, showing no apparent precipitation of peptide or decrease in NMR signal intensity over the entire period of study. In all cases sharp lines were observed in 1D spectra, suggesting that the peptides were not aggregated in solution. A variable concentration study showed no change in the 1D spectra of the peptide between 2.5 mM and 0.04 mM peptide concentration under NMR conditions.

Trace amounts of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was added as an internal chemical shift reference. Solvent suppression was achieved by 0.6-1.5 second solvent presaturation during the relaxation delay. Spectral windows of 6600 Hz were used. Standard Varian pulse sequences were used, and data were processed using Varian VNMR 6.1 software and analyzed with the sparky program.[1] Shifted sine bell window functions were generally applied before Fourier transformation. For all samples, GCOSY spectra were obtained in absolute mode with gradient echo coherence selection; TOCSY[2] and ROESY[3] spectra were acquired in the sensitive mode with hypercomplex phase cycling (States-Haberkorn method). Watergate solvent suppression, as implemented in Varian´s Protein Pack suite of experiments, was used instead of presaturation in ROESY and TOCSY experiments. All experiments were performed by collecting 2048 points in f2 and 300-600 points in f1. TOCSY experiments employed a standard MLEV-17 spin lock sequence with a spin lock field of 7-8 KHz and mixing time of 80 ms. ROESY experiments used spinlocking fields of 3 kHzs and mixing times of 200-250 ms. The ^{1}H chemical shift assignment of the peptides was achieved by the sequential assignment procedures.[4]

A2.3.2 NMR Structure Ensemble Generation

Distance constraints were derived from the intensities of NOE cross-peaks in 600 MHz ROESY spectra with 200 ms mixing time at 277 K. Using the $I = c.r^{-6}$ approximation[5] (where I is the NOE intensity, r is the distance between protons, and c is a constant) and the NOE between Tyr aromatic ring protons as an internal standard, the interactions were grouped into very strong (≤ 2.5 Å), strong (≤ 3.0 Å), medium (≤ 3.5 Å), and weak (≤ 4.5 Å) categories. NMR structure ensembles were generated using CNS v1.1.[6] The constraints employed for the ensembles are summarized in Supplementary Section **S4**. Default CNS simulated annealing script, which includes high-temperature dynamics and then a cooling cycle in torsion space (with *KNOE*= 150) followed by further cooling in Cartesian space (*KNOE* = 50) and minimization (*KNOE* = 75), was modified in the following ways: (a) the high-temperature torsion dynamics and cooling phases were each doubled in length (2000 15-fs steps each, initially 50 000 K with cooling during the last 2000 steps); (b) the Cartesian cooling cycle (starting at 2000 K, *KNOE* = 75 or 50 Kcal/Å^2) was doubled in length (6000 5-fs steps), which allowed us to ramp the Erepel scaled factor from 0.2 to 1.0 during the cycle; (c) chemical shift and H-bond constraints were not employed, and the dihedral angle energy function was disabled. In the present study, 500 random structures were calculated, and the 10 lowest energy structures were selected.

A2.4 Observed NOE's and MD statistics of 5, and P-cyc

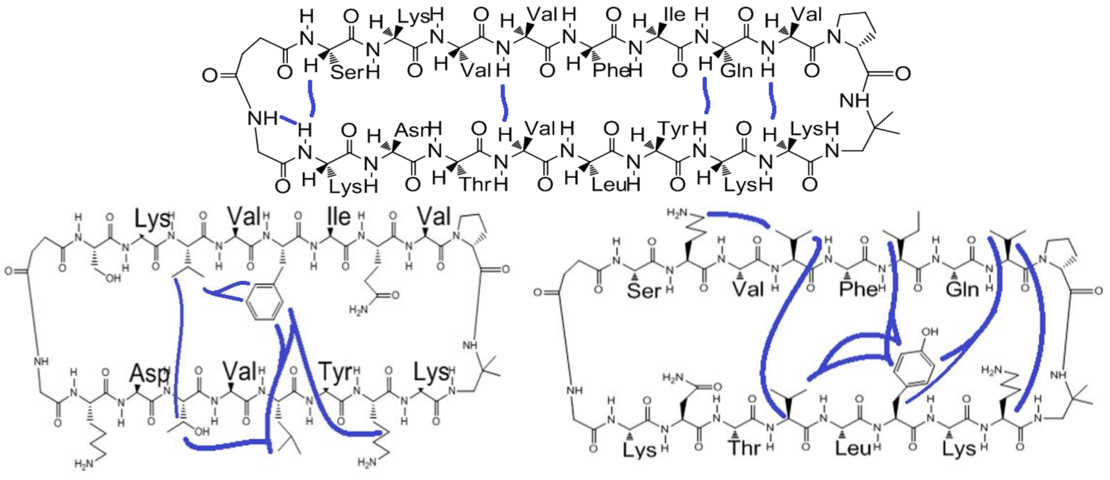


Figure A2.7. Observed Long Range NOE’s for **1**.

Table A2.2. Observed Long Range NOE’s for **1**.

Long Range NOE's		**NOE Strength**
Ser(01) HA	Lys(18) HN	Strong
Val(04) HN	Val(15) HA	Strong
Gln(08) HA	Lys(12) HN	Strong
Val(08) NH	Lys(12) HA	Strong
Asn(17) HA	Val(04) HN	Medium
Val(15) HA	Ile(06) HN	Medium
Ile(06) HB	Tyr(13) HE	Strong
Ile(06) HB	Tyr(13) HD	Strong
Lys(12) HG	Phe(05) HD	Weak
Leu(14) HB	Phe(05) HE	Medium
Val(08) HG	Tyr(13) HD	Medium
Val(08) HG	Tyr(13) HE	Weak
Val(08) HG (or) Ile(06) $HGCH_3$	Tyr(14) HE	Medium
Val(03) HG	Phe(05) HD	Medium
Val(03) HG	Phe(05) HE	Weak
Val(15) HG	Tyr(13) HB	Weak
Val(08) HG	Tyr(13) HB	Medium
Leu(14) HD	Phe(05) HE	Medium
Leu(14) HD	Phe(05) HZ	Weak
Leu(14) HD	Thr(16) HB	Medium
Lys(02) HE (or) Lys(18) HE	Val(04) HG	Medium
Lys(11) HA	Val(08) HB	Strong
Thr(16) HB	Val(03) HG	Weak
Lys(12) HD	Phe(05) HE	Weak
Val(08) HG	^{D}Pro(09) HD	Medium

Table A2.3. NMR and Refinement Statistics for **1**.

NMR distance and dihedral constraints	**Protein**
Distance constraints	
Total NOE	41
Intra-residue	0
Inter-residue	41
Sequential ($\|i - j\| = 1$)	17
Medium-range ($\|i - j\| < 4$)	5
Long-range ($\|i - j\| > 5$)	19
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	
ϕ	0
ψ	0
Structure statistics	
Violations (mean and s.d.)	0
Distance constraints (Å)	0
Dihedral angle constraints (°)	0
Max. dihedral angle violation (°)	0
Max. distance constraint violation (Å)	0
Deviations from idealized geometry	
Bond lengths (Å)	± 0.0059
Bond angles (°)	± 0.885
Impropers (°)	± 0.448
Average pairwise r.m.s. deviation (Å) (10 Structures)	
Heavy	1.8 ± 0.9
Backbone	0.96 ± 0.47

A2.5 Chemical Shift Tables

For all chemical shift tables, the following labels were used to denote proton identities for 1,2-diamino-1,1-dimethylethane (Dadme) within the peptides:

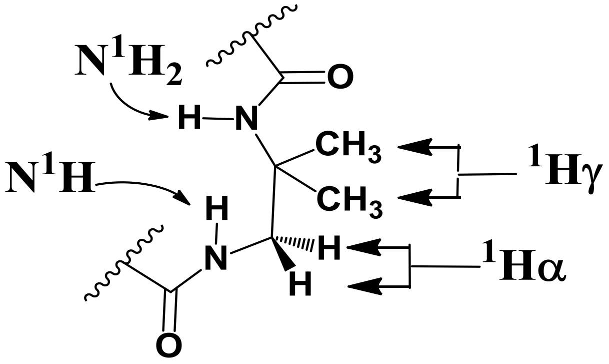


Table A2.4. Proton resonances (ppm) for 1 in 9:1 H2O/D2O, pH=3.8, 100 mM d3-acetate (4oC), 2.5mM peptide.

Res #	**aa**	**N^{1}H**	**^{1}Hα**	**^{1}Hβ**	**^{1}Hγ**	**^{1}Hδ**	**^{1}Hε**	**^{1}HNε**
1	**Ser**	8.538	4.628	3.734				
2	**Lys**	8.32	4.507	1.816,1.712	1.349	1.641	2.944	7.624
3	**Val**	8.29	4.426	1.703	0.759,0.391			
4	**Val**	8.515	4.139	1.866	0.811			
5	**Phe**	8.622	5.325	2.85,2.666		7.05	7.293	7.23
6	**Ile**	8.854	4.24	1.509	CH_3=0.927, CH_2=0.8254 ,0.982	0.7169		
7	**Gln**	8.955	4.875	2.081,1.998	2.259	7.619		
8	**Val**	9.049	4.624	2.075	1.103,0.9283			
9	**^{D}Pro**		4.335	2.288,1.94	2.158,2.014	3.903,3.746		
10	**Dadme**	7.952	3.498, 3.674	[N^1H_2 = 8.134]	1.32, 1.176			
11	**Lys**	8.873	4.735	1.82	1.573,1.417	1.731	3.003	7.678
12	**Lys**	8.987	4.686	1.699,1.575	1.258	1.454	2.583	7.45
13	**Tyr**	8.761	5.403	2.866,2.787		7.079	6.718	
14	**Leu**	8.652	4.543	1.409,1.2	1.309	0.764,0.8619		
15	**Val**	8.453	4.585	1.72	0.7166,0.420			
16	**Thr**	8.557	4.406	3.959	1.098			
17	**Asn**	8.762	5.135	2.614	7.582,6.857			
18	**Lys**	8.228	4.436	1.836,1.766	1.385	1.65	2.959	7.614
19	**Gly**	8.54	3.941					
20	**Suc**		CH_2=2.568,	CH_2=2.721				

Table A2.5. Carbon resonances (ppm) for 1 in 9:1 H2O/D2O, pH=3.8, 100 mM d3-acetate (4oC), 2.5mM peptide.

Res #	**aa**	**^{13}Cα**	**13Cβ**	**13Cγ**	**13Cδ**	**13C-ε**
1	**Ser**	58.1	63.7			
2	**Lys**	55.6		21.2		42
3	**Val**	61.5 (ovlap)				
4	**Val**	60.8	34.5			
5	**Phe**	56.4	40.9			
6	**Ile**	60.1	39.4			
7	**Gln**	55.4	30	34.4		
8	**Val**	59.2	33			
9	**^{D}pro**	64.4	31.3	27.6	51.4	
10	**Dad**	47.7				
11	**Lys**	57				42
12	**Lys**	54.3	35.7			41,7
13	**Tyr**	57.4	39.8			
14	**Leu**	53.3	44.8			
15	**Val**	61.3				
16	**Thr**	61.7*(ovlap)	70.9			
17	**Asn**	unob	39.8			
18	**Lys**	55.8				42
19	**Gly**	45.2				
20	**Suc**					

A2.6 2D IR Data Fitting

2DIR diagonal spectra ($\omega_{pump}=\omega_{probe}$) were fit to 3 gaussian curves using a least-squares fitting routine. The resulting 3 gaussian curves, their sum, and the experimental data are shown in Fig. A2.8. The fit parameters are plotted in Fig. A2.9.

