Vibrational Probe and Methods Development for Studying the Ultrafast Dynamics of Preferential Solvation of Biomolecules by 2D-IR

by

Josef Adrian Dunbar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
in the University of Michigan
2015

Doctoral Committee:
Associate Professor Kevin J. Kubarych, Chair
Professor Eitan Geva
Associate Professor Nicolai Lehnert
Professor Ayyalusamy Ramamoorthy

To my family: my parents who instilled in me a sense of curious-engineering and mathematics, Papa whose love and wonder for the sciences he imparted on me, and Grace who has stood by me, with encouragement, throughout.

Acknowledgements

“If I have seen further it is by standing on the shoulders of giants.” – Issac Newton. The work described here would not have been possible without the help of those around me. My adviser, Kevin Kubarych, has provided me with key guidance throughout my degree, encouraging me to think creatively and critically of problems and providing me with the flexibility to pursue my interests. When I first joined the lab Kevin was instrumental in helping me catchup on the important aspects of our complicated experimental setup and data analysis. I would also like to acknowledge key former and current members in the group including: Ph.D. Derek Osborne whose guidance, over the three years we overlapped, is immeasurable, Ph.D. Matt Ross and Ph.D. Robert (Bobby) McCanne who helped me understand the hardware of the experimental setup, encouraging me to hook up an oscilloscope and monitor the hardware, and Aaron White my friend and colleague who has provided me with a listening ear and guidance when needed. The Biophysics program at Michigan has provided an outstanding environment to perform research and my class mates Matthew Stone, Sima Mofakham, and Edwin Najera, as well as the other student and faculty in the program have been amazing colleagues over the last five years.

My family has always proved a major source of support in my life. My parents, Lee, Harriet, and late mother Anna, provided me with invaluable encouragement throughout my education. My sister Katrina and Grandparents have provided me with counseling and continued support as I pursued my degree, even when it involved me moving to the other side of the country. Grace has been by my side since the start of my time at Michigan and has offered untold support and love over the course of my degree. I am truly thankful for the support I have received, and continue to receive, from my family and colleagues over the last five years.

Table of Contents

List of Figures

List of Appendices

Chapter One
Introduction

1.1 Proteins

At a fundamental level biology relies on the complex orchestration of proteins and their reactions within the cell. Comprised of linear polymers of the twenty naturally-occurring amino acids, proteins offer a highly modular scaffold enabling the catalysis of many of the cellular reactions. In many ways catalytic proteins, enzymes, represent idealized machines performing specially-tailored reactions, often at the speed of substrate and product diffusion. Harnessing the power of enzymes has led to improvements in commercial products where, by stabilizing enzymes for specialty applications, engineers have been able to take advantage of their catalytic power.[1-3] In addition to the utilization of enzymes for their natural reactions it has long been the romance of the chemist and biochemist to utilize the rich body of knowledge on proteins to design customized *de novo* enzymes.[4-7]

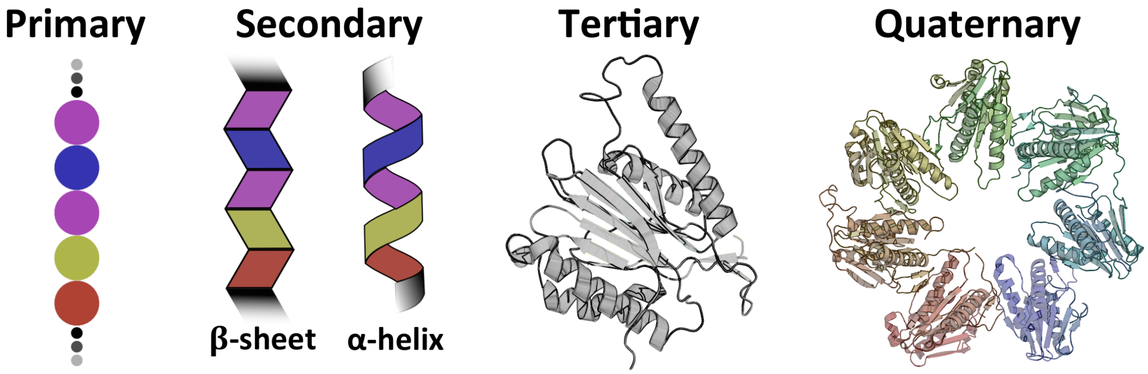

Figure 1.1 Classifications of protein structure from lowest to highest (left to right). The primary structure is composed of the sequence of amino acids in the polypeptide chain. Secondary structures are simple repeating motifs, such as β-sheet and α-helix, which fold to assemble the 3D, tertiary, structure. The quaternary structure defines the interactions between multiple protein molecules (structures from 1FNT.pdb).

Protein structure is often classified into four categories (highlighted in Figure 1.1): primary, secondary, tertiary, and quaternary. These respectively relate to the linear sequence of amino acids, the presence of common folding motifs such as α-helices or β-sheets, the three-dimensional structure of the protein atoms, and finally any interactions of the protein within a complex. Due to the hierarchical nature of these structural properties one is able to control and modulate the higher degrees of protein structure through mutation of the primary amino acid sequence.[8-9] Indeed the natural process of sequence mutation through evolution[10] is evident when comparing proteins of similar functions across organisms, homologs.[11-12] An example of homologous proteins are the avidin and streptavidin proteins from chickens and Streptomyces avidinii bacterium respectively.[13] These proteins (shown in Figure 1.2) both bind the vitamin H, biotin, with a very high affinity and are known to show nearly identical secondary, tertiary, and quaternary structures. A primary sequence comparison between avidin and streptavidin however shows only 30% sequence conservation between these proteins. The conservation of protein structure and function across homologous proteins is commonly seen despite considerable sequence mutations. This is due to the fact that the

Figure 1.2 Comparison of structures for two biotin-binding proteins Avidin (2AVI.pdb, yellow) from chicken and Streptavidin (1STP.pdb, blue) from the Streptomyces avidinii bacterium. These homologous proteins have a 30% sequence conservation and show a striking similarity in their secondary and tertiary structures.

evolutionary pressure to maintain protein function and cell viability is high.

Much effort has been spent studying protein structures and developing predictive models for determining the secondary and higher structure of proteins based on their primary structures.[14-19] While these models have proven useful in predicting some aspects of protein structure, considerable work remains to fully realize their potential. Protein engineering requires a further level of abstraction wherein instead of predicting a protein structure from a primary structure, one tries to determine the appropriate primary structure for a desired reaction site conformation. This formidable challenge is apparent for even small proteins with a 100 amino acid protein possessing $20^{100} \approx 1.3 \times 10^{130}$ primary sequence mutations. Taking a cue from biology this intimidating conformational space can be made more manageable through the utilization of protein scaffolds onto which the catalytic site is engineered. This rational design approach has proven useful in the realization of multiple *de novo* enzymes[20-22], however, without more detail on the effects of protein mutation on the scaffold enzyme, design remains challenging.

1.2 Protein Dynamics

Proteins structures are often determined through x-ray crystallography, where proteins are made to form crystals under narrow solvent conditions[23]. Due to differences between the crystal buffer solution and the cell's environment, and the absence of crystal packing within the cell, it is justified to expect proteins to exhibit a degree of structural heterogeneity *in vivo*.[24-28] Interconversion between these states gives rise to protein dynamics for which multiple biophysical experiments are able to showcase the structural heterogeneity of proteins.[29-30] Protein dynamics span many timescales[31]: from slow global conformational changes on the second timescale to the ultrafast shuttling of energy from photon absorption on the femtosecond, a millionth of a billionth of a second, timescale. These ultrafast dynamics, which further include localized vibrations of the protein structure and the motions of the solvent around the amino acids, serve as the foundation upon which many of the slower dynamics of proteins manifest. Thus to fully understand the

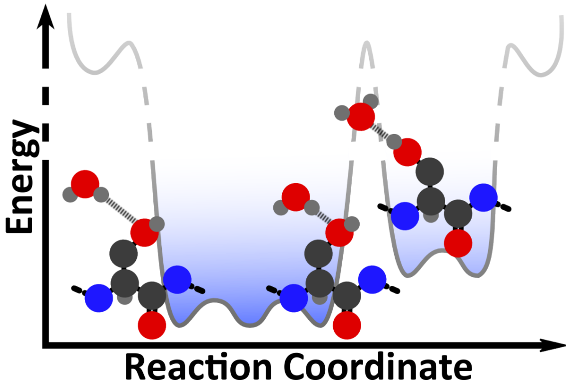

Figure 1.3 Hydrogen bond fluctuations within the local minima structure give rise to ultrafast dynamics of the protein structure. Slower dynamics, involving the rotation of amino acid side chains creates additional heterogeneity of the structure. These additional minima experience separate ultrafast dynamics, specific to their environment.

dynamical nature of proteins one must understand the manifestation and propagation of their ultrafast dynamics.

A classic example of structural heterogeneity that fluctuates on the ultrafast timescale is the hydrogen bond present in liquid water.[32-36] Due to the transient nature of the hydrogen bond and the multiple donors and acceptors present in solution the heterogeneity of hydrogen bonds in solution arises naturally. Similar to hydrogen bonds in water, the secondary and tertiary structures of proteins are often stabilized by hydrogen bonds to the peptide backbone and amino acid side chains.[37-39] These hydrogen bonds, schematically shown in Figure 1.3 fluctuate about their local minima, which in addition to slight conformationally distinct arrangements of the amino acid side chains, gives rise to a heterogeneous ensemble of protein structures.

1.3 Vibrational Spectroscopy

Vibrational spectroscopy offers a unique way of investigating the bond energies of a system. For a typical experiment a molecule is made to interact with infrared (IR) light of a known frequency composition, Figure 1.4a. Similar to other spectroscopic methods, transitions of the molecule that correspond to the energy of incident light allow the molecule to absorb a photon. Photon absorption is accompanied by the transition of the molecule into an excited state that can relax back to the ground state through interactions with its environment. By investigating the frequency composition of the IR light after interaction with the sample one is able to determine the transitional energies in the sample. These energies are well characterized in the IR and are known to correspond to specific bonds between atoms.

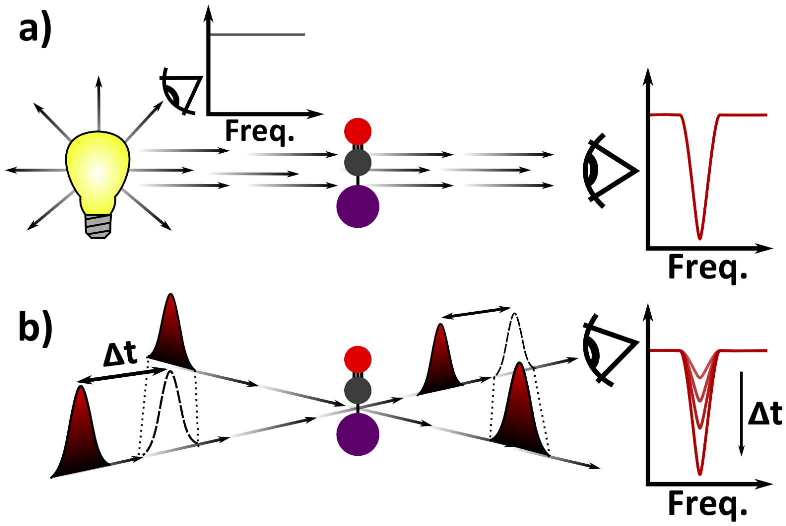

Figure 1.4 a) Traditional vibrational spectroscopy utilizes a continuous IR source to investigate the sample. By comparing the frequency composition of the light before and after interaction with the sample, information is gained on the energetic transitions of the sample which can be used to assign its chemical bonds. b) Pump-Probe vibrational spectroscopy allows for monitoring the relaxation of vibrational energy in the sample with a high degree of temporal resolution. By scanning the time delay between the two pulses, *Δt*, the recovery of the ground state signal can be linked to dynamical motions of the system.

Additional information can be obtained from vibrational spectroscopy if one is able to obtain temporal resolution of the spectra. By simply time-resolving IR spectra, one is able to monitor the progression of slow reactions by monitoring key frequencies. With a significantly high temporal resolution one can monitor the formation and breaking of molecular bonds in a heterogeneous environment, gaining valuable information regarding the timescales of dynamics.[40-43] Temporal resolution can be obtained in multiple ways. For low temporal resolution requirements one can rely on the temporal binning of the IR spectra, processing each spectrum such that it has a known time value. This method scales poorly however as one tries to go towards higher temporal resolution. This is due to many issues including: the detectors ability to monitor and report the desired frequency of light, as well as needing a light source of sufficient power so that enough photons are interacted with the sample and subsequently the detector such that a spectra can be obtained.

To address the issues that arise from trying to increase the temporal resolution of an experiment, pulsed lasers are frequently used in a relatively simple pump-probe setup, Figure 1.4b. These laser systems are commonly capable of probing dynamics on the tens of femtoseconds timescale.[40, 44-45] This is achieved through the interaction of the sample with two light fields separated by a time delay, Δt. Time delays can be made very short through the manipulation of the physical path the light takes to reach the sample. In this two-pulse experiment the first light pulse interacts with the sample exciting the transitions that fall within its frequency spectrum. After a variable time delay the second weaker pulse then interacts with the sample and is spectrally resolved to measure its frequency composition. If the time delay is sufficiently short the molecules will have not been able to relax back to their ground energy state and the resulting spectra will show a depletion of the ground state signal (as well as the presence of an excited state signal). By stepping the delay time between the two pulses, commonly referred to as excitation and detection pulses, one can follow the recovery of the ground state signal (or loss of the excited state signal) as a function of delay time. Applied to proteins pump-probe vibrational spectroscopy has been able to monitor the loss and rebinding of carbon monoxide to Cytochrome c Oxidase on the ultrafast timescale[46-47] as well as gain insight into the energy transport within proteins.[40, 47]

Further spectroscopic information can be obtained through the utilization of two-dimensional infrared spectroscopy (2D-IR).[48-50] 2D-IR incorporates an additional dimension to the Pump-Probe spectroscopy allowing for temporally resolving the dependency of the detection frequency on the excitation frequency. By mapping this dependency of the detection frequency we are able to gain insight into the heterogeneous nature of the microstates of the system. Further information is gained by allowing for interconversion of the system's microstates on the ultrafast time scale. Following how vibrational energy redistributes between the vibrational modes of the system allows for an understanding of the dynamics arising from microstate conversion.

Similar to Pump-Probe spectroscopy, 2D-IR is implemented using ultrafast laser pulses in order to gain the needed temporal resolution, Figure 1.5. In total three laser pulses must interact with the sample in a 2D-IR experiment. The first two pulses of the experiment, k_1 and k_2, represent an excitation pair whose time separation, τ, is scanned for fixed values of T_W, the time delay between the second and third pulse, k_3. These time delays are often referred to as the coherence time and waiting time respectively. After interaction with the three pulses the sample is left in an excited state which radiates a fourth field, k_s, at a time delay T_3. In practice T_3 is detected in the Fourier domain, via a spectrometer, and is recorded as a vibrational echo spectrum. By scanning the coherence time delay for a fixed waiting time and recording the vibrational echo spectra, one is able to apply a Fourier transform over the coherence time dimension to obtain the excitation frequency, $\omega_{excitation}$. The obtained 2D-IR spectrum resolves the dependence of the spectrally resolved echo signal on the excitation frequency.[51-53] This mapping of the detection frequency, $\omega_{detection}$, and excitation frequency is performed for multiple waiting

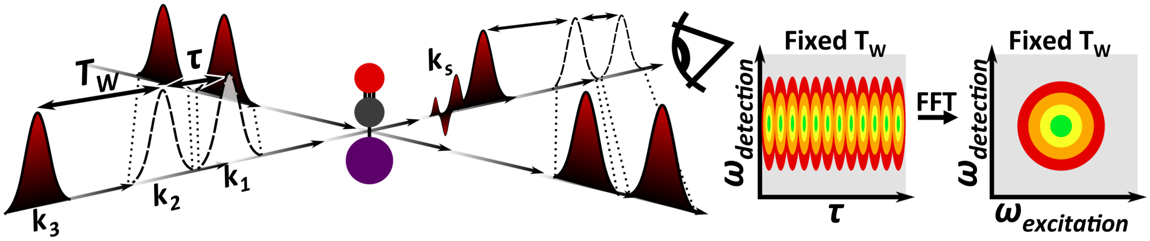

Figure 1.5 2D-IR pulse sequence. Three laser pulses, k_1, k_2, and k_3, interact with the sample to generate the signal field, k_s. The signal is collected in the frequency domain, utilizing a spectrometer (not shown), as a function of the coherence time, τ, which is scanned for each waiting time, T_W. The detection frequency, $\omega_{detection}$, vs. coherence time data is then Fourier transformed over the coherence time dimension to obtain the excitation frequency axis, $\omega_{excitation}$. For each waiting time a 2D spectrum correlating the excitation and detection frequencies is obtained.

times, allowing one to monitor how the vibrational energy redistributes in the system.

1.4 Spectral Diffusion

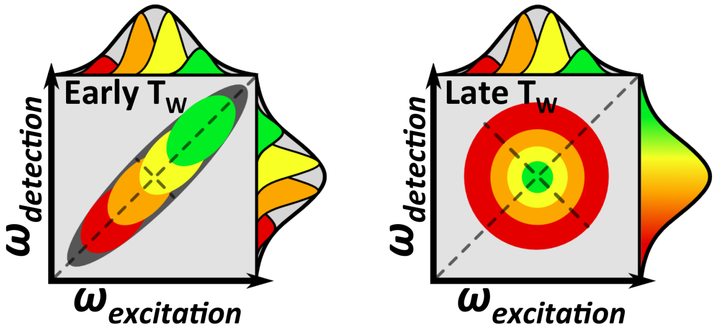

Figure 1.6 Sample 2D-IR spectra showing an inhomogeneously broadened line shape at early waiting time (left) and a homogeneously broadened line shape at a late waiting time (right). By monitoring the ellipticity of the observed peak, the correlation between the excitation and detection frequencies can be quantified as a function of waiting time. This correlation measurement allows for the quantification of the spectral diffusion time scale.

As previously described vibrational spectroscopy allows for the monitoring of the vibrational energies of a given molecule. Due to interactions with their environment each molecule will exhibit a shifted vibrational frequency from the mean ensemble value, leading to a broadening in the sample's vibrational spectra. This broadening, which arises from the different environments of the molecules in the sample, is termed inhomogeneous broadening.[54] As the different environments interchange with one another the frequency of a given oscillator will appear to undergo Brownian diffusion. This diffusion of the frequency, termed spectral diffusion, provides valuable information on the timescales for the molecules of a sample to interchange between different microstates. Applied to biological systems, spectral diffusion has been used to monitor the slowdown of hydrating water at the interface of proteins[55-57] and extended lipid environments.[58]

The traditional 2D-IR method provides an insightful picture of how one can measure the spectral diffusion, however alternative methods exist for extracting this information.[59-61] 2D-IR allows for measuring the detection frequency's dependency on the excitation frequency at a specific waiting time. Examining a 2D-IR spectrum acquired at an early waiting time, Figure 1.6 left, an elongation of the spectrum is observed along the $\omega_{excitation} = \omega_{detection}$ diagonal. Due to the short waiting time, the probed molecules have not had sufficient time to sample alternate microstates in the system, leaving the excitation and detection frequencies highly correlated. The observed elongation of the peak, along the diagonal, represents the inhomogeneous broadening of the vibrational spectrum. As the waiting time is extended the sample will lose its inhomogeneous character and the line shape will become less elongated, appearing rounded in the spectrum. In the late waiting time limit, Figure 1.6 right, the sample has had ample time to explore all possible microstates of the system. This leads to a total loss of the inhomogeneous broadening observed at early waiting time accompanied by the loss of correlation between the excitation and detection frequencies. This correlation loss is quantifiable by fitting the peak shape and determining the ellipticity of the 2D-IR spectrum at each waiting time point.[62]

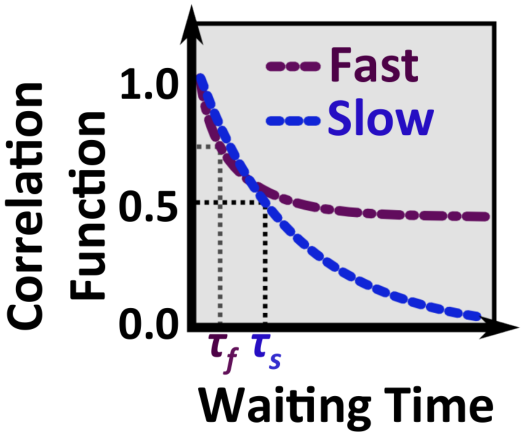

Figure 1.7 Sample Correlation Functions plotted vs. the Waiting Time (T_W) for a system undergoing fast and slow spectral diffusion (purple and blue curves respectively). By fitting the data to an analytical function, such as an exponential decay, one is able to quantify: the degree of system inhomogeneity from the y-intercept, the spectral diffusion lifetime, and the degree of unsampled microstates from the correlation function value at long waiting time.

Plotting this correlation function as a function of the waiting time, see Figure 1.7, provides insight into the spectral diffusion dynamics of the system. In practice the correlation function is fit to an analytical function, commonly an exponential decay, to aid in the interpretation of the data. The y-intercept is a direct measure of the inhomogeneity of the system and allows one to extract the homogeneous and inhomogeneous line width from a linear absorption spectrum. The decay rate provides a measure of the timescale of the systems spectral diffusion, which is relatable to the dynamics of the systems interconversion between microstates. Finally any offset of the correlation function observed at a long waiting time highlights unsampled dynamics that occur on timescales longer than the vibrational lifetime of the probed transition.[62] Together this information provides valuable insight into the nature of the molecules in the sample.

1.5 Site Specific Vibrational Labels

For biophysical systems it is often ideal to possess a high degree of temporal and spatial resolution. Ultrafast laser pulses are able to provide the needed temporal resolution. To achieve a high degree of spatial resolution, however, molecular labels are frequently employed to gain detail at the molecular level.[63-66] The most widely recognized molecular probes in biophysics are arguably the green fluorescent proteins (GFPs).[67-69] These molecule probes can be incorporated into the DNA of an organism, such that when certain proteins are expressed in a cell, the probe is covalently linked to the protein of interest. GFP and its derivatives[70-71] are powerful spectroscopic probes for visible spectroscopy. Multiple other probes exist for alternative wavelengths, providing varying levels of spatial resolution depending on the technique used and the probe. A great advantage of utilizing spectroscopic probes is the ability to study the same probe in a variety of systems. This allows for the characterization of the probe in different environments, allowing one to calibrate the probe’s response to environmental parameters. A common example of this is the frequency shift of a spectroscopic probe due to its environment’s polarity.[72-73] Utilized in this way, probes allow for the monitoring of: site-specific protein folding and unfolding[74-76], ligand binding[77], charge state of neighboring atoms[58], and many other properties.[78]

For vibrational spectroscopy two routes are commonly taken to incorporate spectroscopic probes in biological samples. The first method is the least invasive and involves the isotopically pure synthesis of a system, such as a protein.[79] This method relies on the incorporation of an isotopic label at a specific site of the molecule such that the lighter and heavier isotopes possess a different vibrational frequency and can be isolated in a spectrum. A second approach, more commonly taken in our group, is the introduction of unique spectroscopic probes that exhibit vibrational frequencies outside of the naturally occurring modes found in biology.[57-58, 80] This second method offers a higher degree of sensitivity than the first due to the ability to select vibrational modes with a strong spectroscopic signature in a background free region of the IR spectrum. Additionally the later method allows for the study of larger systems that would prove cost prohibitive to synthesize in most labs.