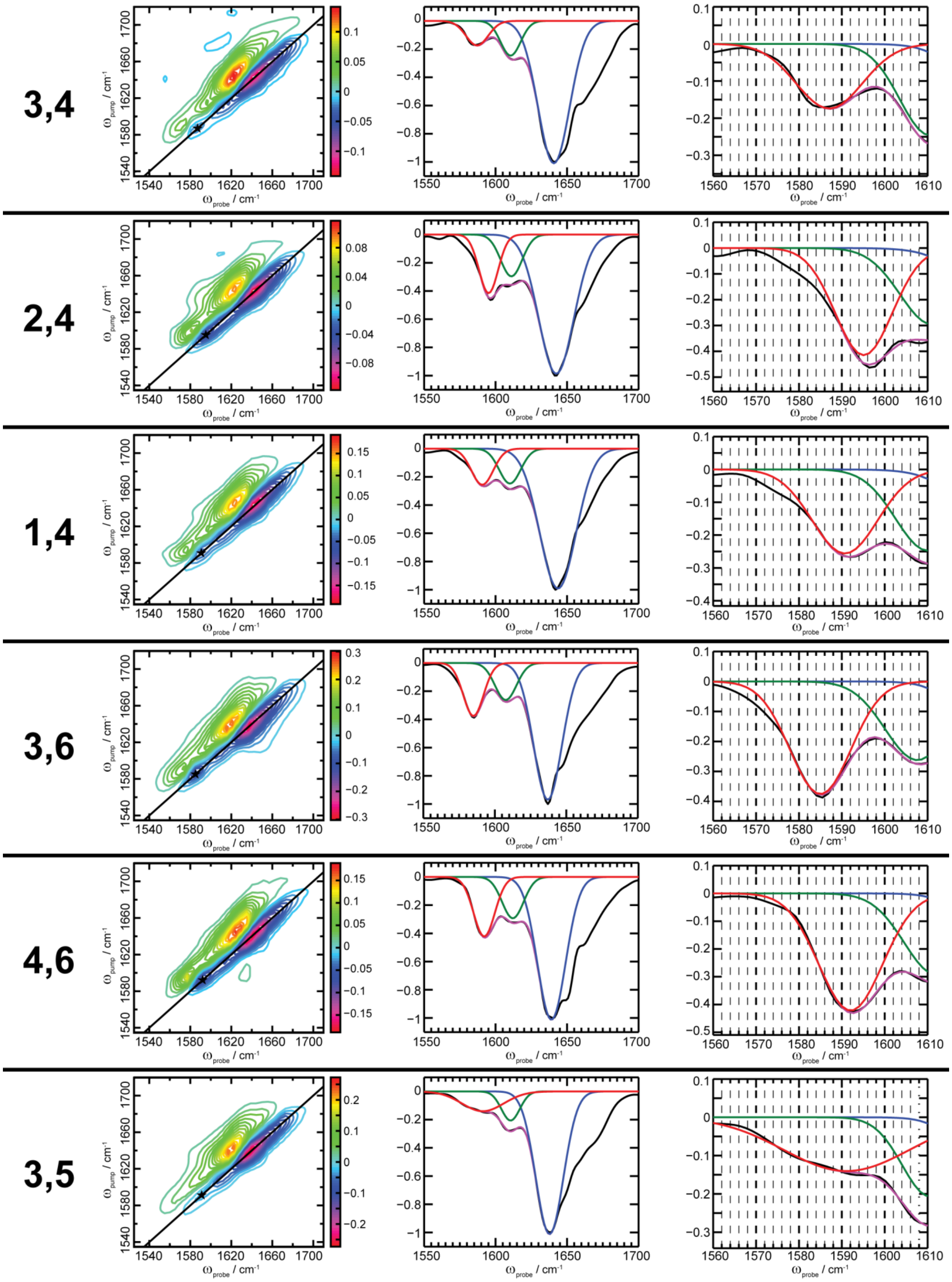


Figure A2.8. 2DIR diagonal spectra for each isotope labeled pair and the result of fitting to three gaussians.

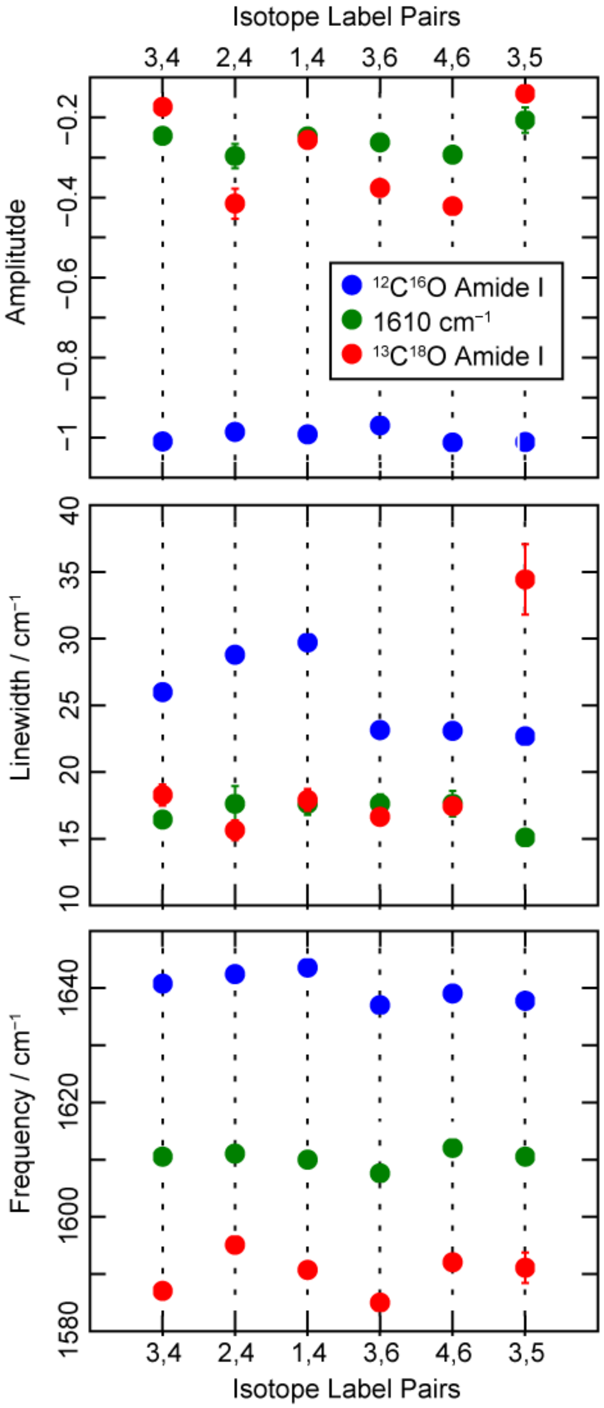


Figure A2.9. 2DIR diagonal spectra for each isotope labeled pair and the result of fitting to three gaussians.

Table A2.6. Population lifetimes from single exponential fit to the decay of 2DIR intensities as waiting time is scanned.

Isotope Labeled Pair	**Unlabeled T_1 / fs**	**$^{13}C^{18}O$ T_1 / fs**
3.4	452	1075
2.4	332	493
1.4	378	385
3.6	216	528
4.6	314	413
3.5	417	365

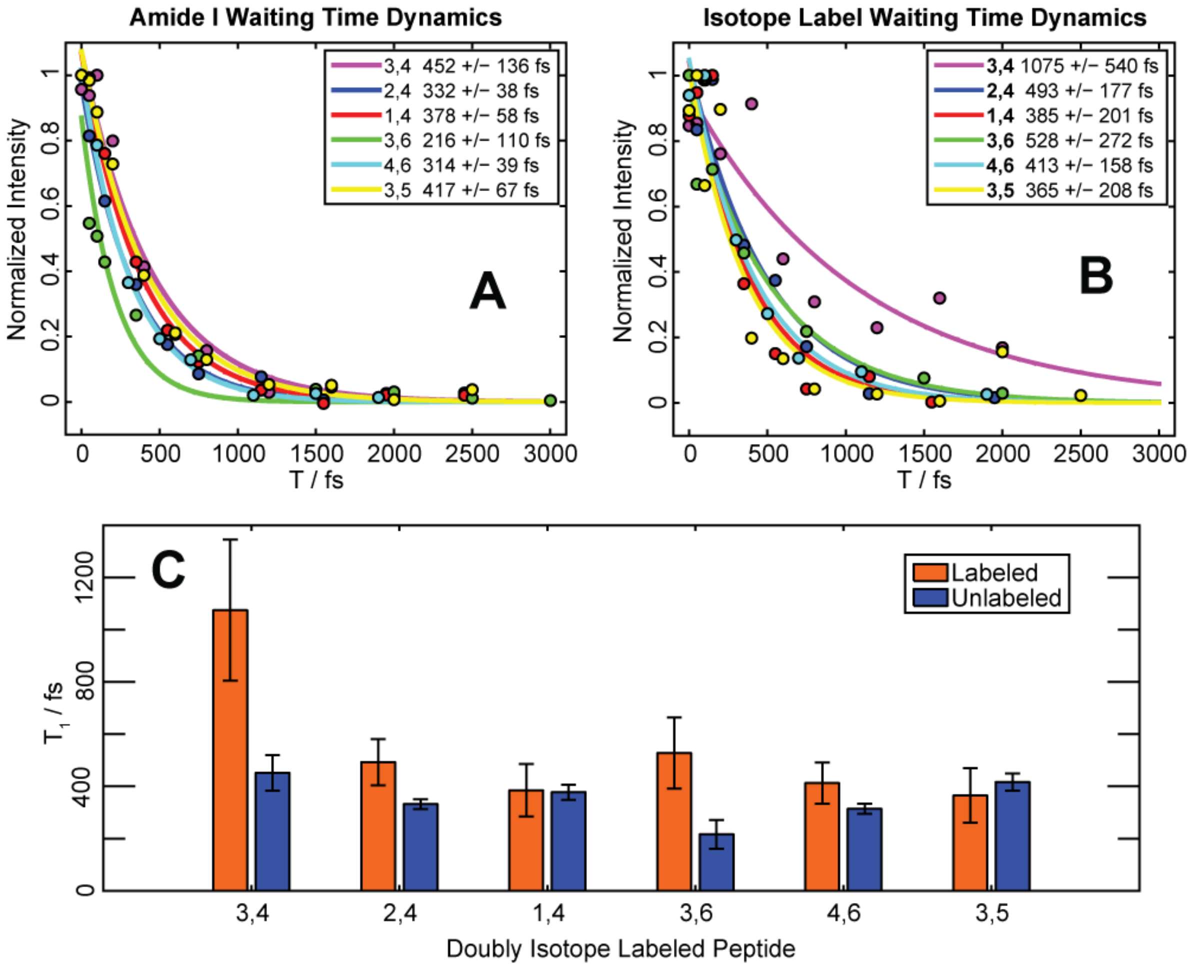


Figure A2.10. Decay of peak intensities over the scan of waiting time fit to single exponential curves for the unlabeled amide I band (A) and the $^{13}C^{18}O$ isotope labeled peak (B).

A2.7 MD Simulations and 2DIR Calculations

REMD simulations were performed with a temperature range of 298 to 380 K. Stable configurations at four different temperatures 320 K, 340 K, 360 K and 380 K were extracted using cluster analysis. Four configurations were extracted for 298 K. 2DIR spectra of the six labeled pairs were calculated starting from the configurations extracted from REMD simulations. The peak positions and diagonal linewidth were extracted and listed in Table A2.7.

From Table A2.7, the 2DIR peak positions and diagonal linewidths do not vary much with respect to temperature, verifying that the macrocycle structure is stable in the simulation.

Table A2.7. Peak frequency and width from REMD simulations

label	Normal simulation	298K_1	298K_2	298K_3	320K	340K	360K	380K
3,4	1583.6	1583.6	1583.2	1584.0	1584.4	1584.4	1584.8	1585.2
2,4	1588.9	1588.1	1587.7	1589.3	1588.9	1588.9	1588.5	1588.5
1,4	1585.6	1585.2	1584.8	1585.2	1585.2	1586.0	1586.4	1587.2
3,6	1577.5	1577.9	1578.3	1578.3	1578.3	1578.3	1579.5	1580.3
4,6	1585.2	1584.0	1584.4	1584.0	1585.6	1585.6	1586.0	1586.4
3,5	1582.4	1584.0	1582.0	1584.0	1581.6	1582.4	1583.6	1584.8

label	Normal simulation	298K_1	298K_2	298K_3	320K	340K	360K	380K
3,4	16.5	16.3	15.3	16.0	15.7	15.7	15.5	14.4
2,4	15.4	14.0	14.9	14.8	15.6	14.2	14.4	13.1
1,4	16.8	16.7	15.7	15.7	17.1	15.4	16.2	14.7
3,6	16.3	16.9	15.6	16.4	16.2	15.1	15.4	13.9
4,6	15.9	14.7	15.2	14.9	15.7	14.9	14.5	13.4
3,5	20.3	20.1	18.1	19.5	20.3	18.5	18.1	17.2

A2.8 References

(1) Goddard, T. D.; Kneller, D. G.