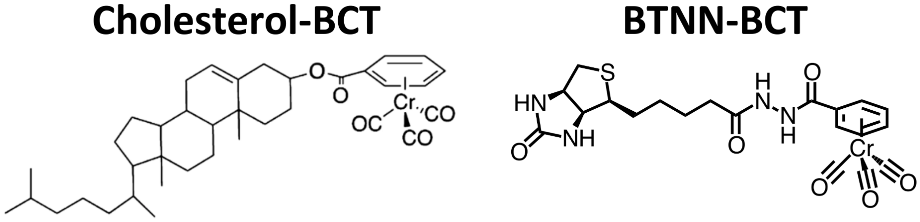

Figure 1.8 Derivatives of benzoyl chromium tricarbonyl (BCT) used as site-specific molecular labels. Cholesterol-BCT has shown to incorporate into lipid bicelles offering insight into the dynamics of water at lipid interfaces. Biotin hydrazide BCT (BTNN-BCT) can be incorporated into the streptavidin family of proteins and offers site-specific information on the protein dynamics.

The vibrational probes commonly used by our group are based on metal carbonyls. These probes offer a strong signal in IR spectroscopy and have been repeatedly shown to serve as sensitive dynamical probes of their environment.[32, 52, 55-58, 81-89] The rich chemistry offered by the organic backbone of some metal carbonyls allows these probes to be specifically synthesized to incorporate with different biological systems. Significant progress has been made to realize the potential of, benzoic chromium tricarbonyl (BCT) as a probe for biological systems, highlighted in Figure 1.8. Utilizing the benzoic acid chemical group as a handle we have linked the probe to site-specific molecules in multiple systems. The first system our group realized utilizing the BCT probe was in the derivatization of BCT with cholesterol, Figure 1.8 left, to create a lipid specific probe of solvation dynamics.[82] This probe demonstrated the dynamical slowdown of water at extended lipid interfaces utilizing low millimolar concentration. Additionally we have shown the derivatization of BCT with a biotin analogue, Figure 1.8 right, to allow for monitoring the dynamical response of site-specific mutations in a protein scaffold (described more in Chapter 4). This labeling chemistry hinges on applying Steglich Esterification to BCT and offers great potential in studying these and future biologically relevant systems.

1.6 Method Development Towards Faster Acquisition

The signal isolation and strength metal carbonyls offer vibrational spectroscopy, compared to isotopic labels, allows them to be utilized at relatively low sample concentrations. This enables the study of precious biological samples by 2D-IR. In addition to the amount of sample needed however, sample stability must also be considered for many biological systems. This pays an effect when trying to study proteins that may denature or come out of solution on the hour time scale required for the acquisition of multiple 2D-IR spectra. Additionally one may wish to average over multiple data sets to improve the signal to noise ratio of the data which, due to its square root scaling with the acquisition number, can prove foreboding when large improvements are desired. To address these common issues in 2D-IR, we have worked on introducing and implementing two complimentary data acquisition methods in addition to our traditional setup. Rapidly Acquired Spectral Diffusion (RASD)[60] serves as the fastest method for acquiring spectral diffusion data, eliminating the need to scan the coherence time delay. Compresses Sensing 2D-IR (CS-2D-IR)[59] allows for sidestepping the Nyquist sampling requirement, offering greatly reduced acquisition times while still providing complete 2D-IR spectra. These two methods, described below, offer significant reduction in the data acquisition time allowing for more spectra to be obtained within a time period and for the study of more transiently stable systems.

1.6.1 Rapidly Acquired Spectral Diffusion

The method of rapidly acquired spectral diffusion (RASD) eliminates the need to scan the coherence time delay in the traditional 2D-IR method. This reduction in the dimensionality of the acquisition provides a significant reduction in the acquisition time of spectral diffusion measurements. Furthermore by eliminating the need to scan the coherence time for each waiting time, we are able to continuously scan the waiting time delay as opposed to stepping the motor. Scanning the delay provides a finer resolution of delay than is available by stepping the motor, allowing for an over ten-fold oversampling of the waiting time while reducing the required acquisition time over sixty-fold.

Compared to the traditional method of 2D-IR, RASD reports on the spectral diffusion of a single excitation and detection frequency as opposed to monitoring the entire spectral band. The excitation frequency is determined from first collecting a traditional 2D-IR spectrum and integrating over the desired detection frequency window. The signal intensity as a function of the coherence time delay is then calculated to determine the optimal coherence time delay and the corresponding motor positions. Once optimized, the coherence time is set and the waiting time is scanned. By fixing the coherence time a spectral grating is imposed on the sample exciting only specific frequencies in the sample and providing excitation specificity. In practice multiple waiting time scans are collected allowing for the averaging of the measured spectral diffusion which when combined with the waiting time over sampling, allows for an approximate ten-fold increase in signal-to-noise of the data. This increased signal-to-noise of the data has allowed for the detailed analysis of preferential solvation of a bio-inspired molecular probe, described in detail in Chapter 3. The greatly reduced acquisition time offered by RASD allows for the monitoring of transiently stable systems and when combined with phase control methods should prove useful for monitoring protein aggregation pathways.

1.6.2 Compressed Sensing

Compressed Sensing (CS) is a signal processing method that utilizes data sparseness to reduce the data needed for signal reconstruction and has seen widespread adoption across many disciplines.[90-92] In a typical 2D-IR experiment the coherence time is scanned from 0 – 10 ps with a resolution of ~1 fs. In the frequency domain, after Fourier transformation of the coherence time axis, this corresponds to a frequency range of 0- 16,500 cm^{-1} with a resolution of 3.3 cm^{-1}. Due to limitations of mid-IR pulse generation the laser bandwidth is typically limited to 100 cm^{-1}, centered near 2,000 cm^{-1}. Frequencies outside the laser's bandwidth will not be coherently driven in an experiment and do not contribute to the observed 2D-IR signal. Thus considering the 2D-IR spectrum as being comprised of only 33 of the available 5,000 frequency components, the signal sparseness becomes apparent.

Applying the ansatz of sparseness, CS aims at solving a given system of linear equations by minimizing the $l_1$norm, $\|l\|_1 \equiv \sum_{i=1}^{n}|x_i|$.[92] This optimization problem can be thought of as trying to find the minimal number of terms to reconstruct the measured signal. In 2D-IR the system of equations set up for determining the excitation frequencies can be represented by, $\boldsymbol{b} = A\boldsymbol{x}$. The vector $\boldsymbol{b}$ represents the time domain data acquired during a scan of the coherence time. The basis set used, A, is the discrete Fourier transform matrix and $\boldsymbol{x}$ is the sparse frequency representation of the time domain signal. By considering $\boldsymbol{x}$ to be a sparse vector we are able to greatly reduce the scan range of $\boldsymbol{b}$. This scan range is further reduced when the linear vibrational spectrum of the molecule under investigation is taken into account and the number of frequency components in the basis set, A, is reduced. In practice we are able to reduce the scanned range twenty-fold, leading to a significant speedup of the acquisition. As described in detail in Chapter 2, CS allows for further speedup of the 2D-IR method and is able to reproduce the full signal amplitude with high fidelity. Similar to RASD, applications of CS-2D-IR will allow for the study of more transiently stable samples by 2D-IR. By providing the complete 2D-IR spectrum CS allows for a more detailed picture of the system than is provided by RASD however one loses the advantage offered by continuously scanning the waiting time delay.

1.7 Summary of Chapters

2D-IR spectroscopy is a powerful technique, capable of investigating the ultrafast dynamical motions that are fundamental to many processes in biology. The following chapters highlight a significant body of work in developing new 2D-IR methods and probes. Chapter 2 is provided, along with Appendix A, as a detailed description of the application of compressed sensing to 2D-IR. Further details of the implementation of CS-2D-IR are provided in these sections as well, including the full code of the algorithmic solver and a description of the pulse scanning utilized to realize the full advantages of compressed sensing.

Chapters 3 and 4 respectively introduce BTNN-BCT as a molecular probe and demonstrate its utilization in a biological setting. Chapter 3 specifically looks at monitoring the ultrafast dynamics of spectral diffusion to study solvation effects arising from preferential interactions between the probe and cosolvent. This work is supported by a significant body of molecular dynamics simulation data, detailed in chapter 3. The characterization of the preferential solvation of the BTNN-BCT probe, outlined in chapter 3, is interpreted in the context of a protein environment in chapter 4. In the protein system we find that the observed dynamics in the protein system are consistent with the interpretation framework established in chapter 3. Together, these works highlight the use of 2D-IR and metal carbonyls as a sensitive, site-specific, method for the investigation of dynamics in biological systems.

The final chapter of this manuscript, Chapter 5, summarizes the advancements made on the application of 2D-IR towards studying biological systems, detailed in the previous chapters. A forward-looking summary is provided in this chapter taking a prospective look at future improvements of the 2D-IR method made possible by the adoption of new acquisition software and hardware in the lab. 2D-IR as well as other multidimensional spectroscopic methods have seen rapid adoption as the technologies to implement these complex systems are fine-tuned. With the recent realization of a turnkey 2D-IR setup, the application of 2D-IR towards biological systems is expected to grow. Building upon the growing body of work, 2D-IR will offer significant insight on characterizing and understanding the dynamics of biological systems complementary to existing methods.

1.8 References:

1. Brannigan, J. A.; Wilkinson, A. J., Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **2002,** *3* (12), 964-970.

2. Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A., Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **2014,** *32* (2), 99-109.

3. Cracknell, J. A.; Vincent, K. A.; Armstrong, F. A., Enzymes as working or inspirational electrocatalysts for fuel cells and electrolysis. *Chem. Rev.* **2008,** *108* (7), 2439-2461.

4. Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D., De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011,** *6* (5).

5. Kries, H.; Blomberg, R.; Hilvert, D., De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **2013,** *17* (2), 221-228.

6. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De novo computational design of retro-aldol enzymes. *Science* **2008,** *319* (5868), 1387-1391.

7. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A., De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999,** *68*, 779-819.

8. Lim, K. H.; Huang, H.; Pralle, A.; Park, S., Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnol. Bioeng.* **2013,** *110* (1), 57-67.

9. Giese, K. C.; Vierling, E., Mutants in a small heat shock protein that affect the oligomeric state - Analysis and allele-specific suppression. *J. Biol. Chem.* **2004,** *279* (31), 32674-32683.

10. Pal, C.; Papp, B.; Lercher, M. J., An integrated view of protein evolution. *Nat. Rev. Genet.* **2006,** *7* (5), 337-348.

11. Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A., Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **2004,** *14* (2), 208-216.

12. Daugaard, M.; Rohde, M.; Jaattela, M., The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Lett.* **2007,** *581* (19), 3702-3710.

13. Green, N. M., Avidin and Streptavidin. *Methods Enzymol.* **1990,** *184*, 51-67.

14. Hopp, T. P.; Woods, K. R., Prediction of Protein Antigenic Determinants from Amino-Acid-Sequences. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1981,** *78* (6), 3824-3828.

15. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L., Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001,** *305* (3), 567-580.

16. Garnier, J.; Osguthorpe, D. J.; Robson, B., Analysis of Accuracy and Implications of Simple Methods for Predicting Secondary Structure of Globular Proteins. *J. Mol. Biol.* **1978,** *120* (1), 97-120.

17. Sali, A.; Blundell, T. L., Comparative Protein Modeling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993,** *234* (3), 779-815.

18. Kim, D. E.; Chivian, D.; Baker, D., Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004,** *32*, W526-W531.

19. Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D., Protein structure prediction using rosetta. *Numerical Computer Methods, Pt D* **2004,** *383*, 66.

20. Lutz, S., Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **2010,** *21* (6), 734-743.

21. Hellinga, H. W., Rational protein design: Combining theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94* (19), 10015-10017.

22. Bornscheuer, U. T.; Pohl, M., Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **2001,** *5* (2), 137-143.

23. Durbin, S. D.; Feher, G., Protein crystallization. *Annu. Rev. Phys. Chem.* **1996,** *47*, 171-204.

24. Udgaonkar, J. B., Multiple routes and structural heterogeneity in protein folding. *Annual Review of Biophysics* **2008,** *37*, 489-510.

25. Smith, J. L.; Hendrickson, W. A.; Honzatko, R. B.; Sheriff, S., Structural Heterogeneity in Protein Crystals. *Biochemistry* **1986,** *25* (18), 5018-5027.

26. Dominguez, L.; Foster, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D., Structural Heterogeneity in Transmembrane Amyloid Precursor Protein Homodimer Is a Consequence of Environmental Selection. *J. Am. Chem. Soc.* **2014,** *136* (27), 9619-9626.

27. Dhanik, A.; van den Bedem, H.; Deacon, A.; Latombe, J. C., Modeling Structural Heterogeneity in Proteins from X-Ray Data. *Algorithmic Foundations of Robotics Viii* **2010,** *57*, 551-566.

28. DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L., Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004,** *12* (5), 831-838.

29. Plaxco, K. W.; Dobson, C. M., Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.* **1996,** *6* (5), 630-636.

30. Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M., Carbonic anhydrase as a model for biophysical

and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008,** *108* (3), 946-1051.

31. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450* (7171), 913.

32. King, J. T.; Ross, M. R.; Kubarych, K. J., Water-Assisted Vibrational Relaxation of a Metal Carbonyl Complex Studied with Ultrafast 2D-IR. *J. Phys. Chem. B* **2012,** *116* (12), 3754-3759.

33. Loparo, J.; Roberts, S.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. I. Vibrational dynamics in two-dimensional IR line shapes. *J. Chem. Phys.* **2006,** *125* (19), 194521.

34. Loparo, J.; Roberts, S.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. II. Hydrogen bond switching dynamics. *J. Chem. Phys.* **2006,** *125* (19), 194522.

35. Asbury, J.; Steinel, T.; Kwak, K.; Corcelli, S.; Lawrence, C.; Skinner, J.; Fayer, M., Dynamics of water probed with vibrational echo correlation spectroscopy. *J. Chem. Phys.* **2004,** *121*, 12431-12446.

36. Cowan, M.; Bruner, B.; Huse, N.; Dwyer, J.; Chugh, B.; Nibbering, E.; Elsaesser, T.; Miller, R., Ultrafast memory loss and energy redistribution in the hydrogen bond network of liquid H2O. *NATURE* **2005,** *434* (7030), 199-202.

37. Jones, S.; Thornton, J. M., Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (1), 13-20.

38. Dill, K. A., Dominant Forces in Protein Folding. *Biochemistry* **1990,** *29* (31), 7133-7155.

39. Baker, E. N.; Hubbard, R. E., Hydrogen-Bonding in Globular-Proteins. *Prog. Biophys. Mol. Biol.* **1984,** *44* (2), 97-179.

40. Hamm, P.; Lim, M.; Hochstrasser, R., Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **1998,** *102* (31), 6123-6138.

41. GRAENER, H.; SEIFERT, G.; LAUBEREAU, A., NEW SPECTROSCOPY OF WATER USING TUNABLE PICOSECOND PULSES IN THE INFRARED. *Phys. Rev. Lett.* **1991,** *66* (16), 2092-2095.

42. Woutersen, S.; Bakker, H., Resonant intermolecular transfer of vibrational energy in liquid water. *Nature* **1999,** *402* (6761), 507-509.

43. Gale, G.; Gallot, G.; Hache, F.; Lascoux, N.; Bratos, S.; Leicknam, J., Femtosecond dynamics of hydrogen bonds in liquid water: A real time study. *Phys. Rev. Lett.* **1999,** *82* (5), 1068-1071.

44. Stratt, R.; Maroncelli, M., Nonreactive dynamics in solution: The emerging molecular view of solvation dynamics and vibrational relaxation. *J. Phys. Chem.* **1996,** *100* (31), 12981-12996.

45. Sundstrom, V.; Pullerits, T.; van Grondelle, R., Photosynthetic light-harvesting: Reconciling dynamics and structure of purple bacterial LH2 reveals function of photosynthetic unit. *J. Phys. Chem. B* **1999,** *103* (13), 2327-2346.

46. Treuffet, J.; Kubarych, K. J.; Lambry, J. C.; Pilet, E.; Masson, J. B.; Martin, J. L.; Vos, M. H.; Joffre, M.; Alexandrou, A., Direct observation of ligand transfer and bond formation in cytochrome c oxidase by using mid-infrared chirped-pulse upconversion. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (40), 15705-15710.

47. Xie, A. H.; van der Meer, L.; Hoff, W.; Austin, R. H., Long-lived amide I vibrational modes in myoglobin. *Phys. Rev. Lett.* **2000,** *84* (23), 5435-5438.

48. Asplund, M.; Zanni, M.; Hochstrasser, R., Two-dimensional infrared spectroscopy of peptides by phase- controlled femtosecond vibrational photon echoes. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (15), 8219-8224.

49. Golonzka, O.; Khalil, M.; Demirdoven, N.; Tokmakoff, A., Vibrational anharmonicities revealed by coherent two- dimensional infrared spectroscopy. *Phys. Rev. Lett.* **2001,** *86* (10), 2154-2157.

50. Khalil, M.; Demirdoven, N.; Tokmakoff, A., Obtaining absorptive line shapes in two-dimensional infrared vibrational correlation spectra. *Phys. Rev. Lett.* **2003,** *90* (4), 047401.

51. Nee, M. J.; Baiz, C. R.; Anna, J. M.; McCanne, R.; Kubarych, K. J., Multilevel vibrational coherence transfer and wavepacket dynamics probed with multidimensional IR spectroscopy. *J. Chem. Phys.* **2008,** *129* (8), 084503.

52. Baiz, C. R.; Mcrobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J., Two-Dimensional Infrared Spectroscopy of Metal Carbonyls. *Acc. Chem. Res.* **2009,** *42* (9), 1395-1404.

53. Baiz, C.; Nee, M.; McCanne, R.; Kubarych, K., Ultrafast nonequilibrium Fourier-transform two-dimensional infrared spectroscopy. *Opt. Lett.* **2008,** *33* (21), 2533-2535.

54. Jonas, D., Two-dimensional femtosecond spectroscopy. *Annu. Rev. Phys. Chem.* **2003,** *54*, 425-463.

55. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Crowding Induced Collective Hydration of Biological Macromolecules over Extended Distances. *J. Am. Chem. Soc.* **2014,** *136* (1), 188-194.

56. King, J. T.; Kubarych, K. J., Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy. *J. Am. Chem. Soc.* **2012,** *134* (45), 18705-12.

57. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B* **2012,** *116* (19), 5604-5611.

58. Osborne, D. G.; Dunbar, J. A.; Lapping, J. G.; White, A. M.; Kubarych, K. J., Site-Specific Measurements of Lipid Membrane Interfacial Water Dynamics with Multidimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2013,** *117* (49), 15407-15414.

59. Dunbar, J. A.; Osborne, D. G.; Anna, J. M.; Kubarych, K. J., Accelerated 2D-IR Using Compressed Sensing. *J. Phys. Chem. Lett.* **2013,** *4* (15), 2489-2492.

60. Osborne, D. G.; Kubarych, K. J., Rapid and accurate measurement of the frequency-frequency correlation function *J. Phys. Chem. A* **2013,** *117* (29), 5891-8.

61. Roberts, S.; Loparo, J.; Tokmakoff, A., Characterization of spectral diffusion from two-dimensional line shapes. *J. Chem. Phys.* **2006,** *125* (8), 084502.

62. Kwak, K.; Park, S.; Finkelstein, I.; Fayer, M., Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: A new approach. *J. Chem. Phys.* **2007,** *127*, 124503.

63. Marion, D.; Wuthrich, K., Application of Phase Sensitive Two-Dimensional Correlated Spectroscopy (Cosy) for Measurements of H-1-H-1 Spin-Spin Coupling-Constants in Proteins. *Biochem. Biophys. Res. Commun.* **1983,** *113* (3), 967-974.

64. Sinkeldam, R. W.; Greco, N. J.; Tor, Y., Fluorescent Analogs of Biomolecular Building Blocks: Design, Properties, and Applications. *Chem. Rev.* **2010,** *110* (5), 2579-2619.

65. Weiss, S., Fluorescence spectroscopy of single biomolecules. *Science* **1999,** *283* (5408), 1676-1683.

66. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G., Site-Specific Incorporation of Biophysical Probes into Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91* (8), 2910-2914.

67. Tsien, R. Y., The green fluorescent protein. *Annu. Rev. Biochem.* **1998,** *67*, 509-544.

68. Cormack, B. P.; Valdivia, R. H.; Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** *173* (1), 33-38.

69. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green Fluorescent Protein as a Marker for Gene-Expression. *Science* **1994,** *263* (5148), 802-805.

70. Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y., Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **2008,** *5* (6), 545-551.

71. Follenius-Wund, A.; Bourotte, M.; Schmitt, M.; Iyice, F.; Lami, H.; Bourguignon, J. J.; Haiech, J.; Pigault, C., Fluorescent derivatives of the GFP chromophore give a new insight into the GFP fluorescence process. *Biophys. J.* **2003,** *85* (3), 1839-1850.

72. Horng, M. L.; Gardecki, J. A.; Papazyan, A.; Maroncelli, M., Subpicosecond Measurements of Polar Solvation Dynamics - Coumarin-153 Revisited. *J. Phys. Chem.* **1995,** *99* (48), 17311-17337.

73. Fischer, K.; Prause, S.; Spange, S.; Cichos, F.; Von Borczyskowski, C., Surface polarity of cellulose derivates observed by coumarin 151 and 153 as solvatochromic and fluorochromic probes. *Journal of Polymer Science Part B-Polymer Physics* **2003,** *41* (11), 1210-1218.

74. Vallee-Belisle, A.; Michnick, S. W., Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat. Struct. Mol. Biol.* **2012,** *19* (7), 731.

75. Moon, C. P.; Fleming, K. G., Using Tryptophan Fluorescence to Measure the Stability of Membrane Proteins Folded in Liposomes. *Methods in Enzymology: Biothermodynamics, Vol 492, Pt D* **2011,** *492*, 189-211.

76. Vallee-Belisle, A.; Michnick, S. W., Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol.* **2007,** *374* (3), 791-805.

77. Chao, W. C.; Lu, J. F.; Wang, J. S.; Yang, H. C.; Pan, T. A.; Chou, S. C. W.; Wang, L. H.; Chou, P. T., Probing Ligand Binding to Thromboxane Synthase. *Biochemistry* **2013,** *52* (6), 1113-1121.

78. Viriot, M. L.; Carre, M. C.; Geoffroy-Chapotot, C.; Brembilla, A.; Muller, S.; Stoltz, J. F., Molecular rotors as fluorescent probes for biological studies. *Clinical Hemorheology and Microcirculation* **1998,** *19* (2), 151-160.

79. Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T., Two-dimensional IR spectroscopy and segmental C-13 labeling reveals the domain structure of human gamma D-crystallin amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (9), 3329-3334.

80. Woys, A. M.; Mukherjee, S. S.; Skoff, D. R.; Moran, S. D.; Zanni, M. T., A Strongly Absorbing Class of Non-Natural Labels for Probing Protein Electrostatics and Solvation with FTIR and 2D IR Spectroscopies. *J. Phys. Chem. B* **2013,** *117* (17), 5009-5018.

81. Nilsen, I. A.; Osborne, D. G.; White, A. M.; Anna, J. M.; Kubarych, K. J., Monitoring equilibrium reaction dynamics of a nearly barrierless molecular rotor using ultrafast vibrational echoes. *J. Chem. Phys.* **2014,** *141* (13).

82. Osborne, D. G.; King, J. T.; Dunbar, J. A.; White, A. M.; Kubarych, K. J., Ultrafast 2DIR Probe of a Host-Guest Inclusion Complex: Structural and Dynamical Constraints of Nanoconfinement. *J. Chem. Phys.* **2013,** *138*, 144501.

83. King, J. T.; Ross, M. R.; Kubarych, K. J., Ultrafast alpha-Like Relaxation of a Fragile Glass-Forming Liquid Measured Using Two-Dimensional Infrared Spectroscopy. *Phys. Rev. Lett.* **2012,** *108* (15), 157401.

84. King, J. T.; Anna, J. M.; Kubarych, K. J., Solvent-hindered intramolecular vibrational redistribution. *Phys. Chem. Chem. Phys.* **2011,** *13* (13), 5579-5583.

85. Baiz, C. R.; McCanne, R.; Kubarych, K. J., *Structurally-Sensitive Rebinding Dynamics of Solvent-Caged Radical Pairs: Exploring the Viscosity Dependence*. Oxford University Press: New York, 2011.

86. Anna, J. M.; King, J. T.; Kubarych, K. J., Multiple Structures and Dynamics of $[CpRu(CO)_2]_2$ and $[CpFe(CO)_2]_2$ in Solution Revealed with Two-Dimensional Infrared Spectroscopy. *Inorg. Chem.* **2011,** *50* (19), 9273-9283.

87. King, J. T.; Baiz, C. R.; Kubarych, K. J., Solvent-Dependent Spectral Diffusion in a Hydrogen Bonded "Vibrational Aggregate". *J. Phys. Chem. A* **2010,** *114* (39), 10590-10604.

88. Anna, J. M.; Kubarych, K. J., Watching solvent friction impede ultrafast barrier crossings: A direct test of Kramers theory. *J. Chem. Phys.* **2010,** *133* (17), 174506.

89. Baiz, C. R.; McRobbie, P. L.; Preketes, N. K.; Kubarych, K. J.; Geva, E., Two-Dimensional Infrared Spectroscopy of Dimanganese Decacarbonyl and Its Photoproducts: An Ab Initio Study. *J. Phys. Chem. A* **2009,** *113* (35), 9617-9623.

90. Yang, J. C.; Wright, J.; Huang, T. S.; Ma, Y., Image Super-Resolution Via Sparse Representation. *Ieee Transactions on Image Processing* **2010,** *19* (11), 2861-2873.

91. Lustig, M.; Donoho, D.; Pauly, J. M., Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* **2007,** *58* (6), 1182-1195.

92. Donoho, D. L., Compressed sensing. *Ieee Transactions on Information Theory* **2006,** *52* (4), 1289-1306.

Chapter Two
Accelerated 2D-IR Using Compressed Sensing

2.1 Introduction

Two-dimensional infrared (2D-IR) spectroscopy is a powerful tool to study the dynamics and structure of molecules in condensed phases. 2D-IR spectra offer important insight into the coupling between molecular vibrations, the dynamics of vibrational energy transfer, vibrational and orientational relaxation, as well as chemical exchange and spectral diffusion [1-10]. Given the increasingly clear chemical applications of 2D-IR spectroscopy, it is desirable to be able to combine traditional thermodynamic variables with 2D-IR spectroscopy. Any acceleration in data acquisition can enable a wide range of investigations which are currently impractical with traditional Fourier scanning methods.

As has become routine in NMR, dynamical information can be obtained using optical pulse sequences that do not provide the full multidimensional response. Recently, for example, we demonstrated a new method for rapidly measuring the frequency-frequency correlation function by combining approaches of three-pulse photon echo spectroscopy with an “inhomogeneity index” inspired by 2D spectroscopy [11]. Compressed sensing (CS) offers another means of obtaining spectral information without the exhaustive sampling required for Fourier transform 2D spectroscopy. CS has already shown promise in many fields including 2D electronic spectroscopy [12-13], impulsive stimulated Raman [14], magnetic resonance imaging [15-16], image processing [17-18], astronomy [19], genomics [20-21], and NMR [22-23]. In this letter, we report the implementation of a new data collection method for accelerated 2D-IR, which can decrease the acquisition time for collecting individual 2D-IR spectra over sixteen-fold using rapidly moving wedges for the coherence time delay. Additionally, we present a detailed comparison (see Appendix A) of compressed sensing (CS) and traditional Fourier analysis on the same data set, finding CS's nearly perfect reproduction of time-dependent peak amplitudes despite using only a small fraction of the measured data.

The bottleneck in Fourier transform 2D spectroscopy is the need to scan the coherence time delay (t_1) between the first two pulses of the echo sequence. In conventional 2D-IR spectroscopy, the excitation frequency resolution $\Delta\omega_1$ is equal to $2\pi / t_1^{max}$, where t_1^{max} is the maximum scanned t_1 delay (shown in Figure 2.1). Intuitively, this temporal duration corresponds to twice the beat period of two waves that differ in frequency by Δω, which beat at a frequency of Δω, and hence

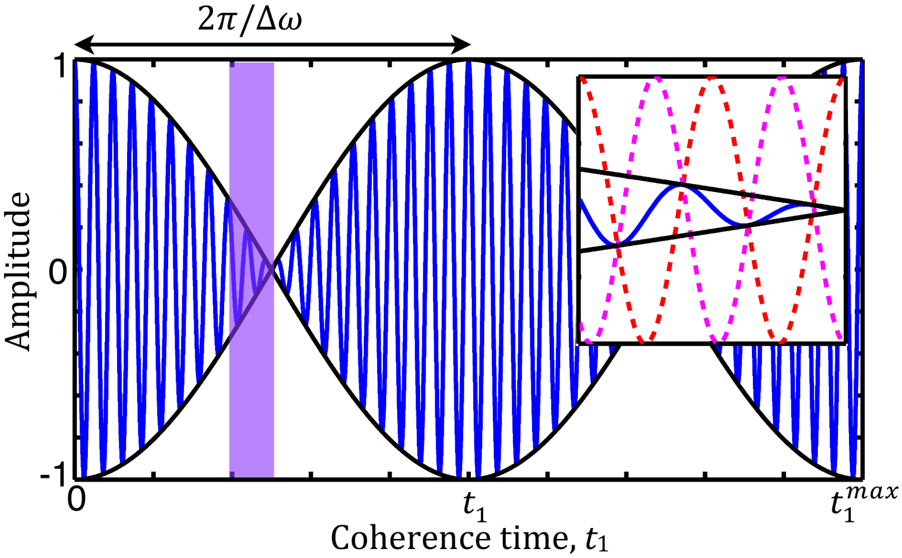

Figure 2.1 Two waves differing in frequency by Δω produce a wave (blue line) whose amplitude is modulated by the difference frequency. Due to the amplitude modulation (black line) during a relatively short time window highlighted in purple (see inset), Compressed Sensing is able to determine the frequency composition of the signal (pink and red dashed lines).

have a period of 2π/Δω. The FT resolution amounts to the minimum beat frequency that fits in half the temporal sampling window. If one measures only a small fraction of the full window, it is not possible to determine the two underlying frequencies deterministically. Nevertheless, a signal exhibiting amplitude modulation indicates that at least one other frequency is present and is causing interference. Compressed sensing exploits the amplitude modulation, and aims to find the smallest set of frequencies that can produce a given signal. Intuitively, it is clear that the beating of two frequencies is most apparent near the nodes of the modulation (purple window in Figure 2.1). CS is capable of providing accurate signal reproduction with effective spectral resolution much higher than would be possible by directly Fourier transforming the temporal data.

In FT spectroscopy, the discrete Fourier transform (DFT) matrix is used to calculate the Fourier transform (FT). In CS a DFT matrix is chosen such that it provides finer frequency spacing than would be justified using Fourier analysis of the smaller data set. Since the time domain signal is not artificially extended, the DFT matrix used is no longer square, resulting in an underdetermined system with more unknown frequency components than time data points. The essence of CS reconstruction is to apply an ansatz of sparseness in the frequency domain, where sparseness is invoked merely as an algorithmic requirement, not because the underlying spectral features are truly sparse. Viewed in another way, the very short temporal window used in CS effectively pretends that the dephasing is very (or infinitely) slow. In the context of 2D-IR spectroscopy, the relative peak amplitudes of the frequency data and their waiting time dynamics can be obtained from data sets with many fewer t_1 points using CS than would be required for a conventional FT analysis. Our current implementation of CS-accelerated 2D-IR is not able to recover the 2D spectral peak shape, but does accurately reproduce peak amplitudes (see Appendix A). In 2D-IR, CS permits a remarkable reduction of the scanned coherence time window, allowing for the implementation of new experimental approaches to chemical problems, where ultrafast dynamical measurements can be recorded under varying thermodynamic conditions.

2.2 Experimental Methods

2.2.1 Sample Preparation

For all experiments the well characterized benchmark system dicarbonyl-acetylacetonato-rhodium(I) (RDC) [2, 24-26] was studied in a hexane solution. RDC was purchased from Sigma Aldrich and was used without further purification. Approximately 8 mM solutions were prepared and filtered to remove particulates before use. The sample cell used a 100 μm Teflon spacer between two 3 mm x 25 mm diameter calcium fluoride windows.

2.2.2 2D-IR Setup

The 2D-IR setup has been previously described. [27-28] Three fields E_1, E_2, and E_3 are 100 fs infrared pulses generated by a white-light seeded optical parametric amplifier followed by difference frequency generation. The signal is made collinear with a reference pulse before sum-frequency-generation with a chirped 800 nm pulse to allow for heterodyne detection on a 1340 x 100 pixel CCD camera synchronized to the laser at 1 kHz. The coherence time, t_1, is scanned continuously with a pair of 7° apex ZnSe wedges mounted to a translation stage using an optically encoded, 7 nm resolution, DC motor. A digital signal processor (DSP) monitors the value of the encoder as well as controls the movement of the motor. The position of the motor is recorded with a National Instruments Data Acquisition Card (DAC) that records the position value of the motor when triggered.

A typical complete Fourier transform 2D-IR experiment involves collecting over 10,000 1D spectra over a coherence time of approximately 10 ps. Using Fourier transform methods this range of coherence times allows 2D-IR to detect frequencies from 0 cm^{-1} to approximately 16,700 cm^{-1} with a 3.3 cm^{-1} frequency resolution. The 2D-IR spectrum usually focuses on a relatively narrow frequency window of about 100 cm^{-1}, limited by the bandwidth of the pulses used as well as the finite phase matching bandwidth of the upconversion detection. Windowing the spectra means only a fraction of the possible frequencies have meaningful amplitude. By considering the frequencies outside the range of interest to have zero amplitude, the sparseness of 2DIR is apparent. The goal of compressed sensing is to restrict attention to only the frequencies that are contained in the signal, assuming that all other frequency components are of zero amplitude.

2.2.3 Compressed Sensing Pulse Sequence

Details of the new method used in this paper to collect data for CS analysis follow (see Figure 2.2 for a graphical representation of the method). First the t_1

motor is set to scan continuously between two points using the DSP. A range in the middle of the scan corresponding to approximately 500 fs (117,000 encoder values) is chosen, for which a digital output voltage is set to high on the DSP. The collection range is overscanned to avoid artifacts caused by abrupt changes in motor direction. The digital output pin from the DSP is logically combined with the 1 kHz pulse train of the laser, using an AND gate, to trigger the camera and DAC to record spectra and positions, respectively.

A second DSP is also programed to monitor the state of first DSP's logic pin, and upon detecting a downward edge (indicating the t_1 motor has moved out of range) steps the waiting time motor. This sets up the three-pulse sequence where t_2 is held fixed while the t_1 time delay is continually scanned in a loop; t_2 is stepped after each scan of t_1. A 2DIR spectrum can be generated from scanning the motor in either direction, which in principle could reduce the acquisition time by a factor of two. With the motors and stages we employ, however, significant amplitude differences in forward-scanned and backward-scanned 2D spectra require that we only analyze one set of data. We anticipate that choosing mechanical components

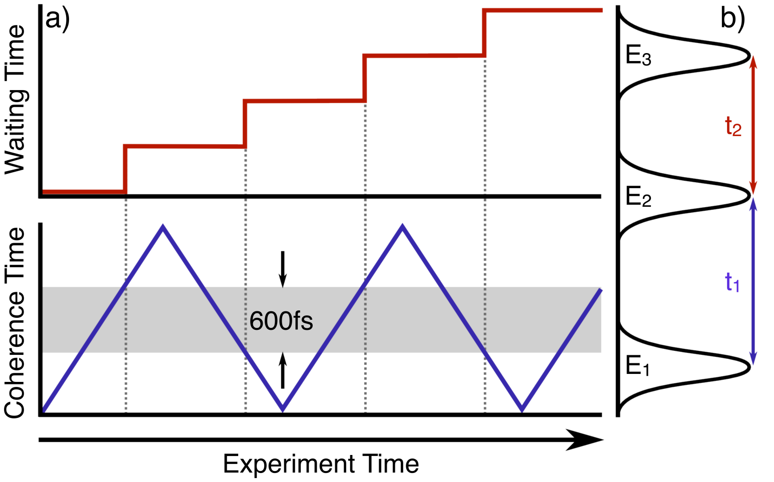

Figure 2.2 Graphical depiction of data collection method for CS analysis. The coherence time is continuously scanned (blue line) between two set points. A range corresponding to 600 fs (grey region) is used to collect spectra. When the coherence time motor leaves the collection range (dashed lines) the waiting time motor is stepped (red line). The three-pulse sequence that is created by this method is depicted to the right (b).

appropriate for the new demands presented by compressed sensing would allow recording reliable spectra by scanning in both directions. Using the digital pin of the DSP to gate when spectra are recorded a full set of waiting times (104 2D-spectra) can be stored in the camera's memory, allowing the camera buffer to only be read off once per experiment, further reducing the time needed to obtain data.

2.2.4 Compressed Sensing Data Analysis

The single file of spectra is parsed into a three-dimensional array (detection frequency x coherence time x waiting time) prior to analysis with the MP algorithm. Parsing is done by using the coherence motor position data acquired by the DAQ for each camera frame to determine the frames corresponding to the coherence motor switching directions. Grouping sequential frames between the turning points of the coherence motor we build a slice of our data array at a single waiting time. Waiting times are assigned by using the waiting time motor position recorded by the DAQ for each spectra and using a known calibration to convert the motor position to a time delay.

The assignment of excitation frequencies in the implementation of CS as described only requires knowledge of coherence time step size, Δt_1, obtained from a calibration of the t_1 motor delay. Using the Discrete Fourier Transform (DFT) matrix as the basis set, frequencies can be assigned to the columns by linearly interpolating a vector spanning from 0 Hz to $2\pi/\Delta t_1$ Hz so that the vector's length corresponds to the number of columns in the matrix. One can also trim the DFT matrix, removing columns (frequencies) that are known not to contribute to the signal, in order to speed up calculations. In this case assignment of frequencies requires knowledge of Δt_1, the size of the original DFT matrix, and which columns of the original matrix were used in the CS algorithm.

2.3 Results and Discussion

Compressed sensing aims to find the smallest set of frequencies that will reproduce a time domain signal, which in the present case is the t_1 dependent signal recorded at each detection frequency ω_3. CS reconstructions rely on sparseness, where given a particular basis set, a signal can be constructed such that most of the coefficients are zero. Sparseness allows for one to approximately solve an underdetermined system of equations by taking the solution with the smallest number of non-zero coefficients [29-30]. The number of non-zero values in a set is defined as the l_0-norm. Hence, CS algorithms minimize the l_0-norm, a combinatorial problem requiring all possible solutions to be enumerated before the problem can be solved. Due to the cost of enumerating all possible solutions, the minimal l_0-norm is often approximated by various methods when implementing CS [31].

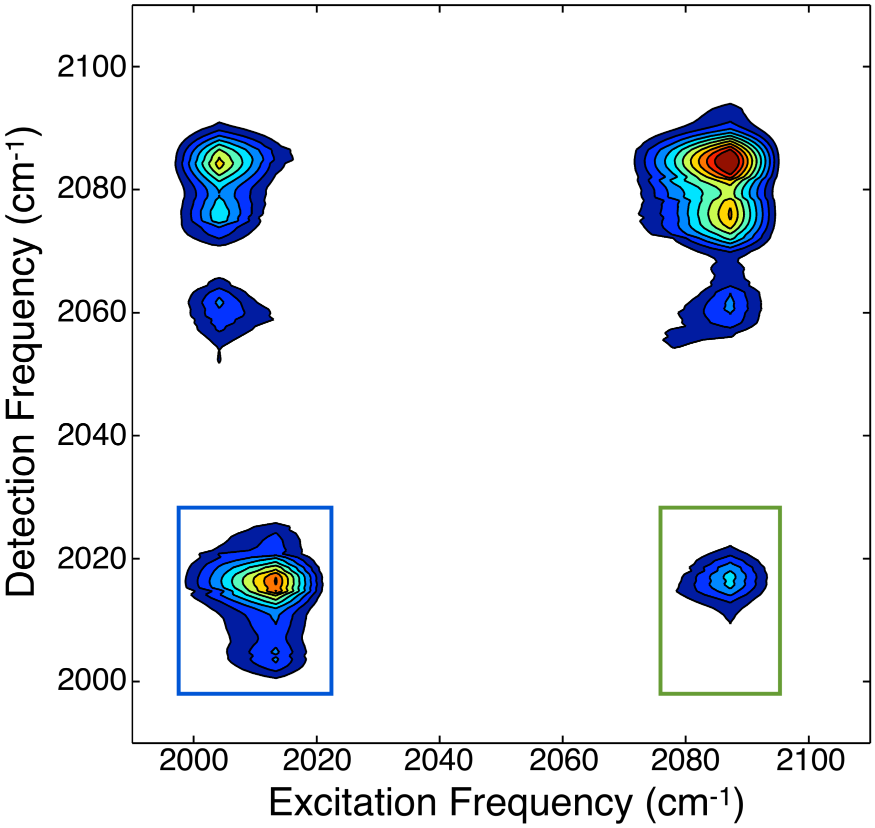

Figure 2.3 2D-IR rephasing spectrum of RDC in hexane at 8.9 ps waiting time. The spectrum was constructed using the MP algorithm with a 503 fs coherence time window and ε =1x10^{-3}. A 7 cm^{-1} shift is observed in the excitation frequency of the upper cross peak.

In this work we used a Matching Pursuit (MP) [31-33] solver to approximate the l_0-norm. This solver uses a "greedy" algorithm to iteratively reconstruct the spectra from the collected data and serves as a drop-in replacement for the FT used traditionally (see Appendix A for code and mathematical formulation). While it is known that the MP algorithm does not explicitly solve the system for the minimum l_0-norm [31-32], it proved to be much less computationally costly and to yield more accurate reconstructed spectra than the popular SPGL1 solver [34]. The MP algorithm requires four input parameters: a reduced DFT matrix, t_1 data for each detection frequency, a noise threshold (ε), and the number of iterations per reconstruction. To generate the reduced DFT matrix, rows of the DFT matrix are truncated to match the length of the time domain signal. For the input data, the algorithm is given a matrix from a single waiting time that contains the coherence time data at each detection frequency. The noise threshold and number of iterations typically depend on the details of the experimental data, although it was found that spectra typically converge in fewer than 200 iterations using $\varepsilon = 5\times10^{-3}$.

To validate the use of CS in the calculation of 2D-IR spectra, we tested the ability of the new method to reproduce the well-characterized response of RDC. The spectrum in Figure 2.3was produced by scanning a 503 fs coherence time window, t_1 = 477 fs to t_1 = 980 fs corresponding to scanning through a node in the amplitude modulation, sampled uniformly with steps of 0.791 fs at a waiting time of 8.9 ps and was analyzed with the MP algorithm. By oversampling t_1 using continually scanned wedges we avoid the need to sample our time points randomly to account for frequency components below the actual sampling rate [12, 35], as is often done in CS. The CS-accelerated 2D-IR spectrum reproduces well the features of the response of RDC. It was noticed that excitation frequencies of the cross peaks became slightly shifted with shorter t_1 windows but still exhibited waiting time dynamics similar to the full FT 2D spectra (see SI). This shifting is consistent with previously reported frequency variations[12, 36] and is seen in the 7 cm^{-1} shift of the upper left cross peak in Figure 2.3. To further test our CS implementation, we analyzed various peak volumes as a function of the waiting time. Figure 2.4a shows the rephasing amplitude trace for the low frequency diagonal peak, Figure 2.3 blue box, as well as the lower cross peak, Figure 2.3 green box. Fitting the diagonal peak rephasing amplitudes with a double exponential yields a decay time of 60 ps, which agrees well with the value of 60 ps obtained using traditional Fourier analysis [26]. Further validation of the CS analysis can be found in the quantum-beat modulated cross peak volumes where coherent beating is clearly seen between the high and low frequency modes of RDC. Fourier transforming the coherence signal returns the splitting between the two fundamental vibrations to be 69 cm^{-1} (Figure 2.4b), which is again consistent with known values.

The applicability of CS to 2D-IR spectroscopy as a replacement for the FT is apparent in its success in reproducing both the spectrum and t_2 dependent peak volumes. We have shown that CS offers a promising alternative to the Fourier transform in 2D-IR spectroscopy by analyzing conventionally measured data (see SI) and data collected specifically for CS analysis using the accelerated scanning approach introduced here. CS accurately reproduces the main features of a 2D-IR spectrum using less than one-sixteenth the temporal data traditionally needed. The accuracy of CS-determined peak frequencies is known to be dependent on the length of the input time signal and care must be taken to avoid undersampling the data.

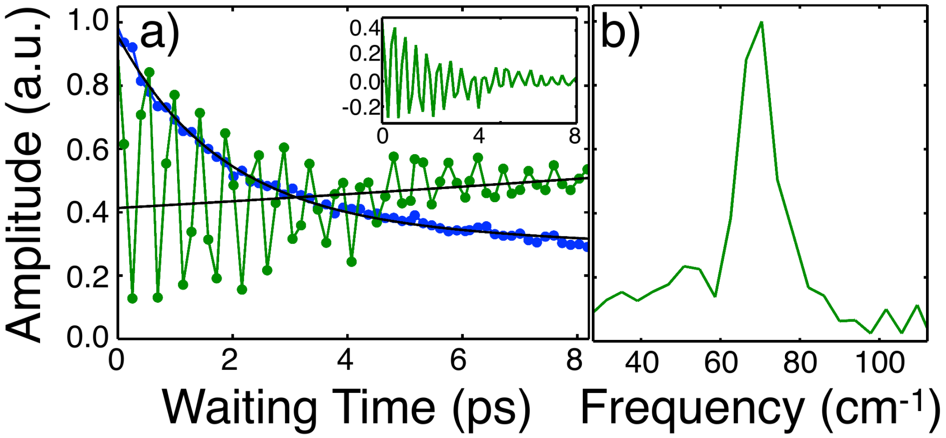

Figure 2.4 (a) Integrated rephasing amplitudes for the low frequency diagonal cross peak (blue) and the (2015, 2084 cm^{-1}) cross peak (green), with biexponential fits (lines). The coherence beating pattern in the cross peak is apparent by looking at the residuals of the fit (inset). (b) The Fourier transform of the quantum beats reveals the 69 cm^{-1} splitting between the two modes.