(2) Bax, A.; Davis, D. G. *Journal of Magnetic Resonance (1969)* **1985**, *65*, 355.

(3) Bothner-By, A. A.; Stephens, R. L.; Lee, J.; Warren, C. D.; Jeanloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811.

(4) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons, 1986.

(5) Bolm, C.; Atodiresei, I.; Schiffers, I. *Org. Synth.* **2005**, *82*, 120.

(6) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *D54*, 905.

APPENDIX 3

Supporting Information for

CHAPTER 5

Synthesis of Cysteine-Selective Cyclopentadienyl Rhenium Carbonyl Labels for FTIR and 2D IR Probes of Protein Structure[*]

A3.1 Characterization of $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$, $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$, $[\eta^5\text{-}(C_5H_4)CH_2I]Re(CO)_3$, $[\eta^5\text{-}(C_5H_4)CH_2SSO_2CH_3]Re(CO)_3$, V71C *α*–synuclein unlabeled and labeled with $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$

---

[*] This chapter will be submitted to *Organometallics* in April 2012. It was prepared in collaboration with D. Skoff, S. D. Moran, M. T. Zanni.

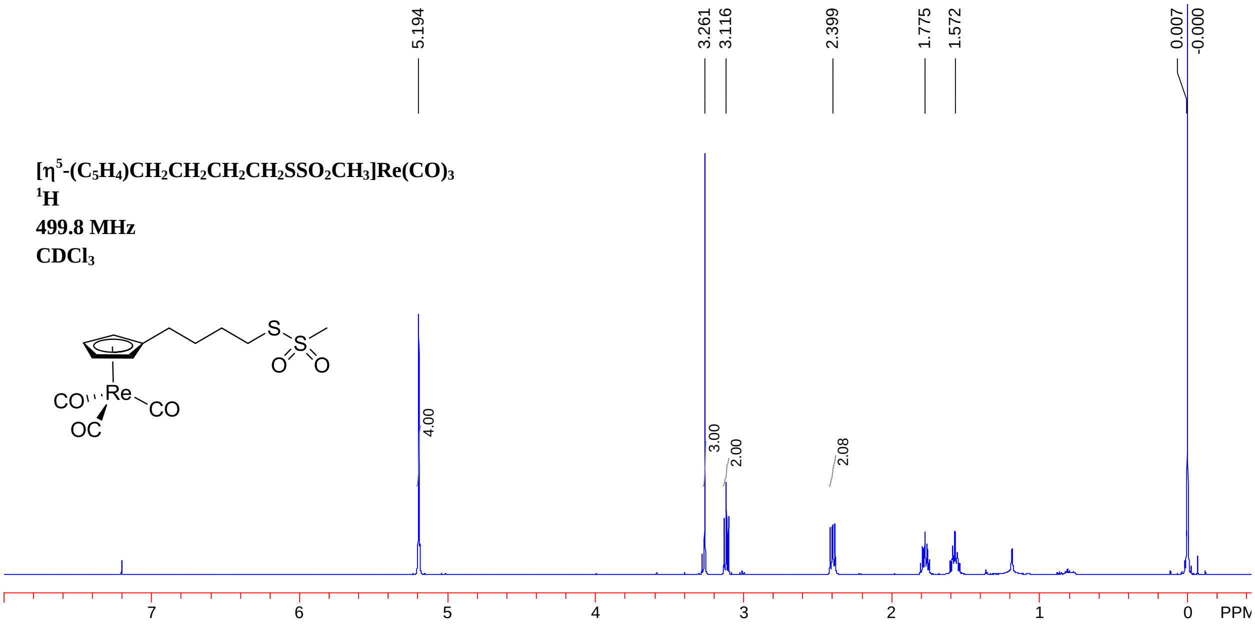


Figure A3.1. $[\eta^5\text{-}(C_5H_4)CH_2CH_2CH_2CH_2SSO_2CH_3]Re(CO)_3$ ^1H NMR: (499.8 MHz, $CDCl_3$) δ 5.195 (Cp, m, 4H), 3.262 (-CH_3, s, 3H), 3.116 (-CH_2S-, t, J = 7.3 Hz, 2H), 2.399 (CpCH_2-, m, 2H), 1.775 (CH_2CH_2S-, p, J = 7.6 Hz, 2H), 1.572 (CpCH_2CH_2-, m, 2H).

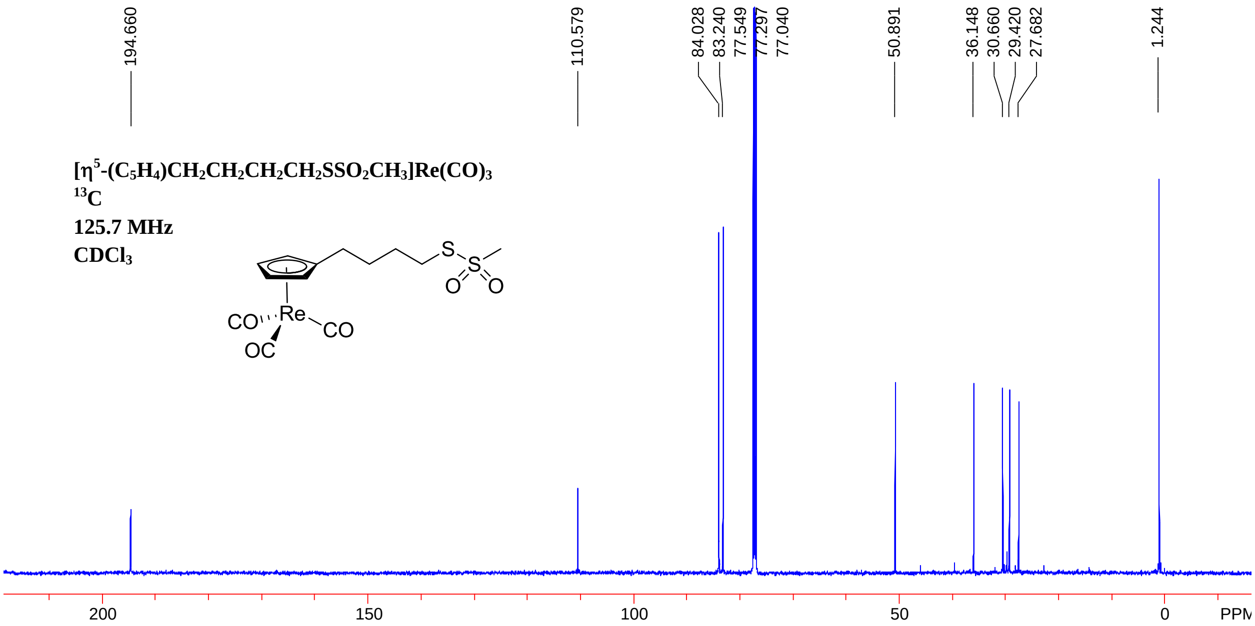


Figure A3.2. [η^5-(C_5H_4)$CH_2CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ ^{13}C NMR: (125.7 MHz, $CDCl_3$) δ 194.7 (*C*O), 110.6 (*ipso*-Cp), 84.0 (Cp), 83.2 (Cp), 50.9 (-*C*H_3), 36.1 (-*C*H_2S-), 30.7 (CpCH$_2$*C*H$_2$-), 29.4 (-*C*H$_2$CH$_2$S-), 27.7 (Cp*C*H$_2$-).

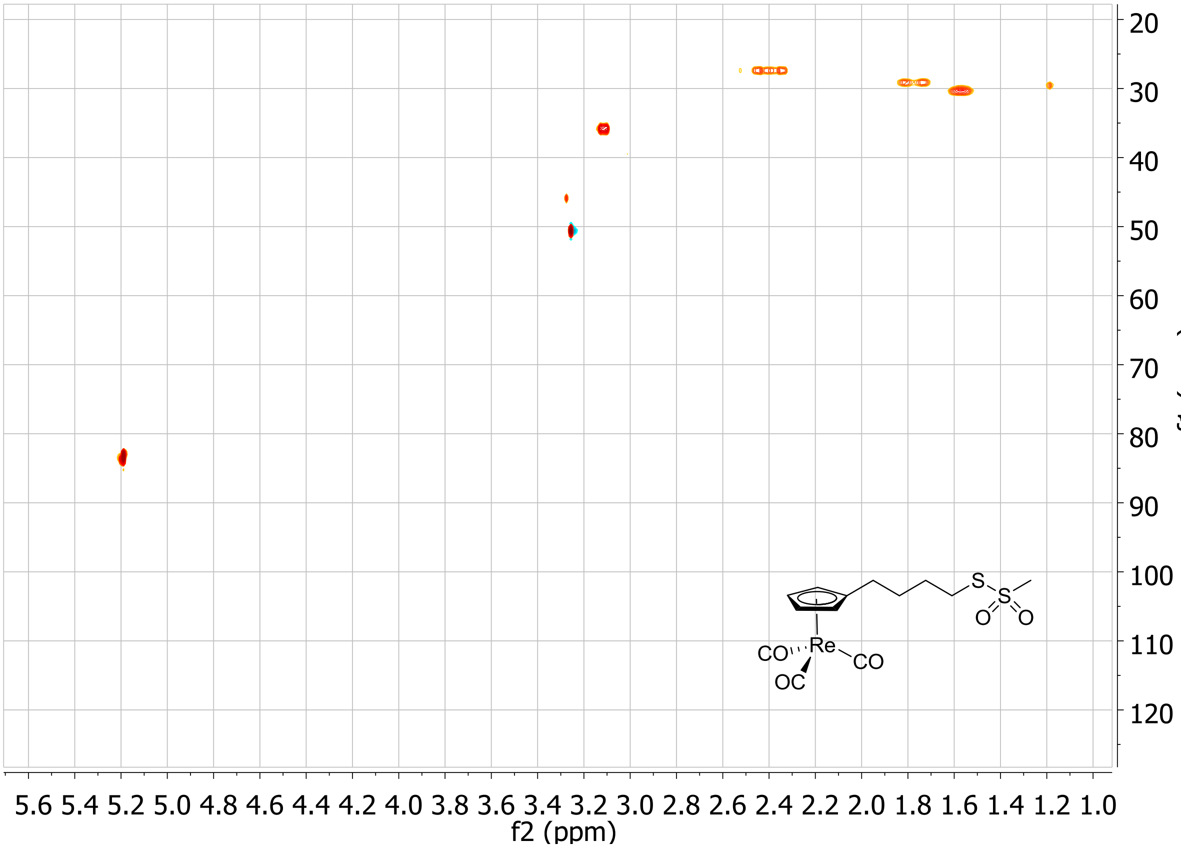


Figure A3.3. [η^5-(C_5H_4)$CH_2CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ HSQC Spectrum, 500.2 MHz, $CDCl_3$

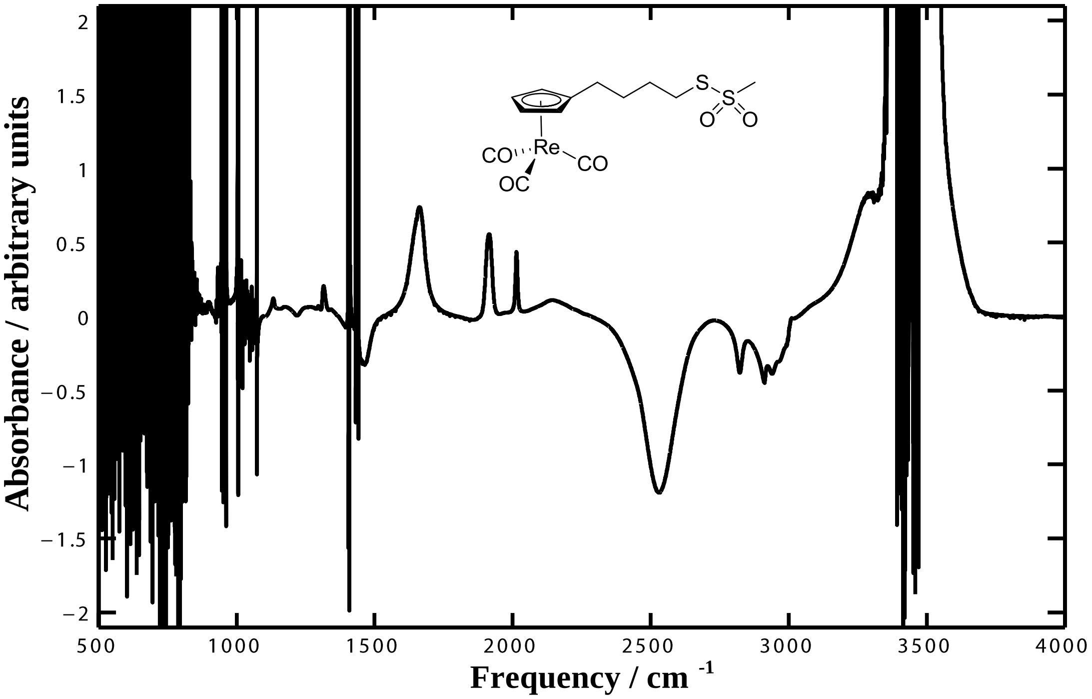


Figure A3.4. [η^5-(C_5H_4)$CH_2CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ FTIR Spectrum in DMSO