2.4 Conclusions

The ability to quickly obtain a single 2D-IR spectrum, less than 0.7 seconds acquisition time on a 1 kHz system, and to collect multiple spectra in rapid succession, over 200 spectra in under 9 minutes, greatly speeds up the acquisition of 2D-IR. Further speedup in collecting successive spectra is straightforward by combining CS with pulse shaping methods [37-38]. The drastic reduction in acquisition time made possible by CS opens up the ability of 2D-IR to easily study the waiting time dependent spectral changes in systems of varying pH, ionic strength, concentration, or other chemically important parameters. We note that the speed improvement found here is fully general, and is not dependent on the specifics of the signal level or spectral region. Any 2D spectrum can benefit from compressed sensing, though the benefits of speed may not outweigh the lack of lineshape information for some applications. Although spectral diffusion measurements were not considered in the context of CS because rapid methods for collecting this information already exist [11], CS should still be amenable to the analysis of spectral diffusion by the inhomogeneity index, where only the integrated peak volumes from the rephasing and non-rephasing spectra are needed [6, 39].

2.5 Acknowledgements

This work was published in the Journal of Physical Chemistry Letters under the title: Accelerated 2D-IR Using Compressed Sensing. *J. Phys. Chem. Lett.* **2013,** *4* (15), 2489-2492.

2.6 References:

1. Anna, J. M.; King, J. T.; Kubarych, K. J., Multiple Structures and Dynamics of [CpRu(CO)2]2 and [CpFe(CO)2]2 in Solution Revealed with Two-Dimensional Infrared Spectroscopy. *Inorg. Chem.* **2011,** *2*, 9273-9283.

2. Anna, J. M.; Kubarych, K. J., Watching solvent friction impede ultrafast barrier crossings: A direct test of Kramers theory. *J. Chem. Phys.* **2010,** *133*, 174506.

3. Anna, J. M.; Ross, M. R.; Kubarych, K. J., Dissecting Enthalpic and Entropic Barriers to Ultrafast Equilibrium Isomerization of a Flexible Molecule Using 2DIR Chemical Exchange Spectroscopy. *J. Phys. Chem. A* **2009,** *113*, 6544-6547.

4. King, J.; Ross, M.; Kubarych, K., Ultrafast α-Like Relaxation of a Fragile Glass-Forming Liquid Measured Using Two-Dimensional Infrared Spectroscopy. *Phys. Rev. Lett.* **2012,** *108* (15), 157401.

5. King, J. T.; Anna, J. M.; Kubarych, K. J., Solvent-hindered intramolecular vibrational redistribution. *Phys. Chem. Chem. Phys.* **2011,** *13*, 5579-5583.

6. King, J. T.; Baiz, C. R.; Kubarych, K. J., Solvent-dependent spectral diffusion in a hydrogen bonded "vibrational aggregate". *J. Phys. Chem. A* **2010,** *114*, 10590-10604.

7. King, J. T.; Ross, M. R.; Kubarych, K. J., Water-Assisted Vibrational Relaxation of a Metal Carbonyl Complex Studied with Ultrafast 2D-IR. *J. Phys. Chem. B* **2012,** *116*, 3754-3759.

8. Loparo, J. J.; Roberts, S. T.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. II. Hydrogen bond switching dynamics. *The Journal of chemical physics* **2006,** *125*, 194522.

9. Nee, M. J.; Baiz, C. R.; Anna, J. M.; McCanne, R.; Kubarych, K. J., Multilevel vibrational coherence transfer and wavepacket dynamics probed with multidimensional IR spectroscopy. *J. Chem. Phys.* **2008,** *129* (8), 084503.

10. Zheng, J.; Kwak, K.; Asbury, J.; Chen, X.; Piletic, I. R.; Fayer, M. D., Ultrafast dynamics of solute-solvent complexation observed at thermal equilibrium in real time. *Science* **2005,** *309*, 1338-1343.

11. Osborne, D. G.; Kubarych, K. J., Rapid and Accurate Measurement of the Frequency-Frequency Correlation Function. *J. Phys. Chem. A* **2012**.

12. Sanders, J. N.; Saikin, S. K.; Mostame, S.; Andrade, X.; Widom, J. R.; Marcus, A. H.; Aspuru-guzik, A., Compressed Sensing for Multidimensional Spectroscopy Experiments. *J. Phys. Chem. Lett.* **2012,** *3*, 2697-2702.

13. Almeida, J.; Prior, J.; Plenio, M. B., Computation of Two-Dimensional Spectra Assisted by Compressed Sampling. *J. Phys. Chem. Lett.* **2012,** *3*, 2692-2696.

14. Katz, O.; Levitt, J. M.; Silberberg, Y., Compressive Fourier Transform Spectroscopy. *arXiv:1006.2553 [physics.optics]* **2010**.

15. Lustig, M.; Donoho, D.; Pauly, J. M., Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **2007,** *58*, 1182-1195.

16. Goldstein, T.; Osher, S., The Split Bregman Method for L1-Regularized Problems. *SIAM J. Imag. Sci.* **2009,** *2*, 323-343.

17. Duarte, M. F.; Davenport, M. A.; Takhar, D.; Laska, J. N.; Sun, T.; Kelly, K. F.; Baraniuk, R. G., Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Process. Mag.* **2008,** (March), 83-91.

18. Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S., Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009,** *31*, 210-227.

19. Bobin, J.; Starck, J.-l.; Ottensamer, R., Compressed Sensing in Astronomy. *IEEE J. Sel. Topics Signal Process.* **2008,** *2*, 718-726.

20. Amir, A.; Zuk, O., Bacterial community reconstruction using compressed sensing. *J. Comp. Bio.* **2011,** *18*, 1723-1741.

21. Erlich, Y.; Gordon, A.; Brand, M.; Hannon, G. J.; Mitra, P. P., Compressed Genotyping. *IEEE Trans. Inf. Theory* **2010,** *56*, 706-723.

22. Holland, D. J.; Bostock, M. J.; Gladden, L. F.; Nietlispach, D., Fast multidimensional NMR spectroscopy using compressed sensing. *Angew. Chem., Int. Ed.* **2011,** *50*, 6548-6551.

23. Kazimierczuk, K.; Orekhov, V. Y., Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem., Int. Ed.* **2011,** *50*, 5556-5559.

24. Demirdoven, N.; Khalil, M.; Golonzka, O.; Tokmakoff, A., Correlation Effects in the Two-Dimensional Vibrational Spectroscopy of Coupled Vibrations. *J. Phys. Chem. A* **2001,** *105*, 8025-8030.

25. Khalil, M.; Demirdöven, N.; Tokmakoff, A., Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution. *J. Phys. Chem. A* **2003,** *107*, 5258-5279.

26. Khalil, M.; Demirdöven, N.; Tokmakoff, A., Vibrational coherence transfer characterized with Fourier-transform 2D IR spectroscopy. *J. Chem. Phys.* **2004,** *121*, 362-373.

27. Anna, J. M.; Nee, M. J.; Baiz, C. R.; Mccanne, R.; Kubarych, K. J., Measuring absorptive two-dimensional infrared spectra using chirped-pulse upconversion detection. *J. Opt. Soc. Am. B: Opt. Phys.* **2010,** *27*, 382-393.

28. Nee, M. J.; Mccanne, R.; Kubarych, K. J., Two-dimensional infrared spectroscopy detected by chirped pulse upconversion. *Opt. Lett.* **2007,** *32*, 713-715.

29. Candes, E. J.; Wakin, M. B., An Introduction To Compressive Sampling. *IEEE Signal Process. Mag.* **2008,** (March), 21-30.

30. Donoho, D. L., Compressed sensing. *IEEE Trans. Inf. Theory* **2006,** *52*, 1289-1306.

31. Chen, S. S.; Donoho, D. L.; Saunders, M. A., Atomic Decomposition by Basis Pursuit. *SIAM Rev.* **2001,** *43*, 129-159.

32. Andrecut, M., Fast GPU Implementation of Sparse Signal Recovery from Random Projections. *Eng. Lett.* **2009,** *17*, 151-158.

33. Mallat, S. G.; Zhang, Z., Matching Pursuit With Time-Frequency Dictionaries. *IEEE Trans. Signal Process.* **1993,** *41*, 3397-3315.

34. Berg, E. V. D.; Friedlander, M. P., Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comp.* **2008,** *31*, 890-912.

35. Hennenfent, G.; Herrmann, F. J., Simply denoise : wavefield reconstruction via jittered undersampling. *Geophysics* **2008**, 19-28.

36. Andrade, X.; Sanders, J. N.; Aspuru-Guzik, A., Application of compressed sensing to the simulation of atomic systems. *Proc. Natl. Acad. Sci. USA* **2012,** *109* (35), 13928-13933.

37. Shim, S.-H.; Strasfeld, D. B.; Ling, Y. L.; Zanni, M. T., Automated 2D IR spectroscopy using a mid-IR pulse shaper and application of this technology to the human islet amyloid polypeptide. *Proc. Natl. Acad. Sci. USA* **2007,** *104*, 14197-14202.

38. Moran, S. D.; Marie, A.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T., Two-dimensional IR spectroscopy and segmental 13C labeling reveals the domain structure of human γD-crystallin amyloid fibrils. *Proc. Natl. Acad. Sci. USA* **2011**, 3329-3334.

39. Roberts, S. T.; Loparo, J. J.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. II. Hydrogen bond switching dynamics. *J. Chem. Phys.* **2006,** *125*, 194522.

Chapter Three

Ultrafast 2D-IR and Simulation Investigations of Preferential Solvation and Cosolvent Exchange Dynamics

3.1 Introduction

Biomolecule hydration is central to macromolecular structure and function in living organisms[1-4], yet detailed experimental probes of interfacial structure and dynamics remain challenging. Considerable recent progress has been made in assessing the degree and nature of biomacromolecular perturbation of hydration water structure and dynamics using spectroscopy[5-7] and computation.[8-11] To probe hydration dynamics at biomolecule interfaces, we have coupled metal carbonyl vibrational labels with ultrafast two-dimensional infrared (2D-IR) spectroscopy, accessing the hydration dynamics of macromolecules such as proteins[12-14] and lipid bilayers.[15] These investigations are consistent with the following picture: in the vicinity of extended interfaces, the orientational motion of hydrating water is modestly slowed by a factor of 2-3 relative to the bulk liquid. The biomolecule's surface limits the availability of three-body transition state configurations, thus raising the free energy barrier to large angular jumps because of the transition state's entropy decrease. Adding cosolutes, such as glycerol, polymers, or more protein, generally slows the water dynamics, but large, crowding species induce a dynamical transition due to the collective hydration of multiple extended interfaces.

In contrast to our previous studies, here we consider a cosolvent system of two liquids, water and *N*,*N*-dimethylformamide (DMF), where we expect a certain degree of preferential solvation by the non-aqueous component. We establish the preference for DMF solvation using a probe solute that is nearly insoluble in pure water, but highly soluble in DMF. We employ a vibrationally labeled flexible biomolecule consisting of biotin derivatized with a tripodal metal carbonyl complex (Figure 3.1), to characterize the solvation dynamics in binary water/DMF solutions.

This ternary system provides a model to investigate several aspects of biomolecule hydration, such as preferential solvation as well as solvent-induced conformational modulation of the solute.

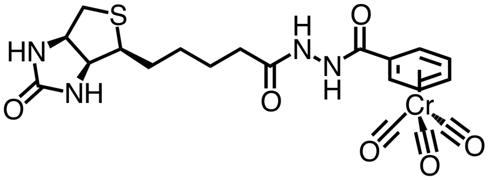

Figure 3.1 Chemical structure of Biotin Hydrazide Benzoyl Chromium Tricarbonyl (BTNN-BCT).

Solvent mixtures appear in numerous chemical and biochemical contexts, from chemical synthesis to protein denaturation. In biochemistry, mixed solvents are frequently employed to increase the solubility of proteins, substrates, and other metabolites. Though cosolvents can act as denaturants or stabilizers of protein structure, a detailed understanding of the nature of these interactions is still incomplete.[16] Since cosolvents, buffers and other osmolytes can associate directly with biomolecule interfaces, a substrate's approach to and exit from an enzyme's active site is determined by the potentially complex transport properties of heterogeneous solvation as well as the dynamical perturbation caused by the extended biomacromolecular interface. In principle, the present biotin/DMF/water system provides a window into both preferential solvation and the associated dynamics of motion within and between solvent shells.

Biotin is a key coenzyme for carboxylase enzymes serving as a shuttle for the bicarbonate ion through the enzyme[17], and is widely recognized for its unique ability to bind streptavidin and its homologues, with femtomolar affinity.[18] Chemically modifying the terminal carbonyl of the biotin tail has made biotin analogues ubiquitous across a wide range of fields ranging from molecular biology to analytical chemistry.[19] In particular, biotin has been used to install precisely anchored organometallic catalysts within the streptavidin host, where a combination of protein engineering and catalyst modifications have enabled a new class of hybrid enzymes.[20-23]

Biotin is amphiphilic but only slightly soluble in aqueous solutions, *ca.* 0.9 mM in water[24], and is often introduced into the aqueous phase through a miscible carrier solvent such as DMF where it shows roughly an order of magnitude greater solubility. In addition to the dynamics under preferential solvation conditions, there is the additional possibility that structural changes associated with varying solution composition might alter the solution dynamics. In previous reports of biotin in aqueous solution using NMR and molecular dynamics simulations, biotin was found to exist in rapid equilibrium between three states: folded, semi-folded, and extended.[25] Hence, the ternary system studied has the potential to use modifications of the solvent environment of the BTNN-BCT solute to manipulate the structural ensemble.

2D-IR is a proven method capable of elucidating the complex dynamics associated with hydration dynamics of water in the bulk[26-30], at surfaces[15, 31], and in other complex environments. Leveraging the well-defined metal carbonyl vibrational modes of BTNN-BCT, 2D-IR allows for the measurement of the timescales for dynamical fluctuations in the vicinity of the probe. 2D-IR correlates excited and detected frequencies, and by analyzing the time-dependent decay of spectral inhomogeneity, it is possible to measure the frequency fluctuation correlation function (FFCF). The FFCF reports on the loss of frequency correlation of the probe transition resulting from equilibrium fluctuations of the probe's environment as well as internal motions, a process denoted "spectral diffusion". Spectral diffusion describes the stochastic sampling of the vibrational line shape as a random walk with steps induced by fluctuations of the solvent and solute. Multiple studies have shown the sensitivity of the FFCF decay to report on the local solvent dynamics[32], including the orientational motion of water at the surface of proteins and biomembranes.[12, 14-15, 33]

By measuring the FFCF at multiple concentrations of DMF/water solutions, we can correlate the observed local dynamical changes to the mole fraction of the DMF cosolvent (x_{DMF}). Indeed, we find the spectral diffusion time scales to exhibit a pronounced DMF concentration dependence, suggesting both a structural and dynamical variation that is dictated by solution composition. Despite the fact that metal carbonyl spectral diffusion in both water and DMF alone occurs on roughly the same time scale, we observe a marked slowdown in spectral diffusion as we dilute the preferred DMF solvent. Molecular dynamics simulations complement the information gained from 2D-IR, providing structural snapshots of the solvation environment around BTNN-BCT probe at varying x_{DMF}. Together these data provide an explanation where the slowdown can be attributed to the exchange of roughly one water-DMF pair.

3.2 Methods

3.2.1 Materials

(Ethyl benzoate) chromium tricarbonyl, N,N-Diisopropylcarbodiimide (DIC), anhydrous N,N-Dimethylformamide (DMF), and deuterium oxide were purchased from Sigma-Aldrich and used without further purification. EZ-Link Hydrazide-Biotin (BTNN, spacer arm 15.7 Å) was purchased from Thermo Scientific and used as received.

3.2.2 BTNN-BCT Synthesis

Biotin Hydrazide Benzoyl Chromium Tricarbonyl (BTNN-BCT) is synthesized using a modified Steglich esterification protocol. (Ethyl benzoate) chromium tricarbonyl, 240 mg, is first subjected to hydrolysis using 400 mg NaOH in 30 mL H_2O with stirring for *ca.* 16 hr. The unreacted ester is extracted through washing the solution 3x with 5 mL diethyl ether. (Benzoate) chromium tricarbonyl is then precipitated from the aqueous phase by addition of HCl and recovered through subsequent extractions with diethyl ether, 5x 3 mL. Combination of the ether fractions and evaporation of the solvent yields (benzoate) chromium tricarbonyl (BCT), dark orange crystals.

To 100 μL anhydrous DMF on ice, 8 mg BCT, 18 mg BTNN, and 11.4 μL DIC was added. The reaction was stirred on ice for one minute before the addition of 0.5 mg of 4-dimethylaminopyridine (DMAP). The reaction was stirred on ice for five additional minutes, followed by stirring at room temperature for three hours.

The BTNN-BCT product was purified from the reaction mixture by separation on an alumina column using a 2:3 volumetric ratio of ethyl acetate and heptane as the mobile phase. The product, fast moving light-orange band, was collected and the solvent was evaporated, yielding the final product. Chemical composition was confirmed by mass spectrometry and infrared spectroscopy.

3.2.3 Sample Preparation

2.5 mM solutions of BTNN-BCT in pure DMF, 75 % (v/v), 50 % (v/v), 35 % (v/v), and 25 % (v/v), DMF in D_2O were prepared using a 10 mM stock solution of BTNN-BCT in DMF. Using the bulk densities for DMF and D_2O, mole fractions were calculated for the solutions. Samples were placed in a custom sample cell utilizing two round 25 mm x 3 mm calcium fluoride windows with a 100 μm Teflon spacer.

3.2.4 Four-Wave Mixing Experiments

Frequency fluctuation correlation function (FFCF) measurements were preformed using rapidly acquired spectral diffusion (RASD) as previously described.[34-35] RASD, a modified implementation of the traditional photon-echo peak shift technique, allows for collection of the FFCF with much faster acquisition times than traditional 2D-IR pulse sequences. This is achieved by holding the first two pulses of a non-collinear 2D-IR pulse sequence at a fixed t_1 time delay, generating a frequency grating in the sample that preferentially excites the chosen transition. Following the two excitation pulses, the third pulse continuously scans the waiting time (t_2). Alternating the time ordering of the first two excitation pulses allows for collection of the rephasing and non-rephasing signals. The inhomogeneity index (*I.I.*), which is related to the FFCF, is given by:

$$I.I. = \frac{A_r - A_n}{A_r + A_n}$$

where A_r and A_n are the absolute moduli of the rephasing and nonrephasing signals, respectively, integrated over the given vibrational band. It is possible to achieve some degree of excitation selectivity by tuning the excitation pulses and by choosing the coherence time delay to suppress the non-detected modes. With typical scanning speeds and the 1 kHz laser repetition rate, we obtain FFCF values with roughly 7 fs waiting time spacing. Since this time spacing is much smaller than our ~100 fs pulse width limited resolution, we apply a 100-fs windowed moving-average to the recorded FFCF results. All correlation functions were fit using a single exponential decay to the non-averaged data, after normalization. We average several waiting time scans to improve the signal-to-noise ratio for the FFCF.

3.2.5 Molecular Dynamics

Simulations of BTNN-BCT in the five experimental solvent conditions: $x_{DMF} = 0.07$, 0.11, 0.19, 0.41, and 1.00 were implemented using the GROningen Machine for Chemical Simulations (GROMACS[36]). For all simulations except $x_{DMF} = 1.00$, the SPC/E[37] water model was used. All simulations were run in a cubic box with at least 10 Å from all BTNN-BCT atoms for 20 ns each (see Appendix B for more detail).

AMBER99SB[38] all-atom force field parameters for DMF and BTNN-BCT were generated using the Antechamber[39] program from the AmberTools12 software package. Molecular structures for these molecules were first optimized in Gaussian 09[40] using the B3LYP functional and the 6-31G(d) basis set. Partial charges of the DMF and the organic backbone of the BTNN-BCT molecule were then assigned using the Restrained Electrostatic Potential (RESP) method. Force field modifications for the metal center and carbonyls were generated using LEaP.[41] A virtual atom was constrained to the center of the benzene ring and used as an anchor for the metal tri-carbonyl group, attaching it to the organic backbone and allowing for rotation of the carbonyls about the chromium – benzene bond.

Since biotin is so sparingly soluble in water, it is tempting to view the solute as being essentially hydrophobic. Many computational investigations of biomolecule hydration dissect the potentially distinct influences of chemical interactions in contrast to the geometrical nature of the constraints placed on water by the extended surface.[42-45] For example, there is a well known de-wetting transition predicted for model hydrophobes as a function of solute size, producing a vapor-like interface due to the expulsion of water near the extended surface. The extent of water vacancy, however, is greatly diminished when attractive interactions are included to mimic, for example, hydrogen bonding and van der Waals attraction.[46] In the biotin case, clearly the whole molecule is not hydrophobic since the ring system is anchored into binding proteins by strong hydrogen bonds, and water can form hydrogen bonds to stabilize the solvation of that fragment. The alkyl chain, on the other hand, is likely the main source of biotin's low solubility, suggesting a heterogeneous hydrophobicity even in a relatively small biomolecule.

To test the effect of chemical interaction as opposed to ideal hydrophobicity,[46] we simulated a model of biotin that lacks partial charges on all of the atoms except the Cr and three carbonyl ligands. All simulations were carried identically for the "hydrophobe" model and for the solute with realistic partial charges determined using RESP.

3.2.6 Radial Distribution Functions

Radial distribution functions (RDF) were calculated from the 20 ns production simulations for all metal-carbonyl oxygen DMF-nitrogen pairs. For each RDF the first and second solvation shells were fit to separate Gaussian functions to obtain the maximum shell density of DMF, Max{g(r)}.

3.3 Results and Discussion

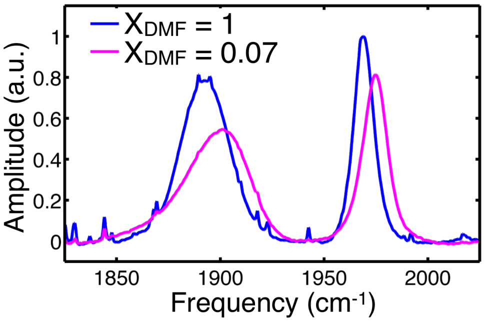

Figure 3.2. FTIR spectra of metal carbonyl region for 2.5 mM BTNN-BCT in pure DMF (blue) and x_{DMF} = 0.07 DMF (magenta), normalized to high frequency peak volume. The data show a 6 cm-1 blue shift of the high frequency mode when solvated in x_{DMF} = 0.07 compared to pure DMF. This blue shift at lower x_{DMF} highlights the increased nonpolar environment of the probe, suggesting preferential solvation by the DMF methyl groups.

3.3.1 Solvent-Dependent FTIR Spectra

The solvatochromic properties of BCT have previously been described[15] and follow a trend typical of metal carbonyls, exhibiting red shifted carbonyl frequencies in increasingly polar solvents. Assuming simple solvation of BTNN-BCT in the DMF ($E_N^T = 0.386$) and D_2O ($E_N^T = 0.991$)[47] mixtures, the spectrum of BTNN-BCT would be expected to red shift with decreasing x_{DMF}. However comparing the linear absorption spectra for 2.5 mM BTNN-BCT in x_{DMF} = 1 and x_{DMF} = 0.07, there is a 6 cm^{-1} blue shift for the symmetric mode upon the addition of D_2O (Figure 3.2), suggesting a decrease in local polarity around the carbonyls with increased bulk polarity. This decrease can be attributed to the preferential solvation of BTNN-BCT by DMF. Similar results have been seen for amphiphilic fluorophores in DMF-H_2O mixtures and readily highlight the complexity of solvation in mixed solvents.

3.3.2 Spectral Diffusion Dynamics

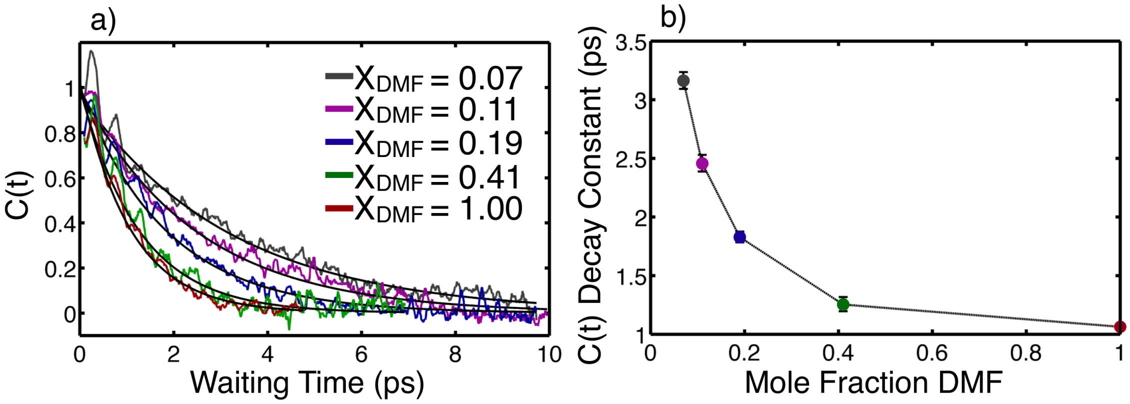

Figure 3.3. a) 100 fs window averaged measured correlation functions for x_{DMF} = 1 (grey), 0.41 (magenta), 0.19 (blue), 0.11 (green), and 0.07 (red) plotted vs. Waiting Time, exponential fits are shown as black solid lines. Spectra were normalized at T_W=0 and scaled to decay to 0 at T_W=∞. b) Correlation lifetimes from exponential fits with error bars plotted vs. Mole Fraction DMF, colors are the same as in a.

The FFCF is a powerful measurement, allowing for quantification of the timescales for a probe to sample the microstates of its system. FFCF measurements have provided valuable insight into the dynamics of water, water-lipid interfaces, inclusion complexes, and many other systems. Normalized FFCF measurements of the symmetric carbonyl mode and corresponding exponential fits are shown in Figure 3.3a for x_{DMF} = 1, 0.41, 0.19, 0.11, and 0.07. Figure 3.3b shows the decay constants obtained from the exponential fits as a function of x_{DMF}. The decays exhibit a highly nonlinear three-fold increase in correlation lifetime in going from x_{DMF} = 1 to x_{DMF} = 0.07. The magnitude of this slowdown is comparable to that seen in multiple model hydration systems previously studied by our group and others. Relative to biomacromolecules which induce a modest slowdown of hydration dynamics, were it possible to study BTNN-BCT in neat water, biotin's relatively small size would not be expected to induce dramatic changes in the hydration dynamics. Thus in analogy with previous studies of binary mixtures of polar solvents, we would expect a smooth variation of solvation time scales between ~1 ps for neat DMF and ~1.5 ps for neat D_2O. In the present case, however, we find a rapid, smooth, spectral diffusion slowdown, apparent for $x_{DMF} < 0.4$. Understanding the molecular nature of this dynamical slowdown is the primary goal of this investigation.

3.3.3 Comparison with Bulk Solution Properties

In order to understand the change in local dynamics of BTNN-BCT in the mixed solvents, we attempted to correlate bulk solution properties to the change in the local dynamics sensed by the carbonyl probes. For other metal carbonyl probes, we have previously observed spectral diffusion timescales to be strongly correlated with the solvent's bulk viscosity.[32] For BTNN-BCT in DMF-D_2O solutions (Figure 3.4a), however, we find no clear trend to the known viscocity.[48] Similar comparisons of spectral diffusion dynamics to empirically measured bulk properties have been carried out for metal-nitrosyls[49], where the acceptor number of the solvent was found to be negatively correlated to the FFCF decay constants. For BTNN-BCT in

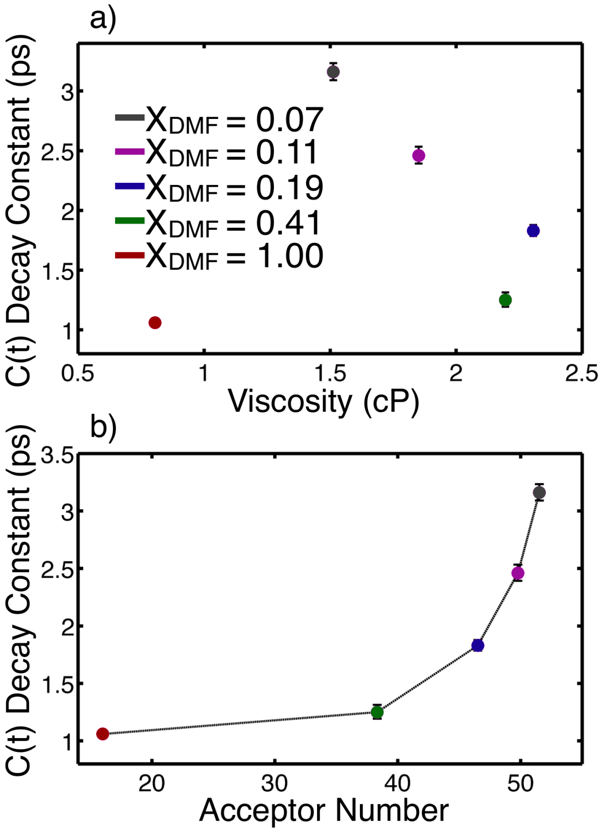

Figure 3.4. Correlation decay constant of BTNN-BCT plotted versus (a) bulk viscosity and (b) acceptor number, same coloring as (a). Error bars for decay constants are shown in both plots.

DMF-D_2O, we find a nonlinear dependence of the spectral diffusion time scale on the known acceptor numbers of the solutions[50] (Figure 3.4b). In contrast to the nitrosyl case, we find that the spectral diffusion is slower with increased bulk solvent acceptor number. The lack of any simple trend in the dynamics, coupled with the blue shifted vibrational bands, suggests that bulk properties will not be applicable in understanding the complex ternary solute-solvent-cosolvent structure and dynamics. Thus, we turn to all-atom molecular dynamics simulations as a way to gain insight into the microscopic behavior.

Mixed-solvent molecular dynamics simulations were run for the system at varying DMF/water compositions. To control for the hydrophobic solvation effects arising from the solvent mixture and to gain further insight into the nature of the BTNN-BCT solvation, complementary simulations were run with and without the RESP charges on the BTNN molecule. In order to preserve the solvent structuring around the carbonyl probes, however, charges were retained for the metal and carbonyl atoms. The simulations enable a direct comparison between solvating a realistic molecule and an ideal hydrophobic model of BTNN-BCT. By simulating this model system we are able to control for the solvation properties resulting from specific interactions between the solute and the solvent, gaining insight into the nature of the DMF-water solution.

3.3.4 Radial Distribution Functions: Preferential Solvation

Since preferential solvation is not unusual in mixed solvent ternary systems, we quantify the solvent structural information provided by the simulations by calculating partial radial distribution functions (RDFs). These allow for the direct comparison of the DMF density (relative to the bulk solution) within specific solvent shells of the metal carbonyls.

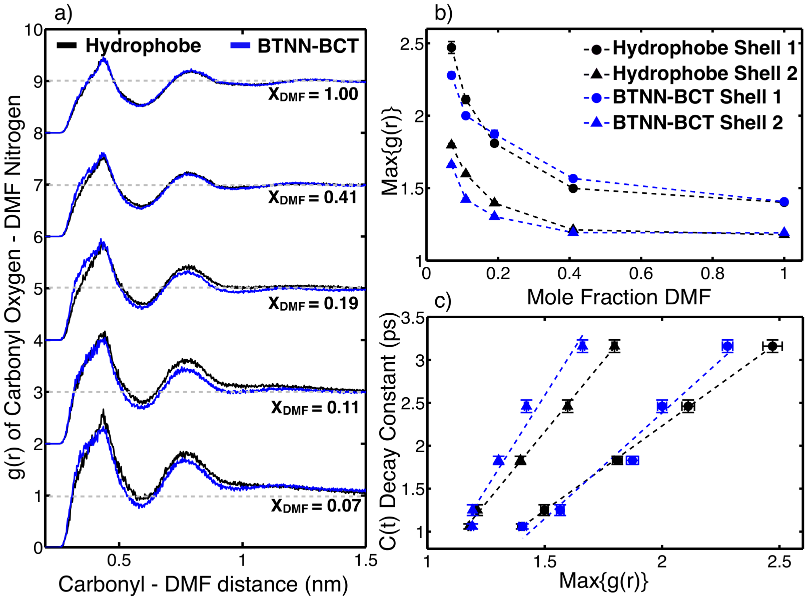

Figure 3.5. a) Radial distribution functions (RDFs) measuring the density of DMF near the carbonyl atoms of the probe were calculated for both series of simulations. RDFs shown for model hydrophobe (black) and BTNN-BCT (blue) for all simulated values of x_{DMF}. At low values of x_{DMF} the model hydrophobe shows a slightly higher density of DMF near the carbonyls than the BTNN-BCT system, these differences are absent in pure DMF. b) Maximum relative density of DMF obtained by fitting the data in (a) to two Gaussian functions, error bars are shown for 95% confidence bounds of fits. The first and second solvation shells (circles and triangles respectively) show similar trends for both the model hydrophobe (black) and BTNN-BCT (blue). c) Comparing the maximum DMF densities to the dynamics observed from the probe, Max{g(r)} is found to correlate strongly with the decay rate for C(t), dashed lines show least squares linear fit to data (same coloring as b).

Comparing the O_{CO}–N_{DMF} RDFs for the model hydrophobe and BTNN-BCT (Figure 3.5a), we find that the main features of the solvation environment are conserved and thus do not depend on the charges on the rest of the molecule. At all DMF concentrations there is an increased density of DMF near the carbonyls relative to the bulk. This preferential solvation is seen to increase with lower mole fractions of DMF, with $x_{DMF} = 0.07$ showing the largest relative DMF density near the solute. At low mole fractions of DMF the model hydrophobe shows increased maximum DMF density in both solvation shells compared to BTNN-BCT (i.e. with full partial charges) though this distinction is absent in pure DMF. The similarity observed between these two models and their RDFs is in contrast to the significant difference in the structural conformations between these two systems (see Appendix B for a detailed analysis of the concentration-dependent conformational changes). The model hydrophobe is found to exist in a predominantly collapsed state at low x_{DMF}, opening up as DMF is added to the system. The BTNN-BCT molecule also samples a larger conformational space as x_{DMF} increases, however the fully collapsed state is not observed at low x_{DMF}.

For both simulated systems the maximum relative density of DMF in the solvent shells was found to correlate strongly with the experimentally observed FFCF decay constant. The model hydrophobe system shows a slightly higher correlation between these parameters than the system with full partial charges (Figure 3.5c). The high correlation found between the degree of preferential solvation and observed FFCF decays highlights the link between the degree of solvent structural ordering around the carbonyl probe and the solvent's dynamical motions.

3.3.5 Preferential Solvation Can Explain the Composition-Dependent Spectral Dynamics

The simulations show clear evidence for preferential solvation of the carbonyls by the DMF solvent molecules, which is also the intuitive result based on the generally poor solubility of transition metal carbonyl complexes in water. Although we were not able to measure the spectral diffusion of the BTNN-BCT in 100% D_2O due to the low solubility, we can compare with our previous report of benzoate chromium tricarbonyl in D_2O, where we found 1.4 ps spectral diffusion.[15] This value agrees quantitatively with 2D-IR measurements of D_2O directly using dilute HOD.[27-29] Thus we conclude based on the present determination of spectral diffusion in pure DMF, that the spectral diffusion dynamics for the pure solvents takes place on the 1-1.5 ps timescale. If the solution were homogeneous, lacking any preferential solvation and concomitant influence on the dynamics, we might expect a monotonic variation from 1.5 to 1.0 ps as the mole fraction of DMF is varied from 0 to 1. Instead, we observe a rather sharp increase in the spectral diffusion time as we increase the water concentration. In analogy with previous solvation dynamics studies of binary solvents[51-57], we propose the following picture. Given preferential solvation by one component of a binary solvent mixture, where both solvents induce dynamics on a similar time scale, there is the possibility that a new, slower time scale emerges due to exchange of distinct solvent species near the solute. The FFCF can be modeled as the sum of three terms:

$$C(t) = w_{D_2O}e^{\left(-t/\tau_{D_2O}\right)} + w_{DMF}e^{\left(-t/\tau_{DMF}\right)} + w_{exch}e^{\left(-t/\tau_{exch}\right)}$$

where the weights of D_2O, DMF, and solvent exchange are given by w_{D_2O}, w_{DMF}, and w_{exch}, respectively. Since the expected bulk water spectral diffusion time scale and that measured for DMF are similar, we can simplify the correlation function to be a sum of two terms:

$$C(t) = w_{solvent}e^{\left(-t/\tau_{solvent}\right)} + w_{exch}e^{\left(-t/\tau_{exch}\right)}$$

where both pure solvents are captured in the $w_{solvent}$ weighted term. Exchange events switching occupancy of D_2O and DMF require mutual translational diffusion, which is expected to be slower than the largely orientational fluctuations that determine neat solution solvation dynamics.[51] For ideally homogeneous solutions one would expect the maximum contribution from exchange to occur near a mole fraction of 0.5, as has been recently reported.[58] With equal numbers of both solvents, fluctuations in solvent shell composition would result in more events that exchange the two solvent species. In contrast to this homogeneous solvation limit, our data shows that the slowest spectral diffusion occurs with the smallest DMF concentration. This finding is consistent with preferential solvation by DMF, which is most pronounced in the solution with the highest water concentration. For a preferentially solvated solute one would expect the greatest contribution from solvent exchange to occur at the smallest mole fraction of the preferred solvent. At low concentrations of the preferred solvent the system has the greatest opportunity to exchange cosolvent molecules near the solute. Figure 3.6 shows a cartoon representation of three solution compositions for a scenario with preferential solvation.

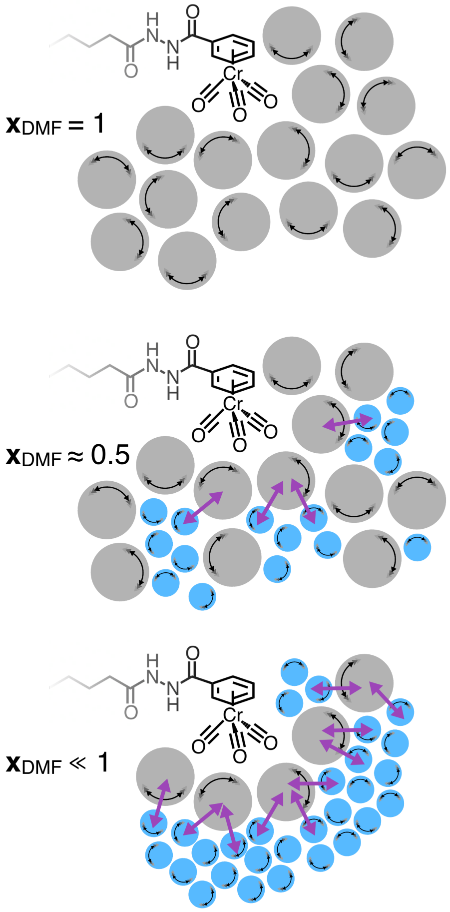

Figure 3.6. The solvent exchange model is able to describe the observed dynamical slowdown caused by preferential solvation of the probe. In pure solvent the fast orientational and librational motions of the solvent cause the spectral diffusion to decay rapidly. As cosolvent is added and the mole fraction of the preferred solvent is lowered, the frequency of exchange events increases, contributing a slower dynamical component to the spectral diffusion. At low concentrations of the preferred solvent these exchange events can contribute to a slowdown of the spectral diffusion of the probe.

Based on the picture that the exchange contribution should be most important at low DMF concentration, we can further refine the model for the correlation function decay. A simple model that incorporates the concentration dependence is

$$C(t) = x_{DMF}e^{\left(-t/\tau_{solvent}\right)} + \frac{1}{\beta}(1 - x_{DMF})\mathrm{e}^{\left(-t/\tau_{exch}\right)}$$

where β is a concentration-*independent* parameter that scales the contribution of the exchange dynamics to the spectral diffusion. There are two justifications for this parameter. First, without a detailed electrostatic mapping of electric field to carbonyl frequency, there is no guarantee that the motional time scale will correspond directly to the spectral diffusion time scale. Second, the degree of preferential solvation must also be considered, since it is unlikely that the first solvation shell is exclusively occupied by DMF. Indeed, it has been shown previously that the solvent exchange contribution to solvation dynamics probed with dynamic Stokes shift methods is considerably smaller in magnitude than the rapid, libration-like motion within the solvation shell.[57] The decay of $C(t)$ in our model is clearly biexponential, however we do not fit biexponentials to our data. Nevertheless, it is possible to obtain adequate fits to $C(t)$ using a single exponential and an offset. As described above, we subtract the offsets from our RASD decays due to the difficulty of determining the offsets reliably in practice,[34-35] thus we believe that our fitting allows for accurate comparison between the model and experimental data.

To compare the predictions of the model, we computed $C(t)$s numerically and then fit them using single exponentials with constant offsets for numerous values of τ_{exch} and β, fixing the $\tau_{solvent}$ at the 1.06 ps value obtained with $x_{\mathrm{DMF}} = 1$. By summing the squares of the residuals between the experimental and modeled decay constants, we identified a range of parameters that are consistent with our data (Figure 3.7). The globally optimal solution corresponds to τ_{exch} = 7.9 ps and β = 6.9, but solutions within the pink region of Figure 3.7 are all consistent with our data. The value of β indicates that the solvent exchange contribution to the overall spectral diffusion is considerably less than the direct, rapid orientational/librational motion of the immediate solvation shell. If the value can be interpreted literally, then the exchange accounts for roughly 14% of the total spectral diffusion dynamics.

With the range of exchange time scales, we can link the model's spectral diffusion dynamics to the motion of DMF within water using the known mutual diffusion coefficient of DMF in water (1×10^{-9} m^2/s).[59] The root mean squared (RMS) displacement is therefore given by

$$\sqrt{\langle r^2 \rangle} = \sqrt{6Dt}$$

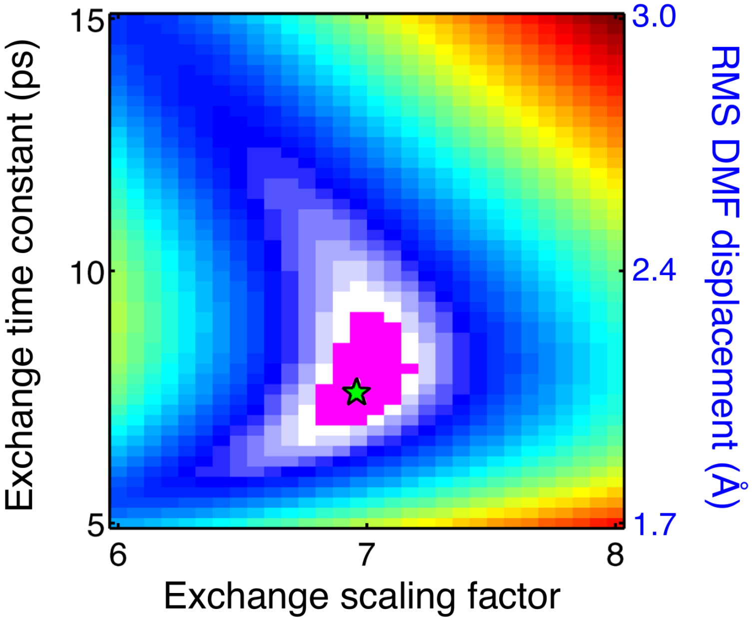

Figure 3.7. The sum of the squares between the model and experiment were calculated, scanning all values of the exchange time constant (τ_{exch}) and the exchange scaling factor (β). The data shows an area of high agreement with the data (pink region) with a global minimum at τ_{exch} = 7.9 ps and β=6.9. Considering the known diffusion constant of DMF in water the root mean square (RMS) displacement of a DMF molecule can be calculated and is found to agree well with the displacement of a single water molecule, ca. 2.8 Å.

from which we find that a time of 7.9 ps corresponds to an RMS displacement of 2.2 Å. Actually, the mutual diffusion coefficient is weakly concentration dependent,[59] but the variation is small, and due to the square root dependence of the RMS displacement, the variation is further suppressed. Indeed all of the good solutions to the model for the spectral diffusion are near or below 2.8 Å, which is the distance to the first maximum in the O-O radial distribution function of liquid water.[60] In other words, the model is consistent with the interpretation that the slowdown in spectral diffusion at low DMF concentration reports the time scale for the exchange of a single pair of DMF and water molecules. The use of the model highlights, however, that one cannot simply read the values of the spectral diffusion directly, since there are two dynamically distinct contributions which are separated in time by roughly an order of magnitude, and the slower contribution is weighted much less than is the faster process.

3.4 Conclusions

Here we show a detailed view of the dynamical effects of preferential solvation in a DMF/water solution. With the 2D-IR method RASD we are able to experimentally determine the spectral diffusion timescale of a small amphiphilic probe molecule in a variety of solvent environments. The data show a pronounced solvent dependence of the spectral diffusion timescale on the solvent composition. It is found that increasing concentrations of water induce a dynamical slowdown, despite the fact that both water and DMF alone produce spectral diffusion on the 1-1.5 ps time scale. With the aid of all-atom molecular dynamics simulations, this slowdown is found to correlate strongly with the degree of preferential solvation of the solute by DMF, as measured with partial RDFs. Remarkably the RDFs between the solute and DMF are found not to depend on the charges or conformation of the solute itself. The lack of correlation between the conformation of the solute and the solvation environment suggests that our probe can be thought as minimally perturbing to the solvent, below the Lum-Chandler-Weeks transition length scale of ~1 nm.[61]

The dynamical slowdown caused by preferential solvation is modeled as the introduction of a slow exchange mechanism between the preferred solvent and the bulk. This exchange model is found to agree well with the experimental data and the optimal 7.9 ps timescale for exchange dynamics is in excellent agreement with the timescale for a DMF molecule to diffuse the distance of one water molecule. Our findings complement the extensive body of work investigating binary solution solvation dynamics using dynamic fluorescence Stokes shift methods, highlighting the sensitivity of nonlinear infrared echo experiments to probe the structural dynamics of cosolvent exchange involving polar molecules.