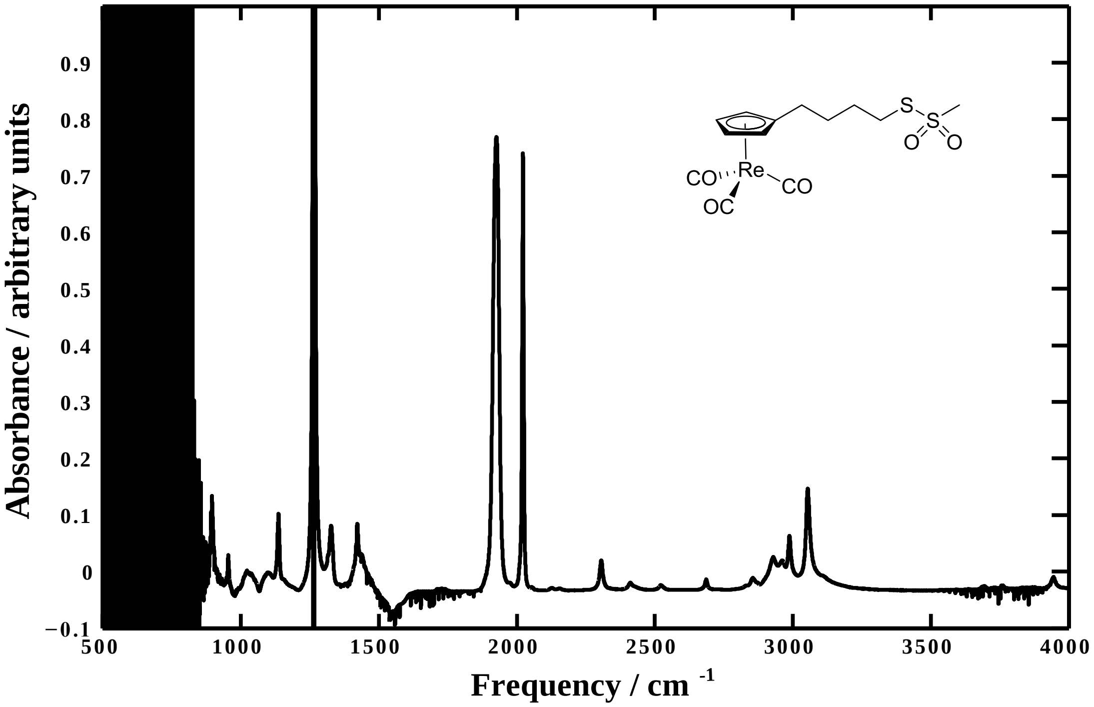


Figure A3.5. [η^5-(C_5H_4)$CH_2CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ FTIR Spectrum in 1:1 CCl_4:CH_2Cl_2

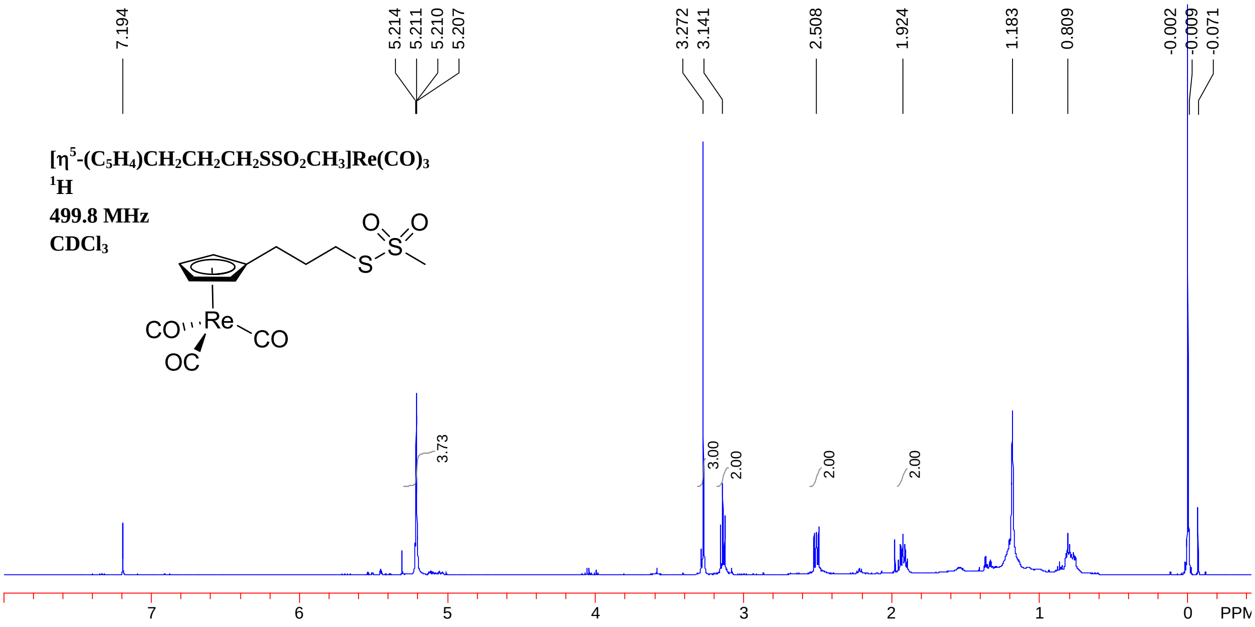


Figure A3.6. [η^5-(C_5H_4)$CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ ^{1}H NMR: (499.8 MHz, $CDCl_3$) δ 5.211 (Cp, m, 4H), 3.272 (-CH_3, s, 3H), 3.141 (-CH_2S-, t, J = 7.2 Hz, 2H), 2.508 (CpCH_2-, m, J(H,H) = 7.8 Hz, 2H), 1.775 ($CH_2C\mathit{H}_2$-CH_2-, p, J = 7.5 Hz, 2H).

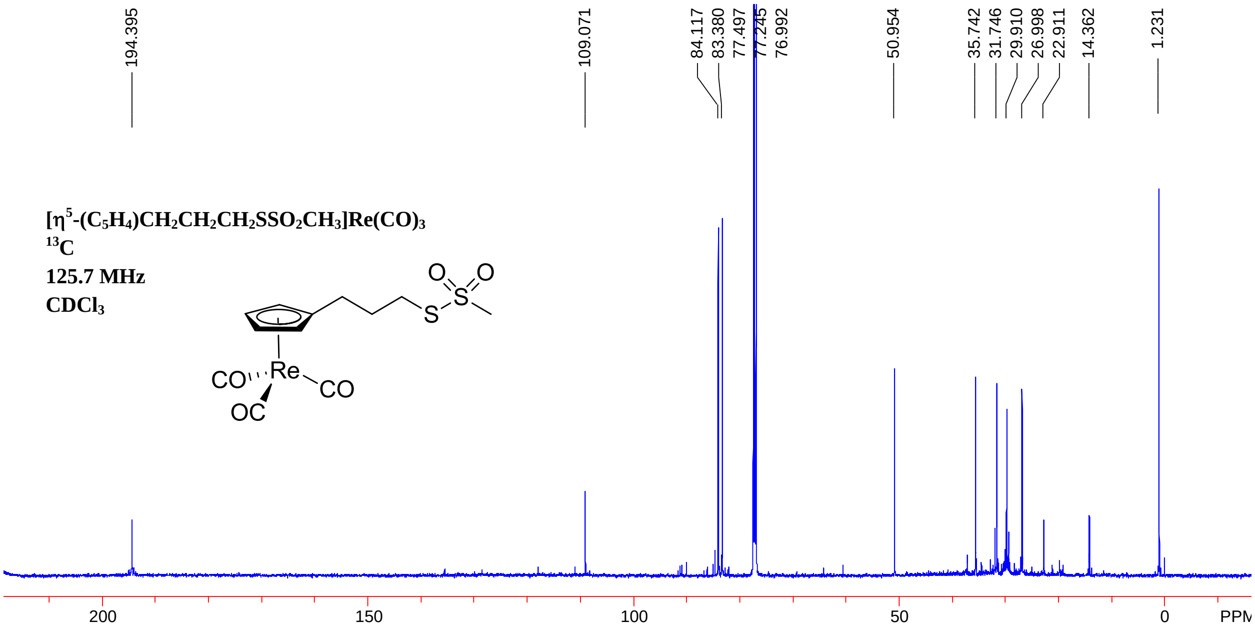


Figure A3.7. [η^5-(C_5H_4)$CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ ^{13}C NMR: (125.7 MHz, $CDCl_3$) δ 194.4 (*C*O), 109.1 (*ipso*-Cp), 84.1 (Cp), 83.4 (Cp), 51.0 (-*C*H_3), 35.7 (-*C*H_2S-), 31.7 (CpCH_2*C*H_2-), 26.998 (Cp*C*H_2-).

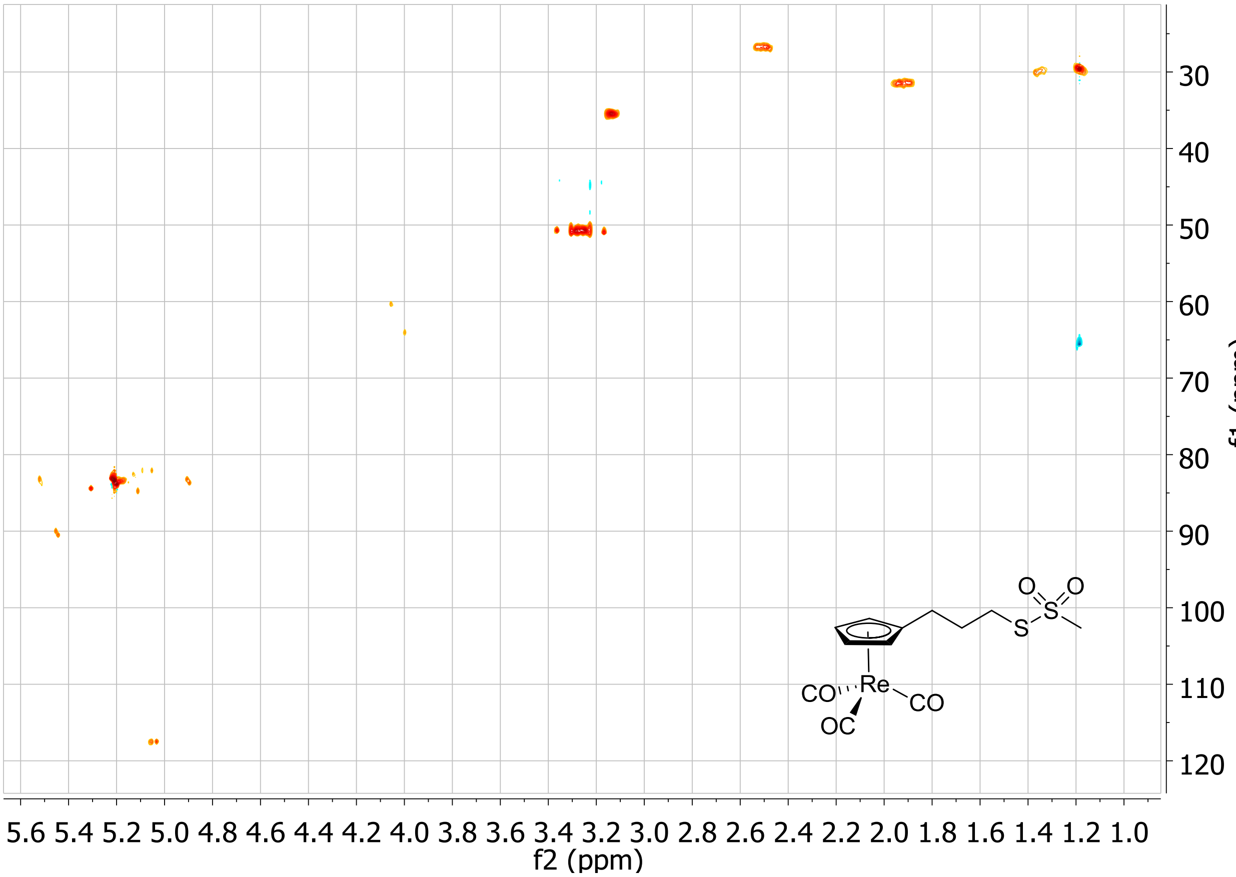


Figure A3.8. [η^5-(C_5H_4) $CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ HSQC Spectrum, 599.7 MHz, $CDCl_3$