By demonstrating the use of a biotin-linked metal carbonyl to monitor solvent shell exchange by 2D-IR, this system is well poised to offer site-specific insight into the dynamical solvent environment near proteins, within cells, and other engineered systems. This work provides an understanding on how dynamical effects arise from a changing environment, which may offer valuable insight into how solvent dynamics can be controlled and leveraged in the design of artificial enzymes.

3.5 Acknowledgements

This work along with its corresponding appendix (Appendix B) was submitted for publication to the Journal of Physical Chemistry B in March 2015.

3.6 References:

1. Rasaiah, J. C.; Garde, S.; Hummer, G., Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **2008,** *59*, 713-740.

2. Pratt, L. R., Molecular theory of hydrophobic effects: "She is too mean to have her name repeated.". *Annu. Rev. Phys. Chem.* **2002,** *53*, 409-436.

3. Tanford, C., How protein chemists learned about the hydrophobic factor. *Protein Sci.* **1997,** *6* (6), 1358-1366.

4. Ball, P., Water as an active constituent in cell biology. *Chem. Rev.* **2008,** *108* (1), 74-108.

5. Otting, G.; Liepinsh, E.; Wuthrich, K., Protein Hydration in Aqueous-Solution. *Science* **1991,** *254* (5034), 974-980.

6. Markelz, A.; Roitberg, A.; Heilweil, E., Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz. *Chem. Phys. Lett.* **2000,** *320* (1-2), 42-48.

7. Bagchi, B., Water dynamics in the hydration layer around proteins and micelles. *Chem. Rev.* **2005,** *105* (9), 3197-3219.

8. Krone, M. G.; Hua, L.; Soto, P.; Zhou, R. H.; Berne, B. J.; Shea, J. E., Role of water in mediating the assembly of Alzheimer amyloid-beta a beta 16-22 protofilaments. *J. Am. Chem. Soc.* **2008,** *130* (33), 11066-11072.

9. Cheung, M. S.; Garcia, A. E.; Onuchic, J. N., Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (2), 685-690.

10. Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G., Water in protein structure prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (10), 3352-3357.

11. Sterpone, F.; Stirnemann, G.; Laage, D., Magnitude and molecular origin of water slowdown next to a protein. *J. Am. Chem. Soc.* **2012,** *134* (9), 4116-9.

12. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B* **2012,** *116* (19), 5604-5611.

13. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Crowding Induced Collective Hydration of Biological Macromolecules over Extended Distances. *J. Am. Chem. Soc.* **2014,** *136* (1), 188-194.

14. King, J. T.; Kubarych, K. J., Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy. *J. Am. Chem. Soc.* **2012,** *134* (45), 18705-12.

15. Osborne, D. G.; Dunbar, J. A.; Lapping, J. G.; White, A. M.; Kubarych, K. J., Site-Specific Measurements of Lipid Membrane Interfacial Water Dynamics with Multidimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2013,** *117* (49), 15407-15414.

16. Pratt, L. R.; Pohorille, A., Hydrophobic effects and modeling of biophysical aqueous solution interfaces. *Chem. Rev.* **2002,** *102* (8), 2671-2691.

17. Bruice, T. C.; Hegarty, A. F., Biotin-Bound Co2 and Mechanism of Enzymatic Carboxylation Reactions. *Proc. Natl. Acad. Sci. U. S. A.* **1970,** *65* (4), 805-809.

18. Green, N. M., Avidin and Streptavidin. *Methods Enzymol.* **1990,** *184*, 51-67.

19. Wilchek, M.; Bayer, E. A., The avidin-biotin complex in bioanalytical applications. *Anal. Biochem.* **1988,** *171* (1), 1-32.

20. Hyster, T. K.; Knorr, L.; Ward, T. R.; Rovis, T., Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C-H Activation. *Science* **2012,** *338* (6106), 500-503.

21. Kohler, V.; Mao, J. C.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knorr, L.; Creus, M.; Prost, J. C.; Schirmer, T.; Ward, T. R., OsO4 center dot Streptavidin: A Tunable Hybrid Catalyst for the Enantioselective cis-Dihydroxylation of Olefins. *Angewandte Chemie-International Edition* **2011,** *50* (46), 10863-10866.

22. Quinto, T.; Schwizer, F.; Zimbron, J. M.; Morina, A.; Kohler, V.; Ward, T. R., Expanding the Chemical Diversity in Artificial Imine Reductases Based on the Biotin- Streptavidin Technology. *Chemcatchem* **2014,** *6* (4), 1010-1014.

23. Sardo, A.; Wohlschlager, T.; Lo, C.; Zoller, H.; Ward, T. R.; Creus, M., Burkavidin: A novel secreted biotin-binding protein from the human pathogen Burkholderia pseudomallei. *Protein Expression Purif.* **2011,** *77* (2), 131-139.

24. Chemistry, R. S. o., Biotin. Merck Sharp & Dohme Corp.: 2013.

25. Lei, Y.; Li, H. R.; Zhang, R.; Han, S. J., Molecular dynamics simulations of biotin in aqueous solution. *J. Phys. Chem. B* **2004,** *108* (28), 10131-10137.

26. King, J. T.; Ross, M. R.; Kubarych, K. J., Water-Assisted Vibrational Relaxation of a Metal Carbonyl Complex Studied with Ultrafast 2D-IR. *J. Phys. Chem. B* **2012,** *116* (12), 3754-3759.

27. Loparo, J.; Roberts, S.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. I. Vibrational dynamics in two-dimensional IR line shapes. *J. Chem. Phys.* **2006,** *125* (19), 194521.

28. Loparo, J.; Roberts, S.; Tokmakoff, A., Multidimensional infrared spectroscopy of water. II. Hydrogen bond switching dynamics. *J. Chem. Phys.* **2006,** *125* (19), 194522.

29. Asbury, J.; Steinel, T.; Kwak, K.; Corcelli, S.; Lawrence, C.; Skinner, J.; Fayer, M., Dynamics of water probed with vibrational echo correlation spectroscopy. *J. Chem. Phys.* **2004,** *121*, 12431-12446.

30. Cowan, M.; Bruner, B.; Huse, N.; Dwyer, J.; Chugh, B.; Nibbering, E.; Elsaesser, T.; Miller, R., Ultrafast memory loss and energy redistribution in the hydrogen bond network of liquid H2O. *NATURE* **2005,** *434* (7030), 199-202.

31. Zhang, Z.; Piatkowski, L.; Bakker, H. J.; Bonn, M., Ultrafast vibrational energy transfer at the water/air interface revealed by two-dimensional surface vibrational spectroscopy. *Nat. Chem.* **2011,** *3* (11), 888-893.

32. King, J. T.; Anna, J. M.; Kubarych, K. J., Solvent-hindered intramolecular vibrational redistribution. *Phys. Chem. Chem. Phys.* **2011,** *13* (13), 5579-5583.

33. Fayer, M. D., Dynamics of Water Interacting with Interfaces, Molecules, and Ions. *Acc. Chem. Res.* **2012,** *45* (1), 3-14.

34. Osborne, D. G.; Kubarych, K. J., Rapid and accurate measurement of the frequency-frequency correlation function *J. Phys. Chem. A* **2013,** *117* (29), 5891-8.

35. Nilsen, I. A.; Osborne, D. G.; White, A. M.; Anna, J. M.; Kubarych, K. J., Monitoring equilibrium reaction dynamics of a nearly barrierless molecular rotor using ultrafast vibrational echoes. *J. Chem. Phys.* **2014,** *141* (13).

36. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

37. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

38. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein

backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006,** *65* (3), 712-725.

39. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006,** *25* (2), 247-260.

40. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

41. Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V., LEAP. University of California, San Francisco: 1995.

42. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005,** *437* (7059), 640-647.

43. Patel, A. J.; Garde, S., Efficient Method To Characterize the Context-Dependent Hydrophobicity of Proteins. *J. Phys. Chem. B* **2014,** *118* (6), 1564-1573.

44. Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S., Sitting at the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Function. *J. Phys. Chem. B* **2012,** *116* (8), 2498-2503.

45. Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Acharya, H.; Garde, S.; Chandler, D., Extended surfaces modulate hydrophobic interactions of neighboring solutes. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (43), 17678-17683.

46. Willard, A. P.; Chandler, D., The molecular structure of the interface between water and a hydrophobic substrate is liquid-vapor like. *J. Chem. Phys.* **2014,** *141* (18), 5.

47. Reichardt, C., *Solvents and Solvent Effects in Organic Chemistry 3rd ed.* Wiley-VCH: 2003.

48. Garcia, B.; Alcalde, R.; Leal, J. M.; Matos, J. S., Solute-solvent interactions in amide-water mixed solvents. *J. Phys. Chem. B* **1997,** *101* (40), 7991-7997.

49. Brookes, J. F.; Slenkamp, K. M.; Lynch, M. S.; Khalil, M., Effect of Solvent Polarity on the Vibrational Dephasing Dynamics of the Nitrosyl Stretch in an Fe-II Complex Revealed by 2D IR Spectroscopy. *J. Phys. Chem. A* **2013,** *117* (29), 6234-6243.

50. Mayer, U.; Gerger, W.; Gutmann, V., Nmr-Spectroscopic Studies on Solvent Electrophilic Properties .2. Binary Aqueous - Non-Aqueous Solvent Systems. *Monatsh. Chem.* **1977,** *108* (2), 489-498.

51. Luther, B. M.; Kimmel, J. R.; Levinger, N. E., Dynamics of polar solvation in acetonitrile-benzene binary mixtures: Role of dipolar and quadrupolar contributions to solvation. *J. Chem. Phys.* **2002,** *116* (8), 3370-3377.

52. Chakrabarty, D.; Chakraborty, A.; Seth, D.; Sarkar, N., Effect of water, methanol, and acetonitrile on solvent relaxation and rotational relaxation of coumarin 153 in neat 1-hexyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. A* **2005,** *109* (9), 1764-1769.

53. Mukherjee, S.; Sahu, K.; Roy, D.; Mondal, S. K.; Bhattacharyya, K., Solvation dynamics of 4-aminophthalimide in dioxane-water mixture. *Chem. Phys. Lett.* **2004,** *384* (1-3), 128-133.

54. Molotsky, T.; Huppert, D., Solvation statics and dynamics of coumarin 153 in dioxane-water solvent mixtures. *J. Phys. Chem. A* **2003,** *107* (41), 8449-8457.

55. Nguyen, C. N.; Stratt, R. M., Preferential solvation dynamics in liquids: How geodesic pathways through the potential energy landscape reveal mechanistic details about solute relaxation in liquids. *J. Chem. Phys.* **2010,** *133* (12).

56. Benjamin, I., Static and dynamic electronic spectroscopy at liquid interfaces. *Chem. Rev.* **2006,** *106* (4), 1212-1233.

57. Agmon, N., The dynamics of preferential solvation. *J. Phys. Chem. A* **2002,** *106* (32), 7256-7260.

58. Gupta, S.; Rafiq, S.; Sen, P., Dynamics of Solvent Response in Methanol-Chloroform Binary Solvent Mixture: A Case of Synergistic Solvation. *J. Phys. Chem. B* **2015**.

59. Biswas, S.; Mallik, B. S., Effects of Temperature on the Structure and Dynamics of Aqueous Mixtures of N,N-Dimethylformamide. *J. Chem. Eng. Data* **2014,** *59* (10), 3250-3257.

60. Ferguson, D. M., Parameterization and Evaluation of a Flexible Water Model. *J. Comput. Chem.* **1995,** *16* (4), 501-511.

61. Lum, K.; Chandler, D.; Weeks, J. D., Hydrophobicity at small and large length scales. *J. Phys. Chem. B* **1999,** *103* (22), 4570-4577.

Chapter Four
Dynamical Effects of Point Mutations in an Engineered Protein Heterogeneous Catalyst

4.1 Introduction

The ability to recombinantly express proteins has offered a myriad of insight into the functioning of these highly complex and tailored molecular machines. Built from a polymeric structure of the twenty natural amino acids, even short proteins can quickly exceed the samplable space for sequence mutations in the lab.[1] Regardless of this formidable conformational space, protein engineering through design and directed mutations remains a highly active field of the biosciences.[2-9] To aid the development of protein engineering there has been a push in the biochemical field towards developing predictive computational models[10-17], based on the known structures of many biological proteins. These tools have led to great successes in the development of *de novo* enzymes[7, 18-20] however there remains much to be desired in their predictive ability.

A main limitation of many of these models is in their representation of the protein solely on structure without respect to any of its conformational dynamics. This is due in part to the difficulty in measuring protein dynamics experimentally and incorporating these parameters into models. Furthermore the vast libraries of structural information provided by protein crystal structures dwarfs the amount of dynamical information available on proteins. Without dynamical considerations designed enzyme scaffolds are engineered to stabilize a transition state of the desired reaction.[21] This method subsequently relies on time-consuming sequence mutation of the protein backbone to modulate the dynamical binding and releasing rates of the enzyme.[22-23] To aid in the directed mutation of these enzymes knowledge of how different amino acids affect their local dynamics is needed. This work utilizes a model engineered protein system[24-26], based on the streptavidin-biotin interaction[27], in combination with a site-specific probe of conformational and solvent dynamics to gain insight into the ultrafast dynamical effects of point mutations in the enzyme scaffold.

Streptavidin and biotin are commonly utilized for their ability to act as molecular Velcro in a variety of environmental conditions.[28-31] The high degree of promiscuity offered by streptavidin in binding biotin analogs has enabled researchers to incorporate organometallic enzymes into the binding site of the protein by attaching them to the biotin tail. This system offers the engineered enzymes a structured and chiral environment allowing for chiral catalysis.[26] Mutations of the protein scaffold in the vicinity of the catalytic site are able to tune the reactivity and enantiomeric selectivity of the catalyst providing many routes for optimization of the enzyme.[24] Utilizing a vibrational molecular probe and 2D-IR we aim to characterize how these mutations affect the conformational dynamics of the protein and surrounding environment. The observed ultrafast dynamics will be compared to molecular dynamics simulations to gain insight into the molecular nature of observed dynamical changes.

2D-IR has repeatedly proven its versatility in measuring the ultrafast dynamics of biological dynamics through the correlation of the inherent frequency inhomogeneities.[32-35] These inhomogeneities arise from the multiple microstates of the system surrounding the vibrational probe. By correlating the time scales for the interchange between these microstates, our group has been able to quantify the degree of water slowdown at the interface if lipid membranes[33] as well as the extent of environmental slowdown as a consequence of preferential solvation of a small amphiphile. 2D-IR allows for the direct measurement of the frequency fluctuation correlation function by temporally measuring the correlation loss between the excitation and detection frequencies for the probe molecules, a process called spectral diffusion.[36-37] If the correlation is measured at an early waiting time, where we have not allowed the microstates of the probe ample time to sample alternate conformations, the spectra will show a high degree of correlation between the excitation and detection frequencies. This correlation is lost as the waiting time is increased and the microstates are allowed more time to interconvert. Using a site-specific probe and molecular dynamics we are able to gain detailed information on the local dynamics of the protein and how they are modulated by single mutations of the peptide backbone by monitoring the spectral diffusion.

4.2 Experimental Methods

4.2.1 Materials

Dimethylformamide (DMF) and deuterium oxide (D_2O) were purchased from Sigma-Aldrich and used without further purification. Wild Type (WT) streptavidin and the two mutants, s112a and s112h, were provided by Professor Thomas Ward at the University of Basel. Biotin hydrazide benzyl chromium tricarbonyl (BTNN-BCT) was prepared as previously described.

4.2.2 Sample Preparation

Samples were prepared to a final concentration of 2.5 mM protein and BTNN BCT in a solution of 25% v/v DMF in D_2O ($x_{DMF} = 0.07$). This was accomplished by the drop-wise addition of 50 μL of 10 mM BTNN-BCT in DMF to a 150 μL solution of 3.3 mM protein dissolved in D_2O. The solution was gently aspirated, to avoid bubble formation, before being placed in a custom sample cell comprised of two 1 in by 3 mm calcium fluoride windows separated by a 100 μm spacer.

4.2.3 FTIR Measurements

FTIR Spectra were collected for the 2.5 mM BTNN-BCT in pure DMF in addition to the three protein-ligand systems studied. Spectra were baseline corrected by subtracting a 4th-order polynomial and normalized to the area of the high-frequency symmetric mode for comparison.

4.2.4 2D-IR Experiments

The details of the 2D-IR methodology and experimental setup have been previously described elsewhere.[38-40] 2D-IR allows for the direct monitoring of spectral diffusion by calculation of the inhomogeneity index ($I.\,I.$) from the absolute moduli of the rephasing (A_r) and nonrephasing (A_n) signals:

$$I.I. = \frac{A_r - A_n}{A_r + A_n}$$

integrating over the appropriate spectral range.[37] The *I. I.* is calculated at multiple waiting times for the probe. From this data we are able to extract the functional form of the FFCF by fitting the *I. I.* data to an exponential decay with a constant offset. The lifetime of the decay reports on the timescale of the microstate sampling of the probe.

4.2.5 Molecular Dynamics Simulations - Setup

Molecular dynamic simulations were run using the GROningen Machine for Chemical Simulation (GROMACS[41]) 4.5.5 software package and the AMBER99SB[42] force field. For all simulations the SPC/E[43] water model was used in addition to DMF for the solvent. The structure of WT homotetrameric streptavidin was obtained by generating symmetry mates in PyMOL[44] from the published 1STP.pdb[45] crystal structure. Starting structures for the s112a and s112h mutants were obtained using the mutagenesis wizard in PyMOL. BTNN-BCT was placed into the cavity of each monomer by aligning the energy-minimized structure of BTNN-BCT with the crystalized biotin position. The protein-ligand complex was placed in an equilibrated, cubic, solvent box with x_{DMF} = 0.07 to correspond to the experimental solvent environment. The box was constructed to allow for a minimum of 1 nm distance between any protein or ligand atom and the edge of the box.

The topology and coordinate files for the protein systems were automatically generated utilizing the *pdb2gmx* function of GROMACS and the pdb structure. BTNN-BCT was previously characterized for the AMBER99SB force field and used as previously described. The charges of the organic backbone of the BTNN-BCT molecule are removed for these simulations. As was shown in the previous chapter, the effect of the charges on the organic backbone minimally effected the solvation of the carbonyl probe. Due to the buried nature of the biotin group in the protein, the lack of formal charges on the organic backbone is expected to have minimal effect on the simulation results due to the short simulation time relative to the time scale of ligand removal. Eight sodium ions were added to each system to neutralize the system charge arising from the protein.

Simulations for each system were first energy minimized using a steepest decent algorithm for 100,000 steps with a step size of 1 fs. These minimized systems were then equilibrated in sequential Canonical (NVT) and isothermal-isobaric (NPT) simulations for 100 ps using step sizes of 0.5 fs and the leapfrog integrator. For these simulations the V-rescale and Parrinello-Rahman thermostat and barostat were used with lifetimes of 100 fs and 2 ps respectively, when appropriate. Production simulations were run for 5 ns with a step size of 1 fs and utilizing the same thermostat and barostat parameters as the isothermal-isobaric simulations. Structures from the simulations were saved every 1 ps of the simulation, generating 5000 structures for each system.

4.2.6 Solvent Sphere Analysis

Solvation spheres were defined around each chromium atom by indexing the molecules within 0.8, 1.1, and 1.8 nm radius of the chromium. This was accomplished by using Matlab to import the full 5 ns trajectories and extract the atoms within the cutoff radius of the chromium. The molecular identity of each extracted atom was determined from the structural information file and the whole molecule was included in reconstructing a reduced trajectory file containing the atoms of interest and their positions. From these reduced trajectories, the numbers of water and DMF molecules were determined, frame-by-frame, though indexing the structural information file with the reduced trajectory.

4.2.7 Spatial Correlation of Solvent Molecules

For each simulated system, the trajectories were processed to maintain a constant frame of reference with respect to the protein, removing the translational and rotation degrees of freedom from the system. The *g_spatial* utility of GROMACS was used to calculate the spatial correlations of the solvent in 0.1 nm x 0.1 nm x 0.1 nm bins for the 5 ns simulations. The correlation maps are overlaid with the initial WT streptavidin and BTNN-BCT structures to highlight the locations of solvent interaction.

4.3 Results and Discussion

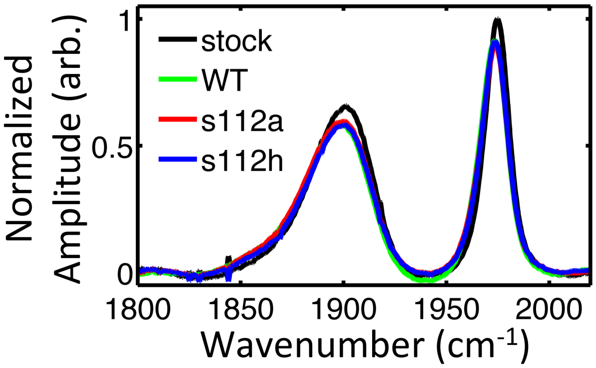

Figure 4.1 FTIR spectra of carbonyl frequencies for BTNN-BCT in pure DMF stock solution (black), WT (green), s112a (red), and s112h (blue) streptavidin systems. The spectra have been baseline corrected and are normalized to the area of the high-frequency symmetric mode. The spectra of the protein samples show a slight increase in peak width for both symmetric and asymmetric modes compared to the stock DMF sample. All three proteins have similar spectra indicating no large differences in electrostatic environments between the different mutants.

4.3.1 Linear IR Spectra

The ability to gain site-specific ultrafast dynamical information from a protein system provides insightful information on the effects of single residue mutations. For our streptavidin system, FTIR and 2D-IR spectra were acquired for the WT, s112a, and s112h mutant systems, utilizing the carbonyl frequencies of the BTNN-BCT probe. Readily apparent from the linear FTIR spectra, shown in Figure 4.1, is the similarity of the spectra in the three protein systems indicating a similar binding environment between the three systems. Compared to the stock DMF solution, the protein spectra show a slight increase in the peak width and a red-shifted frequency. As has been observed for BCT in other systems, this shift is consistent with a more polar solvation environment of the probe molecules bound to the protein. In order to gain dynamical information on these systems 2D-IR was utilized allowing for the investigations of the site-specific vibrational dynamics in each system.