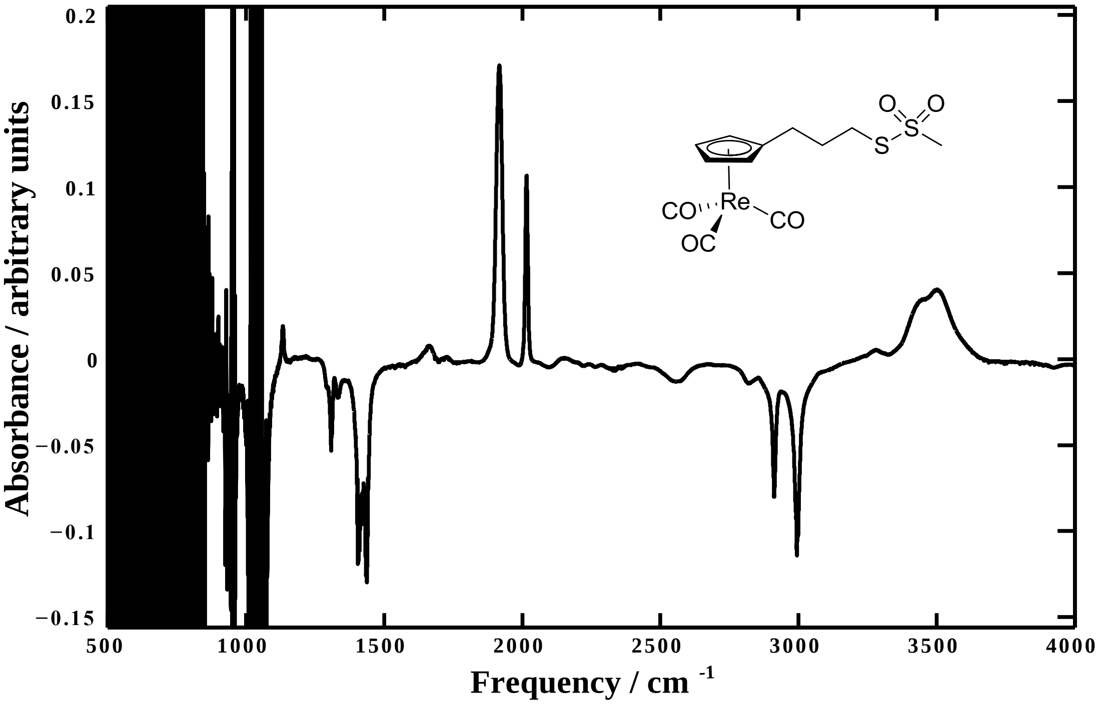


Figure A3.9. [η^5-(C_5H_4) $CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ FTIR Spectrum in DMSO

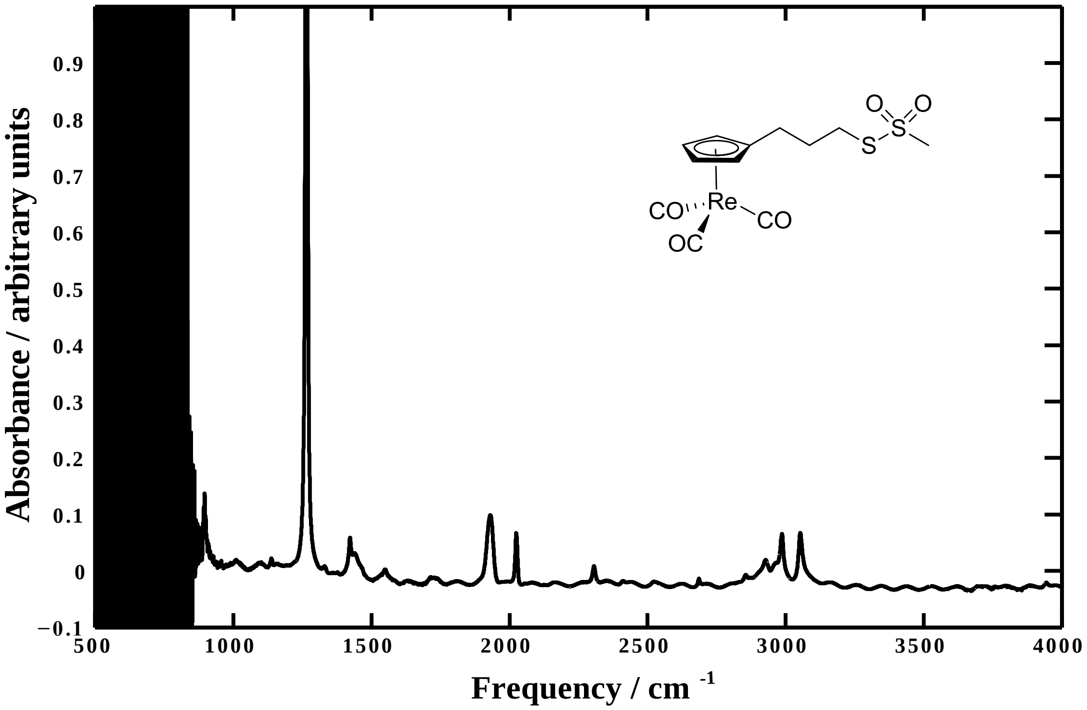


Figure A3.10. [η^5-(C_5H_4) $CH_2CH_2CH_2SSO_2CH_3$]Re(CO)$_3$ FTIR Spectrum in 1:1 CCl_4:CH_2Cl_2

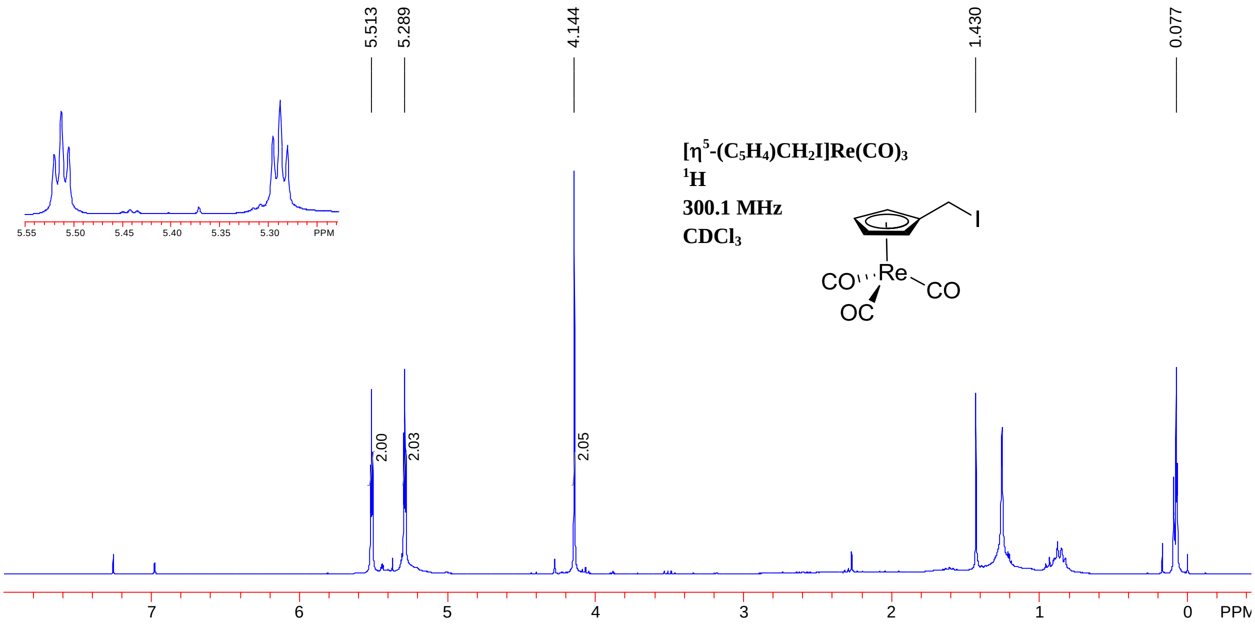


Figure A3.11. [η^5-(C$_5$H$_4$)CH$_2$I]Re(CO)$_3$ ^{1}H NMR: (300.1 MHz, CDCl$_3$) δ 5.513 (Cp, t, J = 2.2 Hz, 2H), 5.289 (Cp, t, J = 2.2 Hz, 2H), 4.1444 (-CH_2-, s, 2H).

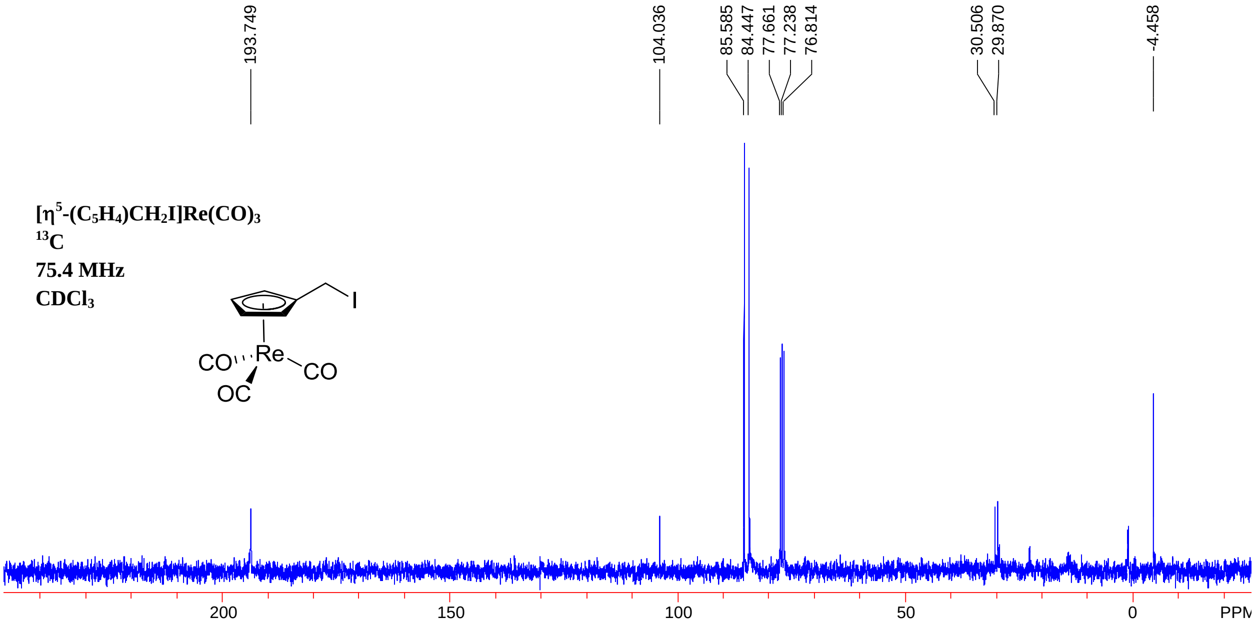


Figure A3.12. [η^5-(C_5H_4)CH_2I]Re(CO)$_3$ ^{13}C NMR: (75.4 MHz, $CDCl_3$) δ 193.7 (*C*O), 104.0 (*ipso*-Cp), 85.6 (Cp), 84.4 (Cp), 29.9 (-*C*H_2S-).