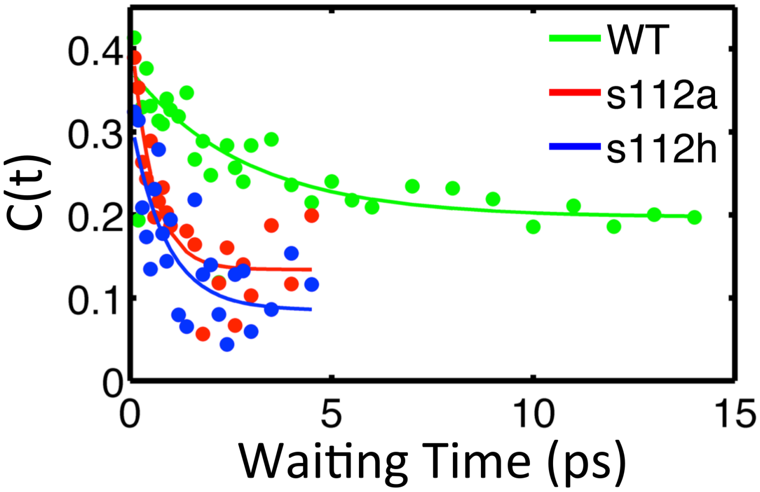

Figure 4.2 Frequency Fluctuation Correlation Function (C(t)) values plotted as a function of waiting time for each WT (green), s112a (red), and s112h (blue) protein system. For each system the experimental time points (circles) were fit to a single exponential with an offset (solid line) to extract the correlation lifetimes: τ_{WT} = 2.86 ps, τ_{s112a} = 0.53 ps, and τ_{s112h} = 0.86 ps. The data show a clear dynamical difference between WT streptavidin and the two mutants studied.

4.3.2 2D-IR Spectra

For each protein sample, 2D-IR spectra were acquired for 2.5 mM BTNN-BCT bound to streptavidin in D_2O with x_{DMF} = 0.07. DMF was utilized as a carrier solvent for the BTNN-BCT probe; see the previous chapter for a full characterization of the DMF-D_2O solvent dynamical effects on the probe molecule. The 2D-IR spectra allow for the quantification of the spectral diffusion dynamics through the FFCF in the different protein systems Figure 4.3. In the presence of protein we see a decrease in the spectral diffusion timescales from those observed in the previously reported x_{DMF} = 0.07system. This difference is the most extreme for the s112a and s112h mutants, which show a significantly faster relaxation (0.53 and 0.86 ps respectively) of their FFCFs compared to the WT protein (2.86 ps). Between protein systems we observe a drastic quickening in the loss of correlation for the two mutant proteins compared to WT streptavidin. Utilizing the same framework established in chapter 3 to understand the preferential solvation dynamics of our probe, we would expect that the faster lifetimes observed for the protein bound probe can be attributed to a loss of the preferential solvation by DMF of the unbound probe. The observed dynamical differences in *C(t)*, considering the similarity in the FTIR spectra, highlights the importance of dynamical information gained from 2D-IR.

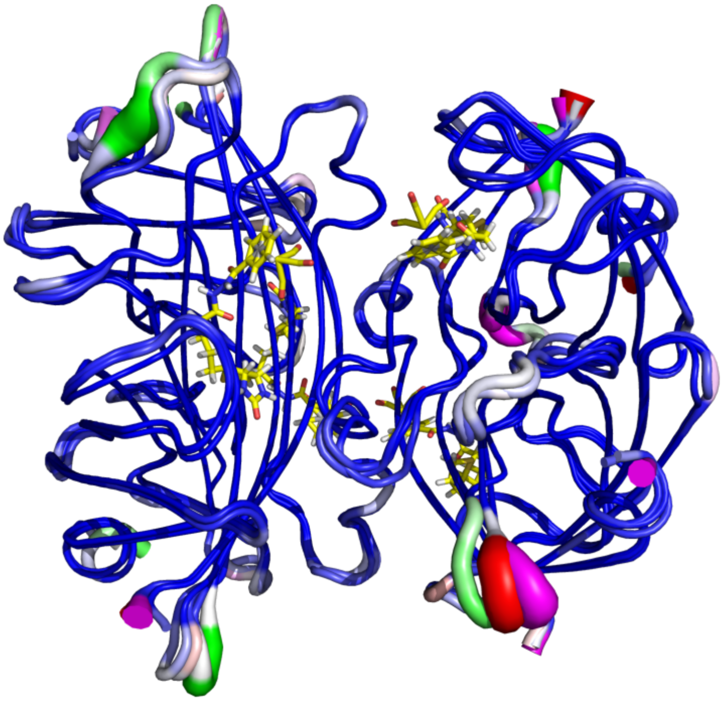

Figure 4.3 Visualization of the root mean square fluctuations (rmsf) about the mean positions of the protein backbone shown for each system, WT (green), s112a (magenta), and s112h (red). Gradient color and line thickness indicate degree of the rmsf value with blue-thin lines showing areas of rigid structure. An overall conservation of the rmsf values in observed in the system with the largest fluctuations occurring in the loop regions of the protein. Comparing the degree of the fluctuations and their locations to the BTNN-BCT positions (shown in yellow) we see that the three proteins show a high degree of similarity in the binding pocket fluctuations.

4.3.3 Molecular Dynamics Simulation of Protein Dynamics

To control for the effect of side chain mutations on the flexibility of the protein backbone, molecular dynamics simulation were utilized to monitor for destabilization of the protein structure. Over the course of the 5 ns simulations the root mean square fluctuations of the backbone were calculated. These results, shown in Figure 4.3, represent the regions of the protein backbone that undergo the largest fluctuations as thicker, colored, sections of the traced mean backbone structure. The data shows that the rigidity of the streptavidin binding pocket remains in the presence of point mutations of the backbone. As would be expected for a rigid structure, the maximal rmsf values are seen to be conserved to the loop regions of the scaffold and do not interact directly with the bound BTNN-BCT ligands. The conservation of the binding site rmsf values indicates an alternative mode of spectral diffusion relaxation than through purely protein dynamics near the binding site.

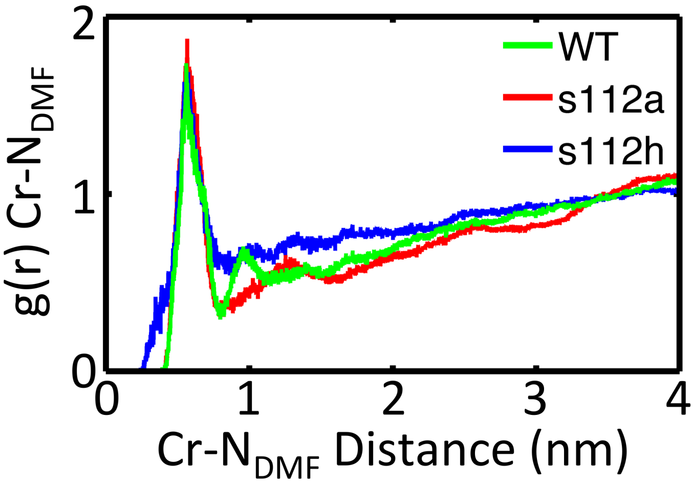

Figure 4.4 Partial Radial Distribution Function (RDFs) calculated between the BTNN-BCT chromium atoms and the DMF cosolvent nitrogens. The RDFs have been normalized to the bulk value of g(r). Present in all mutants is the observed preferential solvation by DMF seen by the peak near 0.6 nm. Besides the primary peak the RDFs for the s112a (red) and s112h (blue) mutants are featureless. A secondary DMF solvation shell is seen for the WT protein (green) at 1 nm.

4.3.4 Solvent Dynamics from Molecular Dynamics

In addition to the dynamical information of the protein provided by the simulations, we are also able to gain insight into the solvation environment around the protein and our probe. The partial radial distribution functions (RDFs) between the BTNN-BCT probes and DMF cosolvent were calculated for each mutant, Figure 4.4. From the RDFs, the DMF is found to preferentially associate with the probe and protein surface for all mutants. Unique to the WT protein system is the structuring of a second solvation shell of DMF emerging from the RDF at a separation distance of 1 nm. As was observed in the ternary BTNN-BCT, DMF, and D_2O system, preferential solvation can contribute a slow timescale to the observed spectral diffusion. This term arises from the exchange of DMF and water in the solvation environment of the probe and has been shown to lead to a slow-down in the observed spectral diffusion.

The spatial correlation of the DMF and water solvent molecules was also calculated separately for each protein mutant. These data highlight the spatial clustering of solvent molecules across multiple frames of the simulation. The heat maps utilize a color scale, with warmer colors representing a higher degree of clustering, to indicate regions of the system that experience a preferential interaction with a particular solvent. From this data, shown in Figure 4.5, two main features of the solvation environment can be observed. The first is the relative

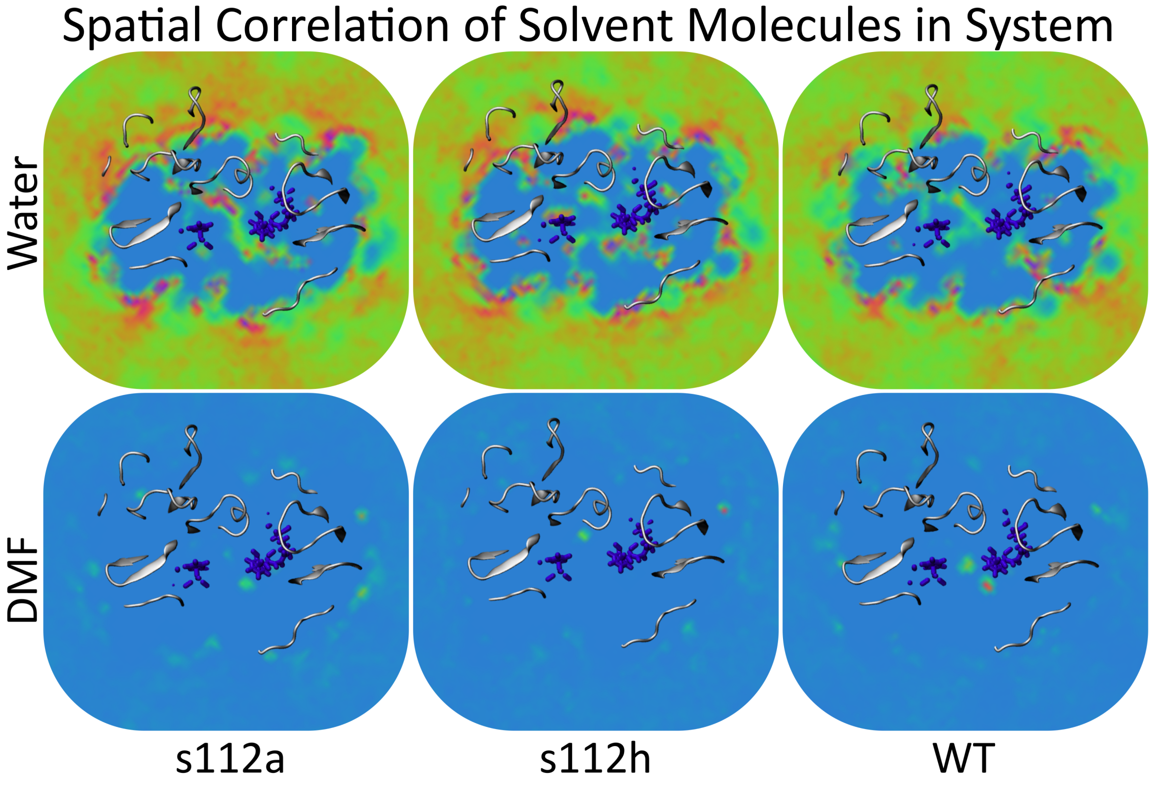

Figure 4.5 Calculated spatial correlations of the different solvents (rows) for each protein system studied (columns). Heat maps represent the degree of spatial correlation with warmer colors indicating a larger degree of correlation. The same heat map is utilized for all figures. To aid visualization the protein structure (grey) and BTNN-BCT (dark blue) have been overlaid with the maps. The data show a decrease in the water interaction near the BTNN-BCT probes for the WT protein system compared to the two mutants studied. The data also shows a relative increase in DMF correlation near the WT binding pocket. This data is consistent with the calculated RDFs however provides more detail on the solvation environment.

decrease of water solvation of the protein surface near the BTNN-BCT probe molecules in the WT protein compared to the two mutant systems. Secondly it can be seen by looking at the spatial correlation of the DMF that the WT experiences a higher degree of preferential interaction of the DMF near the BTNN-BCT probe. This latter trend is consistent with the measured RDFs presented above.

Due to the symmetry of streptavidin we can utilize the same analysis to investigate the solvation environments of the reverse side of the protein Figure 4.6, which contains the other biotin binding sites. The data for this reverse side, Figure 4.7, is consistent with the previously shown spatial correlation maps. A decrease of the water preferential interacting near the probe in the WT protein compared to the two mutants is seen. Additionally a slight increase in the preferential interaction of DMF near the probe is seen in the WT protein though this affect is seen to be smaller than for the other side of the protein.

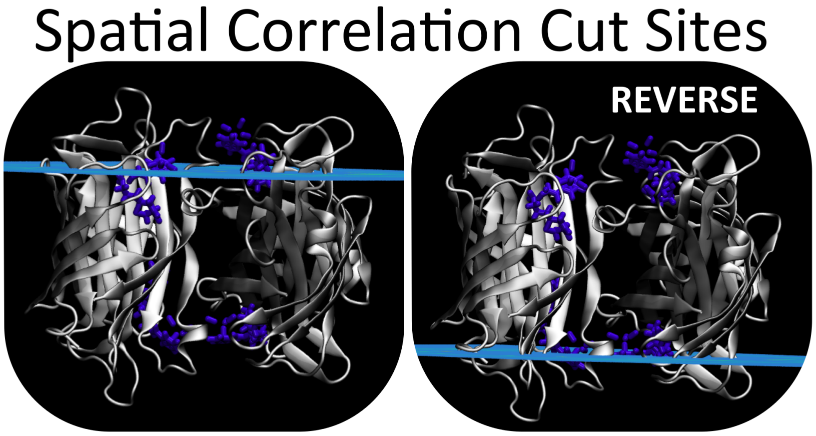

Figure 4.6 Locations of planes used for illustrating the spatial correlation of the solvent in Figure 4.5 (left) and Figure 4.7 (right).

The appearance of the preferential solvation, of the BTNN-BCT by DMF in the WT protein suggests that the solvent exchange dynamics arising from the heterogeneity of the solvation environment is responsible for the observed dynamical slowdown. This effect was previously reported for the solvation dynamics of the BTNN-BCT in a series of DMF – D_2O solutions where the preferred solvent, DFM, preferentially associates with the BTNN-BCT probe. This preferential

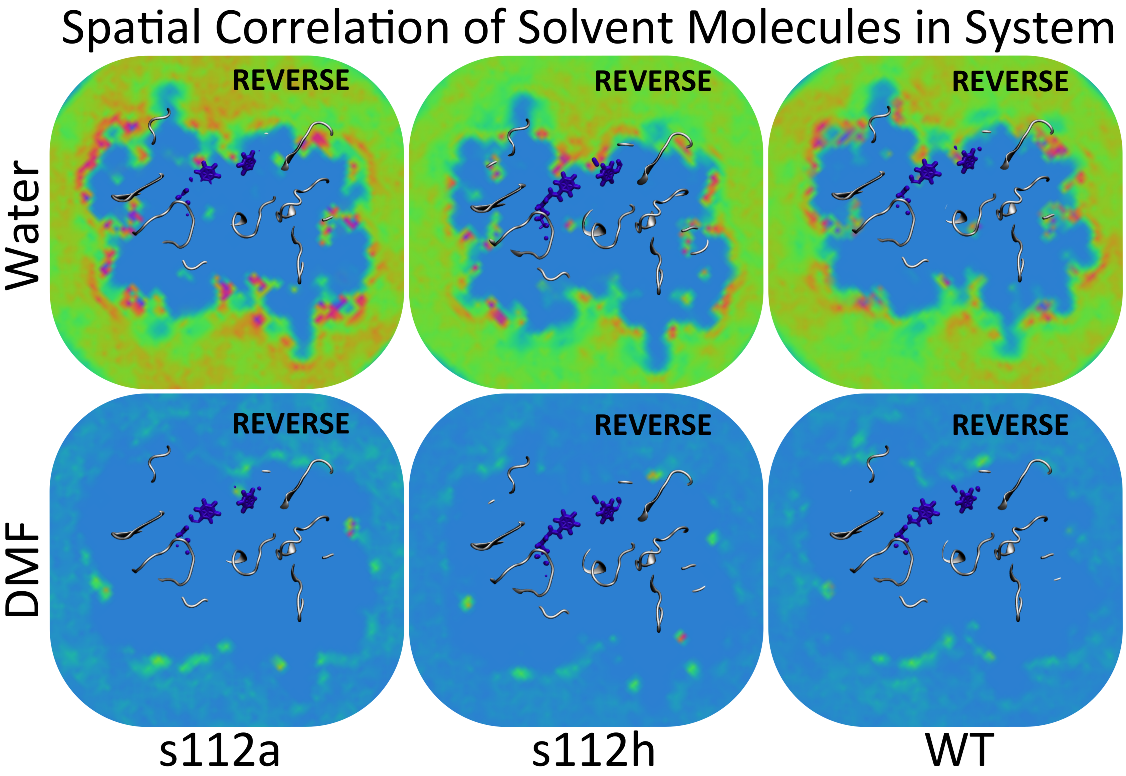

Figure 4.7 Calculated spatial correlations of the different solvents (rows) for each protein system studied (columns) for the reverse side of the protein. The same coloring and map is used as for Figure 4.5. Similar to what was seen on the other side of the protein, the data show a decrease in the water interaction near the BTNN-BCT probes for the WT protein system compared to the two mutants studied. The data also shows a relative increase in DMF correlation near the WT binding pocket, though this affect is seen to be lessened.

solvation leads to a larger contribution of solvent exchange dynamics as the concentration of DMF is decreased and the likelihood of DMF – D_2O exchange increases.

4.3.5 Characterization of Preferential Solvation

The preferential solvation of the BTNN-BCT in the protein can also be visualized by calculating the mole-fraction of DMF within shells of a fixed distance from the BTNN-BCT probe, as shown in **Figure 4.8**a. Looking at the mean x_{DMF} within 0.8, 1.1, and 1.8 nm spheres of each mutant we see that x_{DMF} is increased near the protein surface and tapers off to the bulk value, $x_{DMF} = 0.07$, at large distances. We can continue the analysis of the solvent shells by studying how the solvent within the shell exchanges with the surrounding solvent. From our simulations we are able to quantify the number of solvent molecules within a shell at each time point. By analyzing how the solvent occupancy of the shells fluctuates we are able to compute the solvent occupancy-fluctuation autocorrelation for the solvent shells in each mutant. These autocorrelation functions are subsequently normalized and integrated from 0 - 100 ps to extract the mean lifetime of the correlation. This data shown in **Figure 4.8**b provides insight into the solvent occupancy dynamics around the probe.

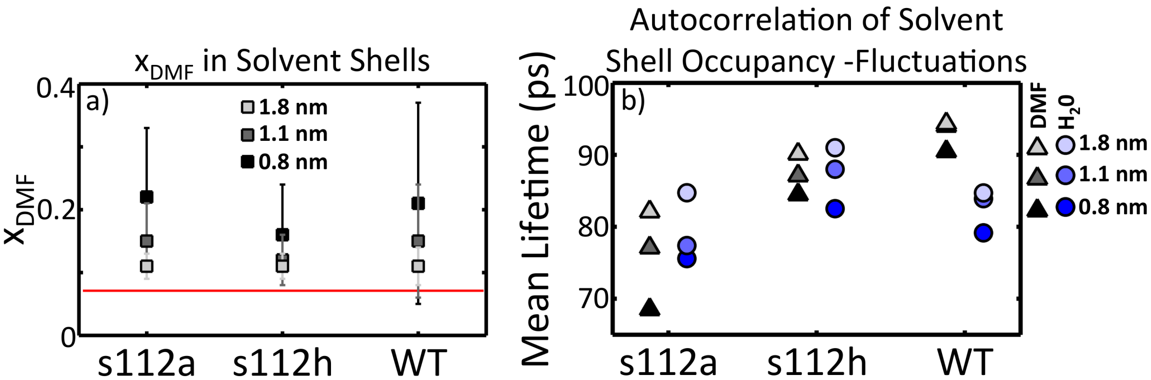

Figure 4.8 a) Calculated mean x_{DMF} values of the solvent within specified distances of chromium atom in BTNN-BCT. Near the probe DMF is found at a higher x_{DMF} than the bulk value (grey dashed line). As the spheres are increased in size the three systems are seen to converge to the bulk value (red line). b) Mean lifetimes from autocorrelation of solvent shell occupancy fluctuations for DMF (triangles) and water (spheres) for shell sizes of 0.8, 1.1, and 1.8 nm (dark to light colors). For the s112a and s112h the lifetimes of the DMF and water fluctuations are seen to coexist on the same time scale. For the WT protein the mean lifetime for the DMF and water seen to separate with the DMF showing slower occupancy fluctuation dynamics than the water.

For s112a and s112h, the water and DMF occupancy fluctuations, spheres and triangles respectively, are found to be comparable across shells of the same size. The similarity of mean lifetimes for these systems suggests a homogeneous solvation environment around the probe; absent of preferential solvation effects that would manifest as a disparity between the timescales of the solvent occupation fluctuations of the two solvents. For the WT protein we see a separation in timescales between the DMF and water occupancy fluctuations. This separation of time scales can be explained form the preferential association of the DMF to the environment around the probe. By preferentially solvating the probe and protein the DMF constrains its motion to near the protein, resulting in an increase in the occupancy fluctuation lifetime. The utilization of the mutants in our experiment offers powerful insight into the nature of this preferential solvation.

4.3.6 Structural Basis for Preferential Solvation

As the effects from preferential solvation are not observed by either mutant studied, the preferential solvation of DMF can be hypothesized to occur due to the presence of the WT s112 residue. Looking at the RDF of the hydroxyl oxygen of s112 and the carbonyl oxygen of DMF, Figure 4.9, we see a classic indication of hydrogen bonding from the peak at 2.7 Å. An additional peak is observed near 4.3 Å corresponding to the association of the DMF methyl groups with the s112 residue highlighting the transient nature of these hydrogen bonds and the sensitivity of our probe to the solvation dynamics surrounding the protein. Together this work demonstrates the ability to control the solvation dynamics of a protein through mutation of its primary structure and the application of 2D-IR in reporting of these dynamical changes.

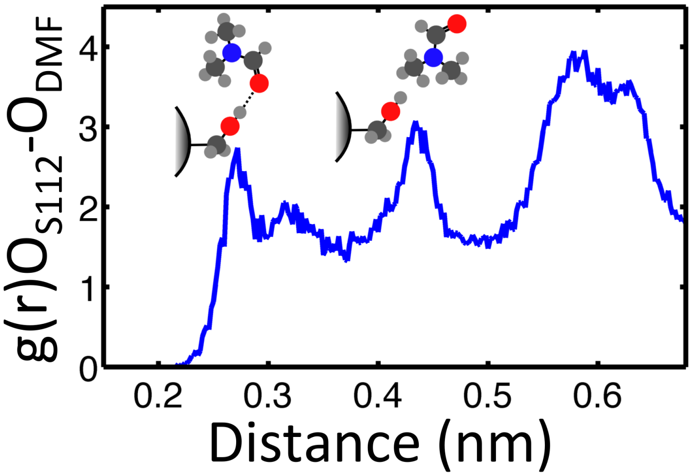

Figure 4.9 RDF between the s112 hydroxyl oxygens and the carbonyl oxygens of DMF. The peak seen at 2.7 Å suggests a hydrogen bonding population of DMF bound to s112. The second peak observed at 4.3 Å shows the association of the methyl groups of DMF with the s112 residues.