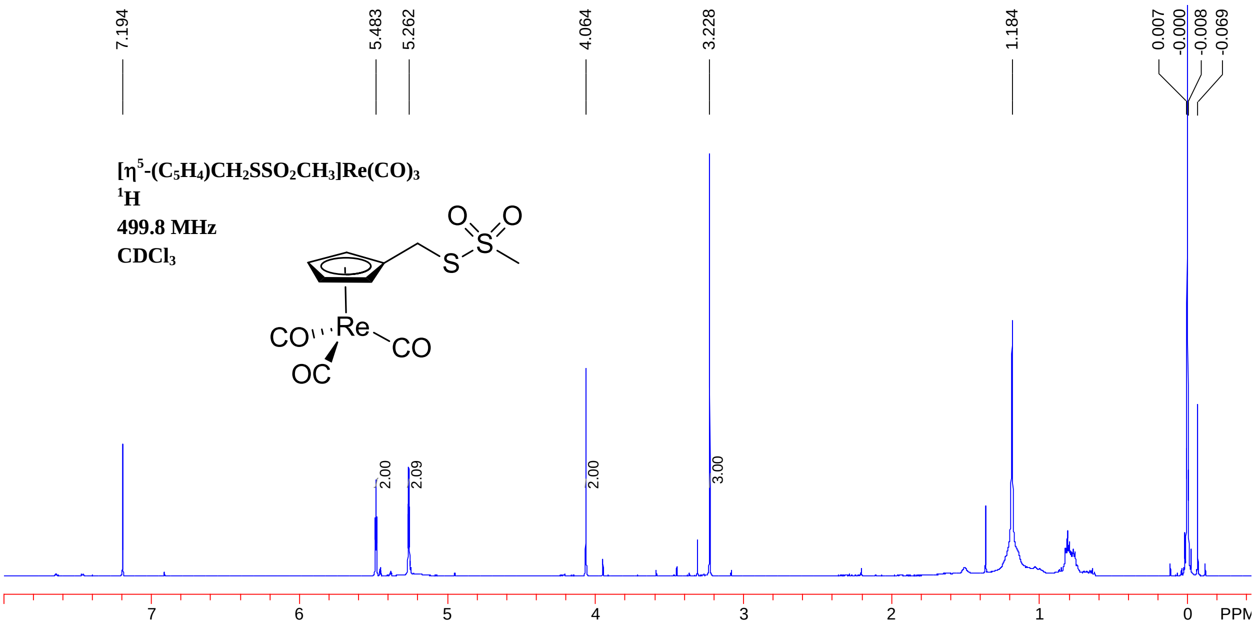


Figure A3.13. $[\eta^5$-$(C_5H_4)CH_2SSO_2CH_3]Re(CO)_3$ ^{1}H NMR: (499.8 MHz, $CDCl_3$) δ 5.483 (Cp, t, J(H,H) = 2.2 Hz, 2H), 5.262 (Cp, t, J(H,H) = 2.2 Hz, 2H), 4.064 (CpCH_2S-, s, 2H), 3.228 (-CH_3, s, 3H).

.

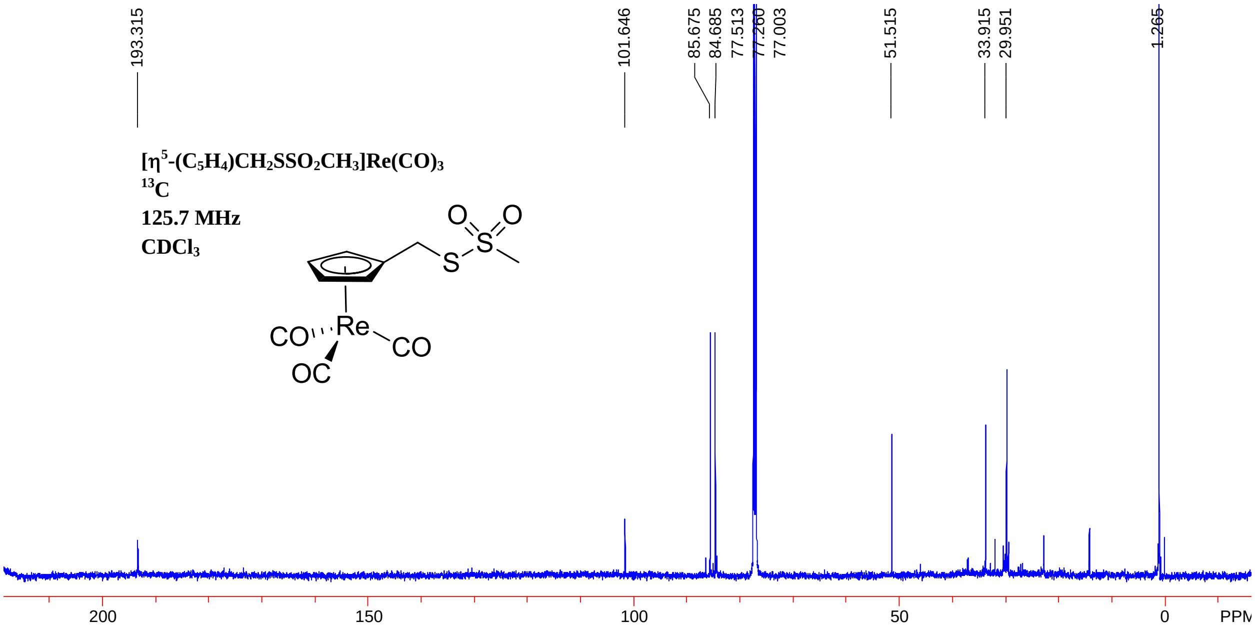


Figure A3.14. [η^5-(C$_5$H$_4$)CH$_2$SSO$_2$CH$_3$]Re(CO)$_3$ ^{13}C NMR: (125.7 MHz, CDCl$_3$) δ 193.3 (*C*O), 101.6 (*ipso*-Cp), 85.7 (Cp), 84.7 (Cp), 51.5 (-*C*H$_3$), 33.9 (-*C*H$_2$S-).

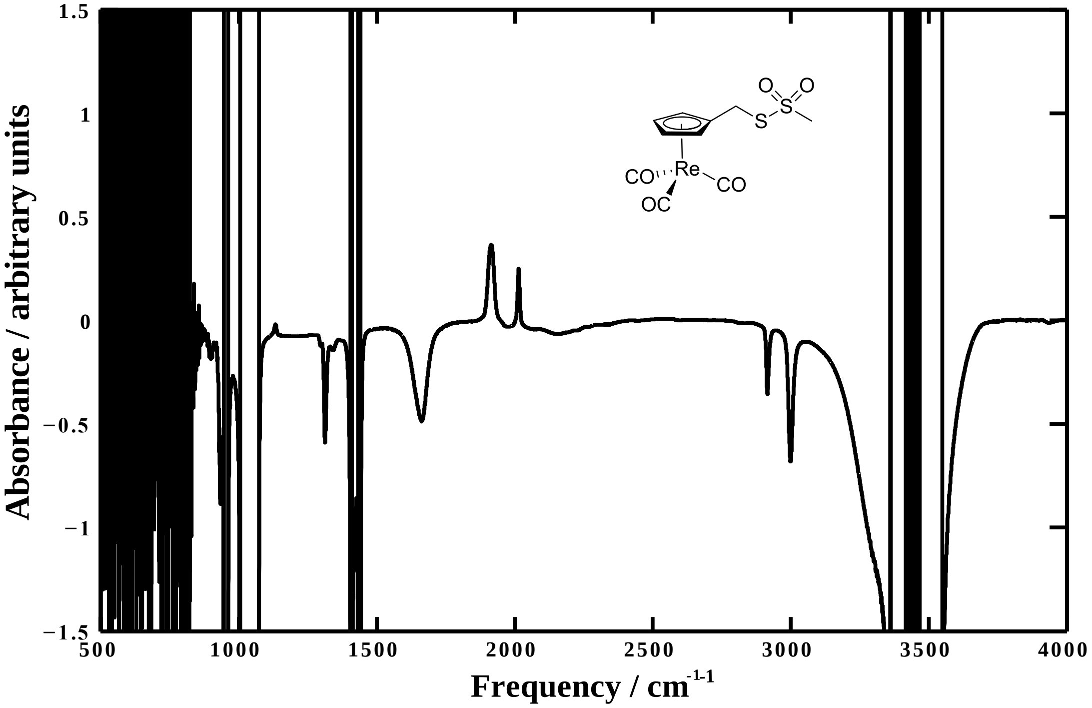


Figure A3.15. $[\eta^5$-$(C_5H_4)CH_2SSO_2CH_3]Re(CO)_3$ FTIR Spectrum in DMSO

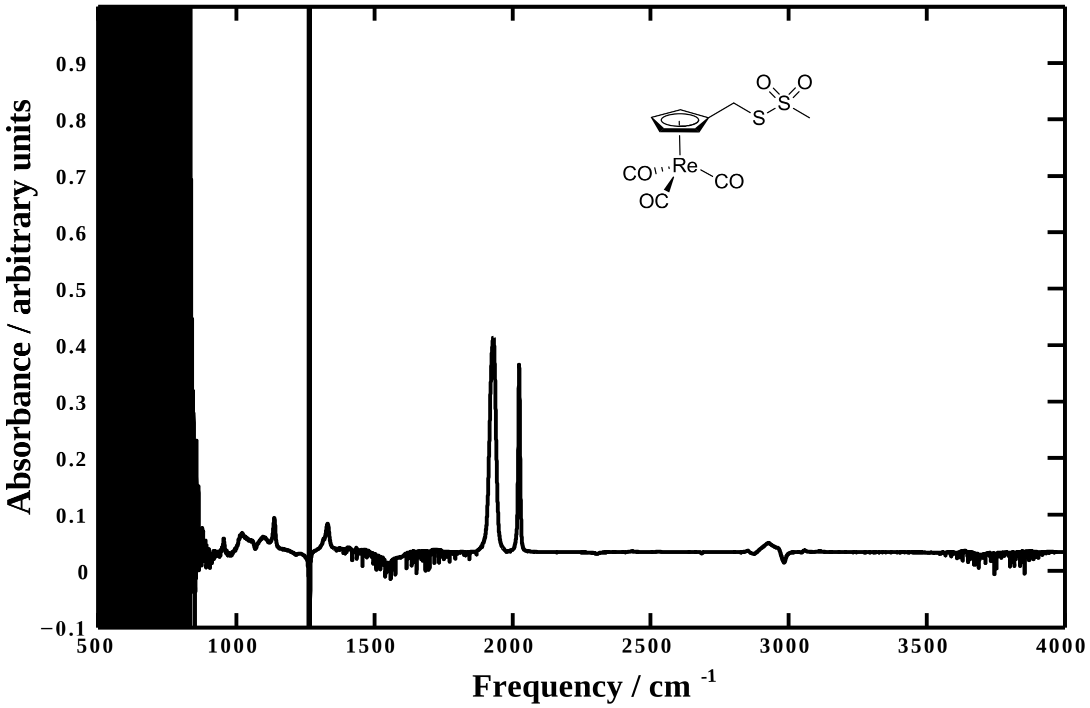


Figure A3.16. [η^5-(C_5H_4)$CH_2SSO_2CH_3$]Re(CO)$_3$ FTIR Spectrum in 1:1 CCl_4:CH_2Cl_2

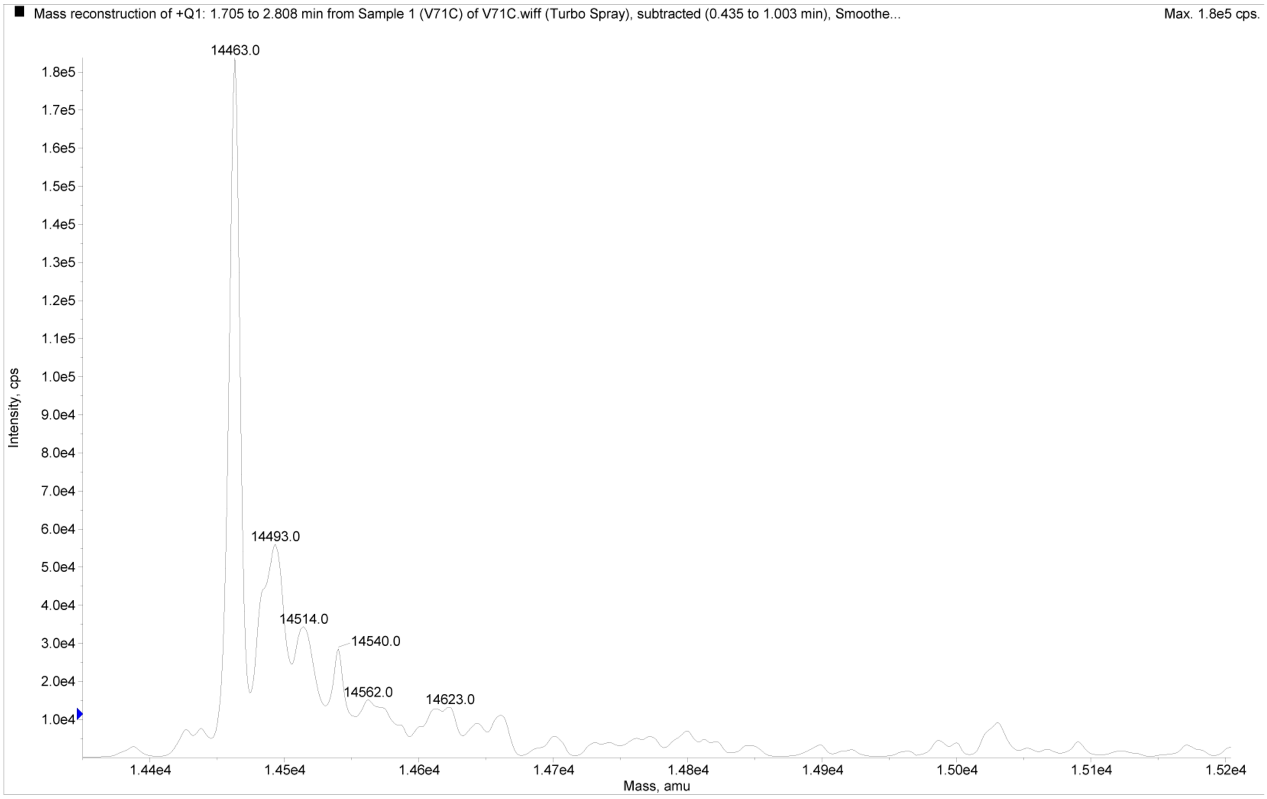


Figure A3.17. Unlabeled V71C α–synuclein MALDI (showing range including unlabeled (M+H+, calc. 14,465.1 Da) and labeled (M+H+, calc. 14,887.1 Da) mass

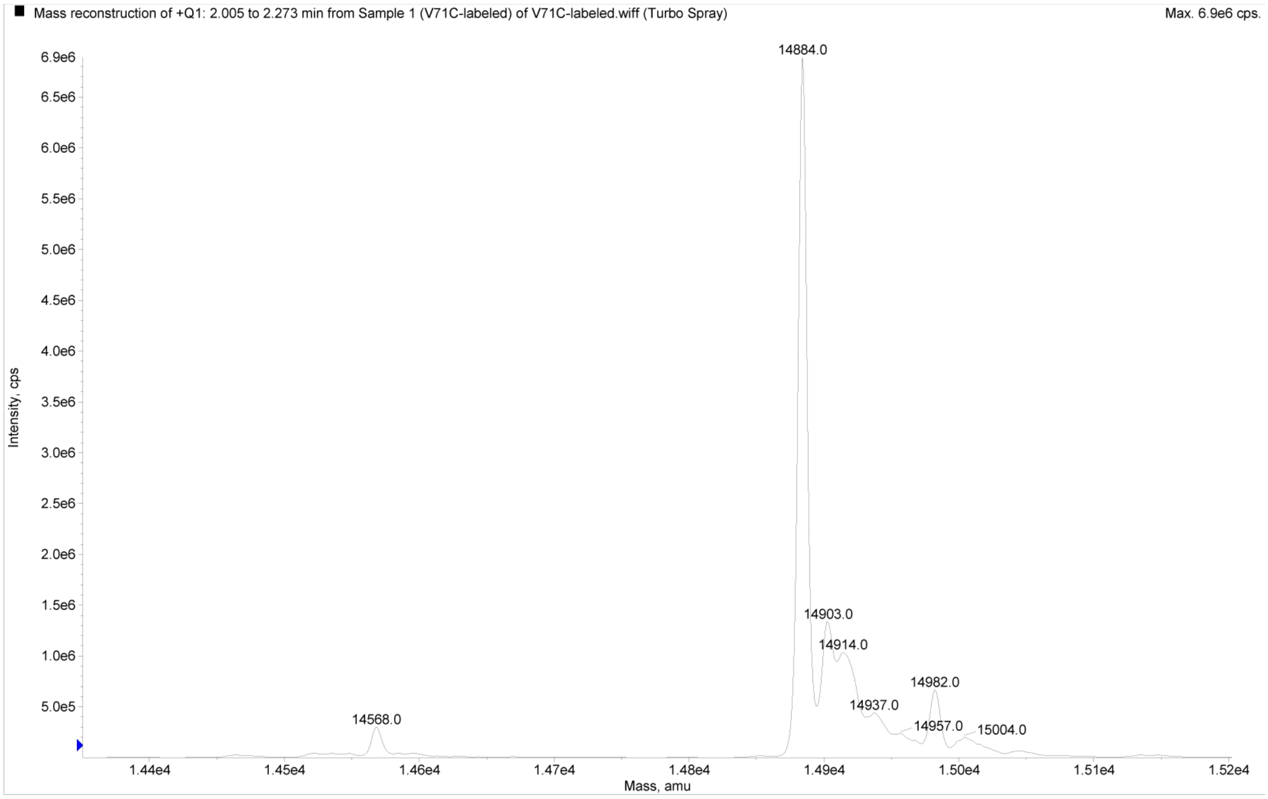


Figure A3.18. V71C α–synuclein labeled with [η^5-(C_5H_4)$CH_2CH_2CH_2CH_2SSO_2CH_3$]Re$(CO)_3$ MALDI spectrum (showing range including unlabeled (M+H$^+$, calc. 14,465.1 Da) and labeled (M+H$^+$, calc. 14,887.1 Da) mass

APPENDIX 4

Supporting Information for

CHAPTER 6

A Non-Natural Infrared Probe of Protein Electrostatics and Solvation[*]

A4.1 Methods

Labeled proteins were characterized by MALDI mass spectrometry (K6C-ReL1 calc. 9035.3, meas. 9042.3 [M+H$^+$]; K6C-ReL3 calc. 9063.3, meas. 9068.2 [M+H$^+$]; K6C-ReL3 calc. 9077.4, meas. 9077.2 [M+H$^+$]; K63C-ReL1 calc. 8919.2, meas. 8920.2 [M+H$^+$]; K63C-ReL3 calc. 8948.3, meas. 9488.9 [M+H$^+$]; K63C-ReL4 calc. 8961.3, meas. 8961.8 [M+H$^+$]).

The waiting time intensities are shown in Fig. A4.4 for the label alone in solvents and in Fig. A4.5 for the label attached to ubiquitin. These intensities were fit to a single exponential. Resulting relaxation times from the fits are given in Table A4.1. Since the lifetime for the fundamental (T_{01}) and for the overtone (T_{12}) are nearly identical for every sample, we reported T_{12} in the main text, since with measurements with the phase cycling scheme, are only susceptible to pump-pump scatter, which means that the T_{12} measurement is likely more accurate.[1]

---

[*] This chapter will be submitted to *J. Am. Chem. Soc.* in April 2012. It was prepared in collaboration with S. Mukherjee and M. T. Zanni.

Fig. 6.4 was drawn with UCSF Chimera. Starting with the ubiquitin structure (PDB ID: 1ubq), the indicated residues were mutated to cysteine, and for K6C, an aspartate residue was added to the end of the sequence. To the cysteine residue, the label was added and drawn from the X-ray coordinates of $CpRe(CO)_3$.[2] The hydrophobic surface was drawn based on the hydrophobicity of each residue as assessed by Kyte and Doolittle,[3] where red is the most hydrophobic and blue is the least. The circles in Fig. 4b-c were estimated by aligning sight through the protein along the α–C to β–C bond axis. The circles were centered over the β–C and have radii of the estimated length from the bond to the carbonyl ligands.

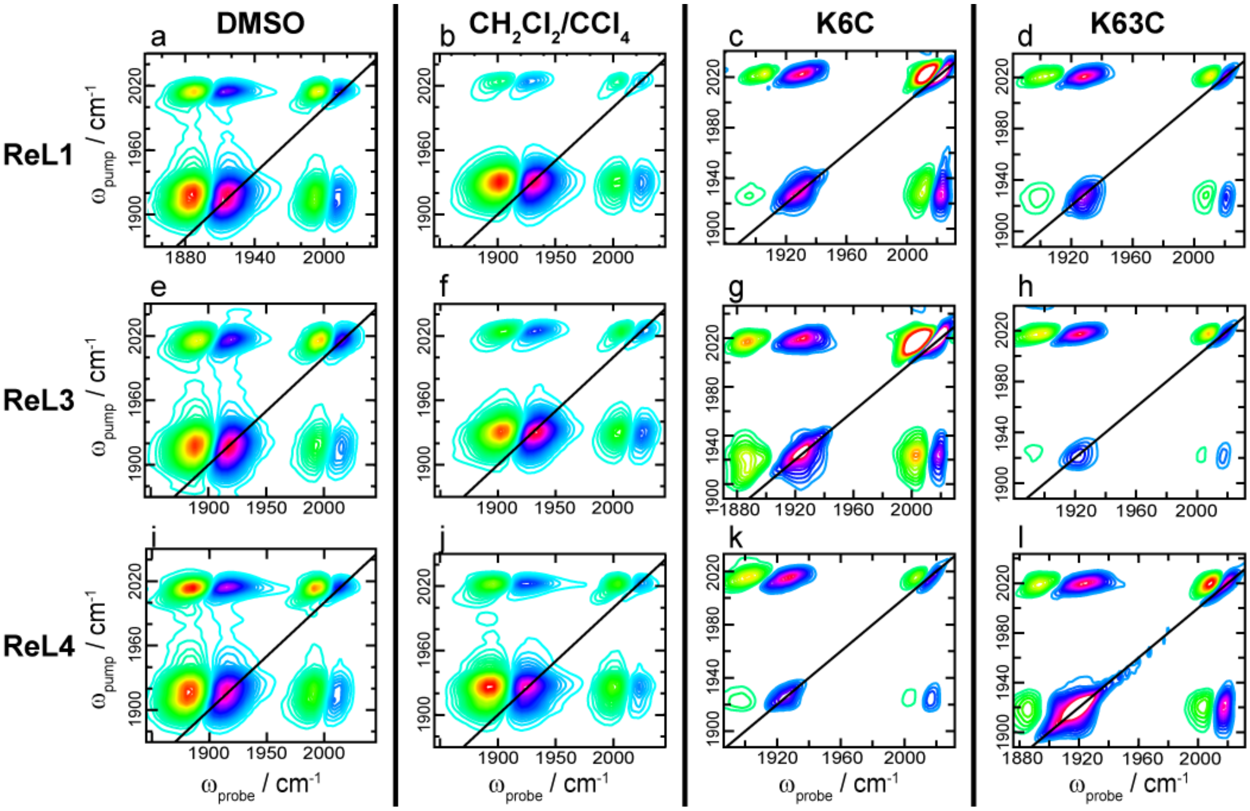


Figure A4.1 - 2D IR spectra for each label in DMSO, CH_2Cl_2/CCl_4, on K6C, and K63C.

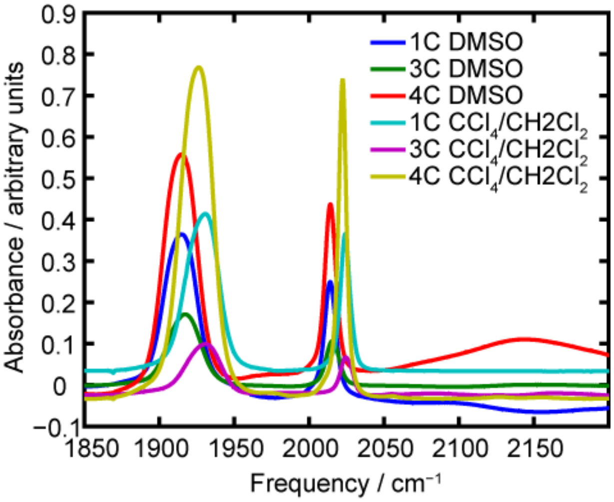


Figure A4.2 - FTIR spectra of the labels in DMSO or 1:1 CH_2Cl_2/CCl_4.

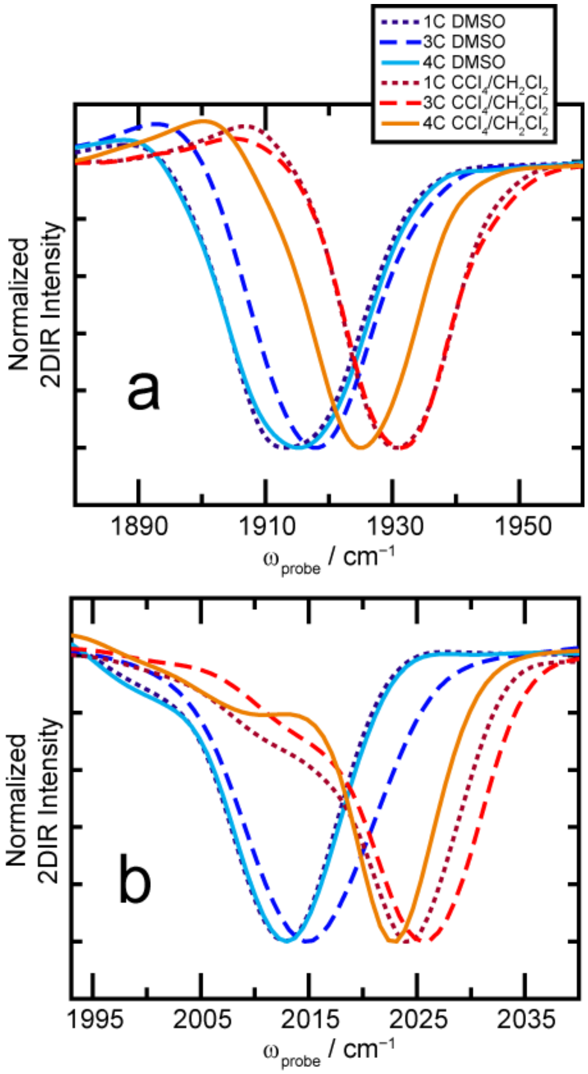


Figure A4.3 - 2D IR diagonal slices for the low (a) and high (b) frequency modes.

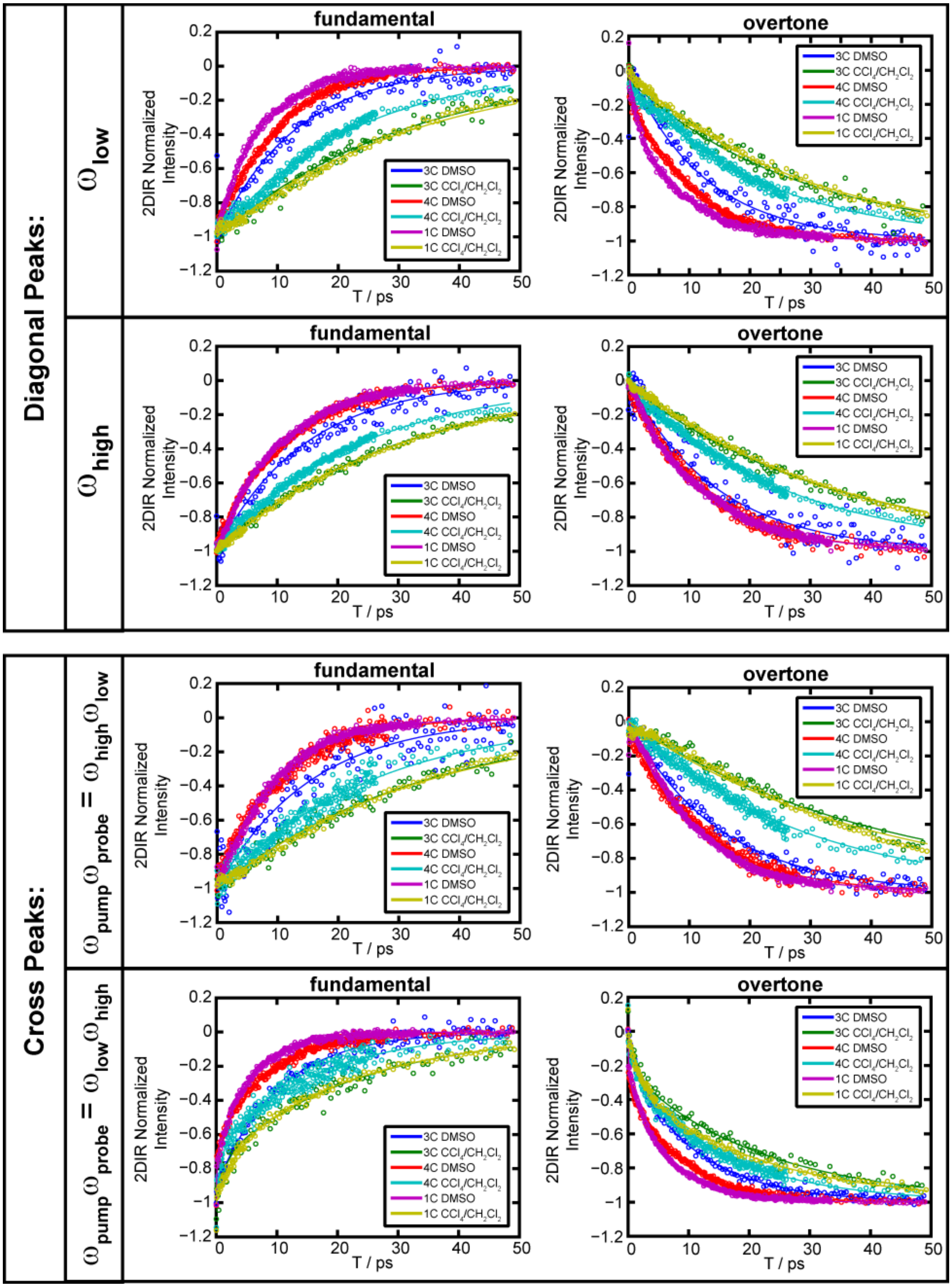


Figure A4.4 - Waiting time data for each label in DMSO or 1:1 CH_2Cl_2/CCl_4.

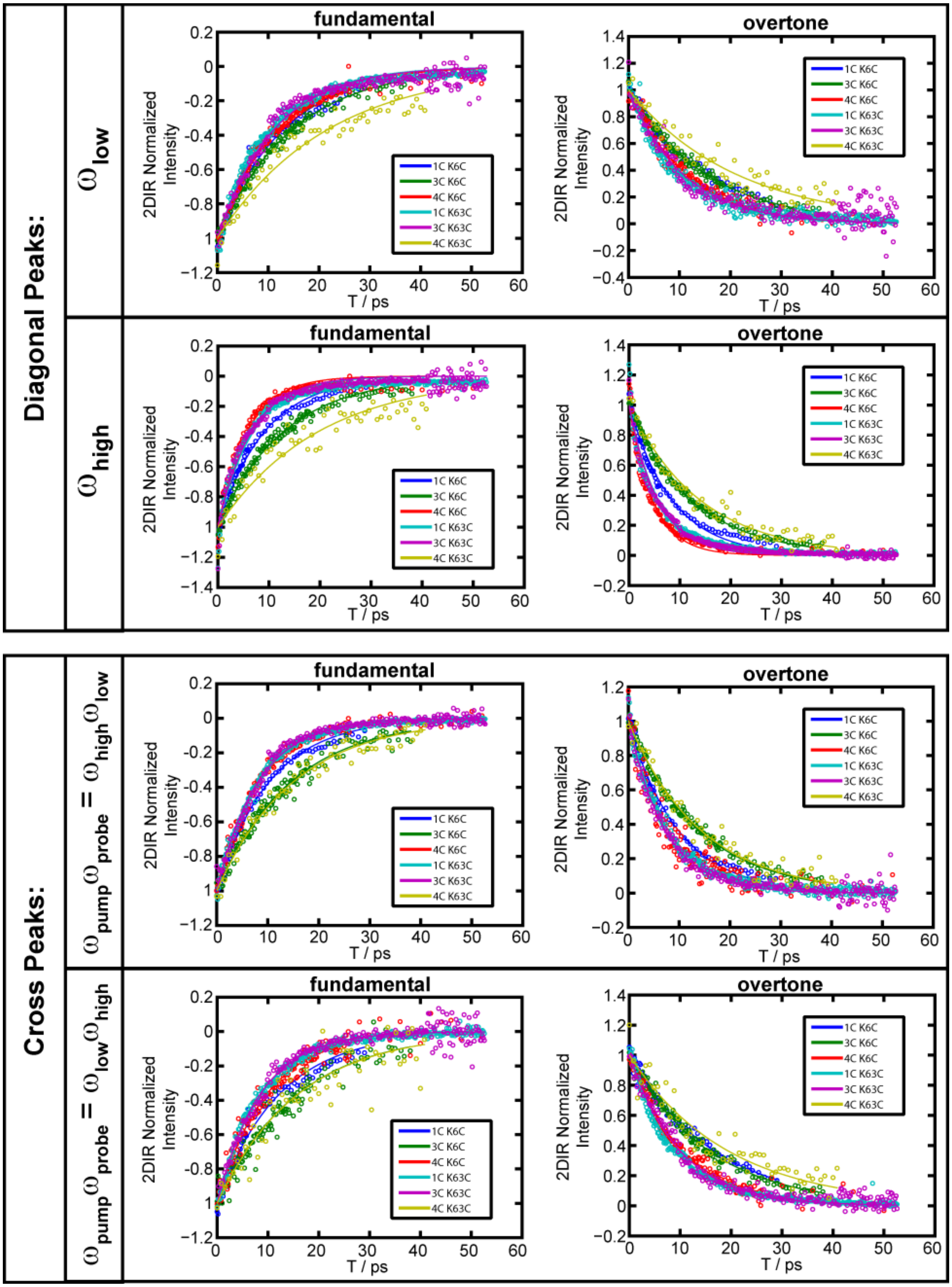


Figure A4.5 - Waiting time data for each label on K6C or K63C.

Table A4.1. Frequencies and lifetimes from 2D IR data.

		Diagonal Peaks						**Cross Peaks ($\omega_{pump}\omega_{probe}$)**			
		Low Frequency			**High Frequency**			**$\omega_{low}\omega_{high}$**		**$\omega_{high}\omega_{low}$**	
		ω / cm^{-1}	**T_{01} / ps**	**T_{12} / ps**	**ω / cm^{-1}**	**T_{01} / ps**	**T_{12} / ps**	**T_{01} / ps**	**T_{12} / ps**	**T_{01} / ps**	**T_{12} / ps**
DMSO	**1C**	1914.9	11	12	2014.0	5.6	6.1	9.6	11	7.6	7.8
	3C	1917.2	15	15	2015.6	10	10	15	15	13	12
	4C	1915.1	12	12	2014.2	7.9	8.4	11	12	10	10
CH_2Cl_2/CCl_4	**1C**	1930.7	31	34	2024.1	18	16	35	39	33	28
	3C	1931.0	30	34	2024.5	19	19	34	42	31	28
	4C	1926.4	24	26	2022.4	14	14	25	28	22	22
K6C	**1C**	1927.7	13	15	2022.2	9.1	8.5	12	17	10	9.7
	3C	1923.6	15	15	2019.4	13	12	16	15	15	14
	4C	1923.3	12	11	2015.4	5.8	4.7	9.7	10	8.3	7.8
K63C	**1C**	1927.6	11	10	2020.8	7.0	5.8	8.6	9.1	8.2	7.8
	3C	1923.2	11.1	10.8	2017.0	6.6	5.7	8.4	9.9	8.1	7.2
	4C	1920.4	21.7	21.2	2020.4	20	14	15.8	19.2	15	14.1

A4.2 References

(1) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* 2010, *52*, 12.

(2) Fitzpatrick, P. J.; Lepage, Y.; Butler, I. S. *Acta Crystallogr B* 1981, *37*, 1052.

(3) Kyte, J.; Doolittle, R. F. *Journal of Molecular Biology* 1982, *157*, 105.