4.4 Conclusions

Here we show how the utilization of site-specific vibrational probes and 2D-IR allows for the study of solvation dynamics near the surface of a protein. The site specificity of the probe and modularity of the protein system allows us to study the solvation dynamics arising from the preferential interaction of cosolvent near the probe's binding site. Comparing the observed dynamics of our BTNN-BCT probe bound to WT streptavidin and two mutants, we see a greater than three-fold quickening of the sensed dynamics due to the mutation of the s112 residue. Through simulations this quickening is found to arise due to the loss of the preferential interaction between DMF and the serine residue, which can associate with DMF through hydrogen bonding.

The loss of the preferential solvation in the system causes a breakdown in the heterogeneity of the solvent environment around the probe. As others and we have previously reported on the dynamics of preferential solvation, this heterogeneity is able to introduce an additional slow term to the measured spectral diffusion, originating from the exchange dynamics of the solvent.[46-52] The localization of DMF near the BTNN-BCT probe greatly increases the likelihood of exchange events between the DMF and water to occur due to the entropic cost of preferential solvation. The fact that this preferential solvation is not observed for either the s112a or s112h mutant highlights how the dynamics near the protein surface can be controlled by mutation of the protein sequence.

2D-IR combined with site-specific dynamical markers allows for the profiling of the dynamics near the protein surface.[32, 34] The solvent dynamics near an enzyme's surface is important for substrate binding and product release from the enzyme.[53] This dynamical information is expected to aid in the design and realization of engineered enzymes, offering a method for calculated tuning of the dynamical environment around a protein. This system offers a high degree of modularity allowing future investigations on how steric effects, charges, and hydrogen bond donor strength can similarly affect the dynamics near the protein surface.

4.5 References:

1. Levinthal, C., Are There Pathways for Protein Folding. *J. Chim. Phys. Phys.-Chim. Biol.* **1968,** *65* (1), 44-45.

2. Winter, G.; Milstein, C., Man-Made Antibodies. *Nature* **1991,** *349* (6307), 293-299.

3. Venkatraman, J.; Shankaramma, S. C.; Balaram, P., Design of folded peptides. *Chem. Rev.* **2001,** *101* (10), 3131-3152.

4. Tsien, R. Y., The green fluorescent protein. *Annu. Rev. Biochem.* **1998,** *67*, 509-544.

5. Holliger, P.; Hudson, P. J., Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **2005,** *23* (9), 1126-1136.

6. Fersht, A. R.; Matouschek, A.; Serrano, L., The Folding of an Enzyme .1. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **1992,** *224* (3), 771-782.

7. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A., De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999,** *68*, 779-819.

8. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, Improving and Using Green Fluorescent Proteins. *Trends Biochem. Sci.* **1995,** *20* (11), 448-455.

9. Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X. X.; Oneil, K. T.; Degrado, W. F., Protein Design - a Hierarchical Approach. *Science* **1995,** *270* (5238), 935-941.

10. Vonheijne, G., Membrane-Protein Structure Prediction - Hydrophobicity Analysis and the Positive-inside Rule. *J. Mol. Biol.* **1992,** *225* (2), 487-494.

11. Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010,** *5* (4), 725-738.

12. Rost, B.; Sander, C., Prediction of Protein Secondary Structure at Better Than 70-Percent Accuracy. *J. Mol. Biol.* **1993,** *232* (2), 584-599.

13. Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D., Protein structure prediction using rosetta. *Numerical Computer Methods, Pt D* **2004,** *383*, 66.

14. McGuffin, L. J.; Bryson, K.; Jones, D. T., The PSIPRED protein structure prediction server. *Bioinformatics* **2000,** *16* (4), 404-405.

15. Kelley, L. A.; Sternberg, M. J. E., Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **2009,** *4* (3), 363-371.

16. Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999,** *292* (2), 195-202.

17. Blom, N.; Gammeltoft, S.; Brunak, S., Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999,** *294* (5), 1351-1362.

18. Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D., De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011,** *6* (5).

19. Kries, H.; Blomberg, R.; Hilvert, D., De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **2013,** *17* (2), 221-228.

20. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De novo computational design of retro-aldol enzymes. *Science* **2008,** *319* (5868), 1387-1391.

21. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F.

E.; Baker, D., Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **2010,** *329* (5989), 309-313.

22. Wu, S. J.; Eiben, C. B.; Carra, J. H.; Huang, I.; Zong, D.; Liu, P. X.; Wu, C. T.; Nivala, J.; Dunbar, J.; Huber, T.; Senft, J.; Schokman, R.; Smith, M. D.; Mills, J. H.; Friedlander, A. M.; Baker, D.; Siegel, J. B., Improvement of a Potential Anthrax Therapeutic by Computational Protein Design. *J. Biol. Chem.* **2011,** *286* (37), 32586-32592.

23. Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Players, F.; Stoddard, B. L.; Popovic, Z.; Baker, D., Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol.* **2012,** *30* (2), 190-192.

24. Quinto, T.; Schwizer, F.; Zimbron, J. M.; Morina, A.; Kohler, V.; Ward, T. R., Expanding the Chemical Diversity in Artificial Imine Reductases Based on the Biotin- Streptavidin Technology. *Chemcatchem* **2014,** *6* (4), 1010-1014.

25. Kohler, V.; Mao, J. C.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knorr, L.; Creus, M.; Prost, J. C.; Schirmer, T.; Ward, T. R., OsO4 center dot Streptavidin: A Tunable Hybrid Catalyst for the Enantioselective cis-Dihydroxylation of Olefins. *Angewandte Chemie-International Edition* **2011,** *50* (46), 10863-10866.

26. Hyster, T. K.; Knorr, L.; Ward, T. R.; Rovis, T., Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C-H Activation. *Science* **2012,** *338* (6106), 500-503.

27. Green, N. M., Avidin and Streptavidin. *Methods Enzymol.* **1990,** *184*, 51-67.

28. Turkova, J., Oriented immobilization of biologically active proteins as a tool for revealing protein interactions and function. *Journal of Chromatography B* **1999,** *722* (1-2), 11-31.

29. Olsvik, O.; Popovic, T.; Skjerve, E.; Cudjoe, K. S.; Hornes, E.; Ugelstad, J.; Uhlen, M., Magnetic Separation Techniques in Diagnostic Microbiology. *Clinical Microbiology Reviews* **1994,** *7* (1), 43-54.

30. Dunbar, S. A.; Vander Zee, C. A.; Oliver, K. G.; Karem, K. L.; Jacobson, J. W., Quantitative, multiplexed detection of bacterial pathogens: DNA and protein applications of the Luminex LabMAP (TM) system. *J. Microbiol. Methods* **2003,** *53* (2), 245-252.

31. Diamandis, E. P.; Christopoulos, T. K., The Biotin (Strept)Avidin System - Principles and Applications in Biotechnology. *Clinical Chemistry* **1991,** *37* (5), 625-636.

32. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Crowding Induced Collective Hydration of Biological Macromolecules over Extended Distances. *J. Am. Chem. Soc.* **2014,** *136* (1), 188-194.

33. Osborne, D. G.; Dunbar, J. A.; Lapping, J. G.; White, A. M.; Kubarych, K. J., Site-Specific Measurements of Lipid Membrane Interfacial Water Dynamics with Multidimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2013,** *117* (49), 15407-15414.

34. King, J. T.; Kubarych, K. J., Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy. *J. Am. Chem. Soc.* **2012,** *134* (45), 18705-12.

35. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B* **2012,** *116* (19), 5604-5611.

36. Roberts, S.; Loparo, J.; Tokmakoff, A., Characterization of spectral diffusion from two-dimensional line shapes. *J. Chem. Phys.* **2006,** *125* (8), 084502.

37. Kwak, K.; Park, S.; Finkelstein, I.; Fayer, M., Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: A new approach. *J. Chem. Phys.* **2007,** *127*, 124503.

38. Baiz, C. R.; Mcrobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J., Two-Dimensional Infrared Spectroscopy of Metal Carbonyls. *Acc. Chem. Res.* **2009,** *42* (9), 1395-1404.

39. Nee, M. J.; Baiz, C. R.; Anna, J. M.; McCanne, R.; Kubarych, K. J., Multilevel vibrational coherence transfer and wavepacket dynamics probed with multidimensional IR spectroscopy. *J. Chem. Phys.* **2008,** *129* (8), 084503.

40. Nee, M. J.; McCanne, R.; Kubarych, K. J.; Joffre, M., Two-dimensional infrared spectroscopy detected by chirped pulse upconversion. *Opt. Lett.* **2007,** *32* (6), 713-5.

41. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

42. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006,** *65* (3), 712-725.

43. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

44. Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

45. Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R., Structural Origins of High-Affinity Biotin Binding to Streptavidin. *Science* **1989,** *243* (4887), 85-88.

46. Luther, B. M.; Kimmel, J. R.; Levinger, N. E., Dynamics of polar solvation in acetonitrile-benzene binary mixtures: Role of dipolar and quadrupolar contributions to solvation. *J. Chem. Phys.* **2002,** *116* (8), 3370-3377.

47. Chakrabarty, D.; Chakraborty, A.; Seth, D.; Sarkar, N., Effect of water, methanol, and acetonitrile on solvent relaxation and rotational relaxation of coumarin 153 in neat 1-hexyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. A* **2005,** *109* (9), 1764-1769.

48. Mukherjee, S.; Sahu, K.; Roy, D.; Mondal, S. K.; Bhattacharyya, K., Solvation dynamics of 4-aminophthalimide in dioxane-water mixture. *Chem. Phys. Lett.* **2004,** *384* (1-3), 128-133.

49. Molotsky, T.; Huppert, D., Solvation statics and dynamics of coumarin 153 in dioxane-water solvent mixtures. *J. Phys. Chem. A* **2003,** *107* (41), 8449-8457.

50. Nguyen, C. N.; Stratt, R. M., Preferential solvation dynamics in liquids: How geodesic pathways through the potential energy landscape reveal mechanistic details about solute relaxation in liquids. *J. Chem. Phys.* **2010,** *133* (12).

51. Agmon, N., The dynamics of preferential solvation. *J. Phys. Chem. A* **2002,** *106* (32), 7256-7260.

52. Gupta, S.; Rafiq, S.; Sen, P., Dynamics of Solvent Response in Methanol-Chloroform Binary Solvent Mixture: A Case of Synergistic Solvation. *J. Phys. Chem. B* **2015**.

53. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450* (7171), 913.

Chapter Five
Conclusions & Looking Forward

5.1 Introduction

The dynamical and structural insight at the molecular level offered by 2D-IR has already provided significant insight into the nature of biomolecules and their environments.[1-13] Despite large advances in implementing and understanding 2D-IR spectroscopy in recent years[14-17], much work remains for 2D-IR to become a staple of the biophysical toolkit. To realize the full potential of this powerful multidimensional spectroscopic technique, this work has demonstrated improvements in the implementation, sensitivity, and understanding of 2D-IR investigations of biomolecules. Implementation improvements are made possible by the utilization of advanced signal processing methods[14] and the introduction of new acquisition hardware and software. Sensitivity improvements are attributed to the utilization and synthesis of strongly absorbing IR probes.[2] The rich body of knowledge provided by studying these probes in a variety of systems lends itself well to their further development and application to biological systems. A greater understanding of 2D-IR is accomplished by the development of explicit molecular dynamics simulations, with a fully flexible and parameterized metal carbonyl probe. These simulations represent the first fully parameterized molecular dynamics simulations of metal carbonyl systems in our group allowing for a detailed molecular picture of the system's dynamics.

Current efforts in the group continue to improve the methodologies of 2D-IR building upon the established modularity of the current setup as well as implementing alternative hardware. These methods are summarized below and include multiple methods for further improving the applicability of 2D-IR to investigating biological samples. As the technology progresses and our understanding of the spectra and measured dynamics continues to mature, 2D-IR will continue to be a powerful analytical method for the investigation of complex systems. Recent funding support of the Laboratory for Multidimensional Optical Spectroscopy (LUMOS) center at Michigan highlights the maturity of the field as well as the versatility of these methods.

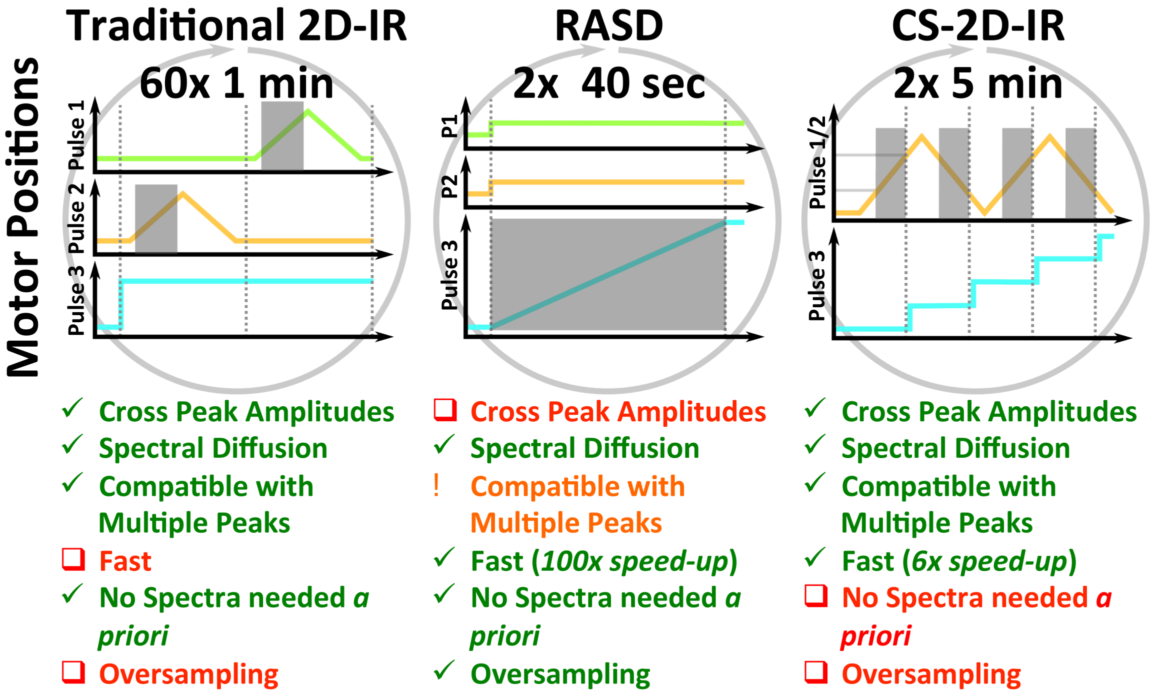

Figure 5.1 Overview of 2D-IR methods implemented in our lab highlighting the pulse sequences used, green, orange, and blue lines, as well as the benefits and limitations of each method. The Grey boxes overlaid on top of the pulses designate the acquisition windows for each method.

5.2 Current Capabilities

Compressed sensing (CS) is a powerful signal processing method that allows for greatly reduced requirements of data acquisition without losing signal information. CS requires that the data which is to be reconstructed is sparse, this is to say that the data can be represented by a relatively few elements from a given basis.[18] The flexability offered by CS as to which basis is used makes this method truly powerful and is attributed to its widespread adoption in the sciences and in other fields.[14, 19-21] CS has been proven applicable to 2D-IR as well as other multidimensional spectroscopies, where the reduction in acquisition time grows as a power of the dimensionality of the approach. By reducing the acquisition time of

2D-IR 6 fold using CS we have improved the ability to study multiple systems as well as systems that are not stable over multiple hours. CS has been shown to provide the same level of detail as traditional 2D-IR spectroscopy utilizing a model system[14], and complements the faster RASD method[15], Figure 5.1, in providing full 2D-IR spectra.

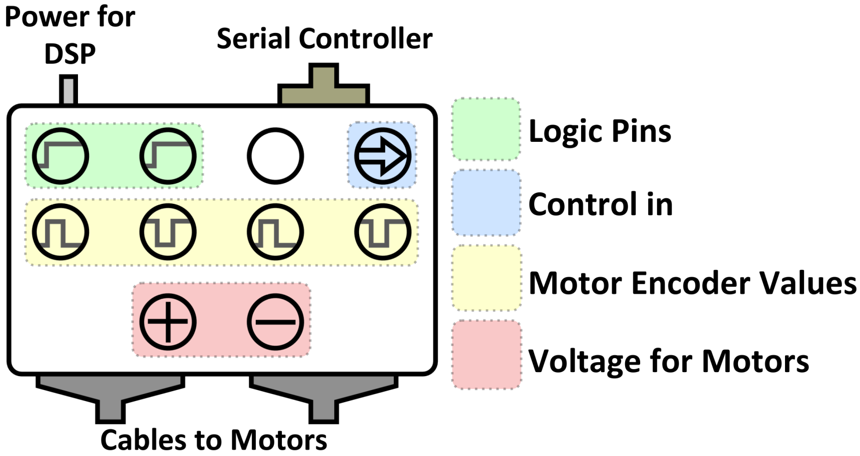

Figure 5.2 Overview of functions enabled by a DSP. The new implementation of the Control in connection (blue box) allows for multiple DSPs to communicate with each other. This enables complex pulse scanning techniques to be employed.

The implementation of CS and RASD in the 2D-IR setup was made possible by software and hardware improvements in the 2D-IR setup. The data acquisition software was modified to allow for full experimental control to occur within LabVIEW. This modification detached the data acquisition's dependency on an outside camera driver, removing the requirement to manually calibrate the camera for acquisition before each experiment. Calibration of the camera is now handled automatically by a variable file in LabVIEW, greatly increasing the day-to-day reliability of the 2D-IR setup as well as providing a higher degree of modularity for future improvements. Additional software improvements were made to the digital signal processors (DSPs), schematically shown in Figure 5.2, which control the motor tracking and motion functions. These were modified to allow for simple real-time communication between multiple DSPs. This allowed us to continuously scan the motor positions of the coherence time motor and send a step signal to the waiting time motor's DSP at the appropriate motor position. Hardware improvements were made in the implementation of a new DAQ card (NI-6320) to the existing 2D-IR setup. The pertinent feature of the card is the ability to monitor simultaneously up 4 motor positions. This has offered us the needed flexibility to record the waiting time delay with a high precision, as is required in RASD, without needing to exchange the wiring of the current setup. These software and hardware improvements have greatly improved the modularity and reliability of our current 2D-IR system creating a stable platform on which to build future developments.

5.3 2D-IR probes for Biology

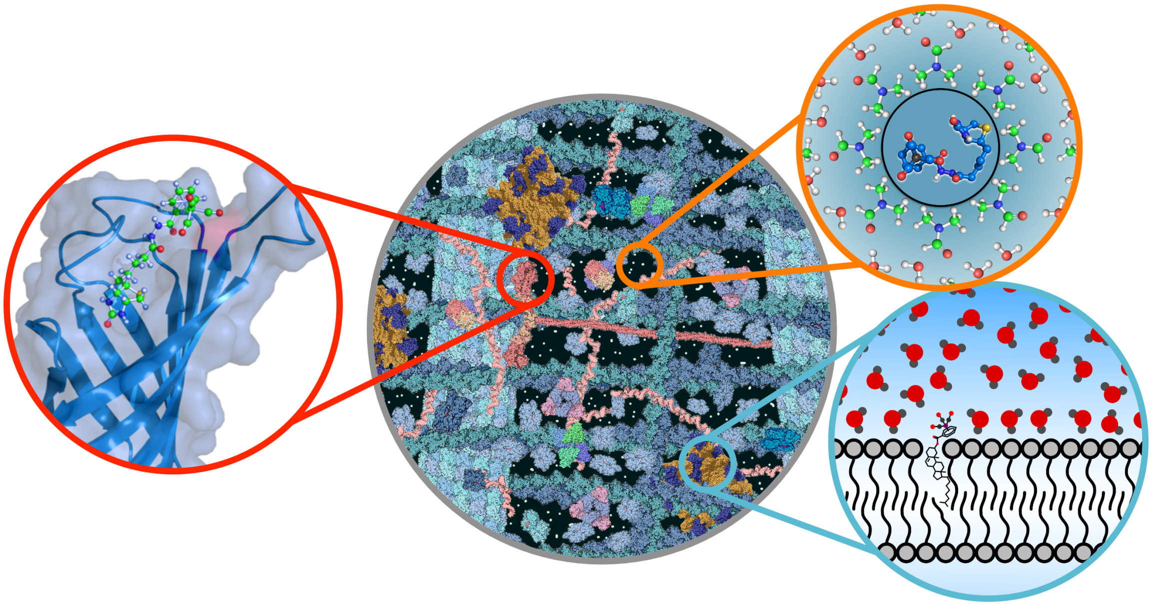

Figure 5.3 2D-IR vibrational probes based off of BCT have been developed to target protein (red), lipid (cyan), and cytosolic (orange) dynamics. BCT probes offer a high degree of modularity allowing for the potential targeting of additional features in the cell.

The biological probes utilized for 2D-IR as described in this text are derivatives of inorganic metal carbonyls, which have served as the foundation for many of the spectroscopic investigations in the group. Benzoic Acid Chromium Tricarbonly (BCT) was selected due to its high availability and the chemistry offered through its acid group. Under mild linking conditions we have shown the ability to create multiple derivatives of this molecular probe. Applying these derivatives to their respective biological settings, Figure 5.3, we have been able to observe the site specific dynamics of solvation at the lipid interface[2] and in the binding site of a protein. Investigations of the protein bound BCT probe show that its local dynamics can be modulated though mutation of a neighboring side chain, offering experimental insight into the rational engineering of these systems towards better catalytic activity. Having demonstrated the ability to link BCT to cholesterol and a biotin analog, we have demonstrated a method that can be adopted to link BCT to other biomolecules including sugars and fatty acids. These derivatives, along with the current probes, could be used to further understand the dynamics of lipid in the cell as well as ligand binding.

The biotin hydrazide BCT (BTNN-BCT) probe described in the text serves as a powerful probe of local protein dynamics. Protein investigations were made possible by our collaborator, Dr. Thomas Ward, who generously provided us with mutants of the protein streptavidin containing a site-specific mutation near the BTNN-BCT carbonyls. As detailed in Chapter 4, utilizing 2D-IR and our probe in this protein system, we are able to discern the change in local dynamics of the protein arising from the mutation of a single amino acid in the protein's binding site. This offers a highly detailed view of the effects of point mutations in the protein system. Characterizing these effects across additional mutations could hold key insight into the rational design of these systems, allowing for control of the binding site dynamics of enzymes through backbone mutations. This work has laid the groundwork for future investigations of the streptavidin BTNN-BCT system and has provided insight on how local protein dynamics are modulated though side chain mutations. Utilizing sensitive dynamical probes, 2D-IR is able to obtain detailed information on biological samples and promises to continue providing insight into these systems' dynamics.

5.4 Preferential Solvation

The processes and reactions that occur within the complex solution environment of the cell must proceed with fidelity to ensure cell viability. Considering the many different types of species in the cytosol of the typical cell, it is clear that certain metabolites will associate preferentially with various proteins and cellular structures. Indeed the often invoked hydrophobic collapse of proteins can

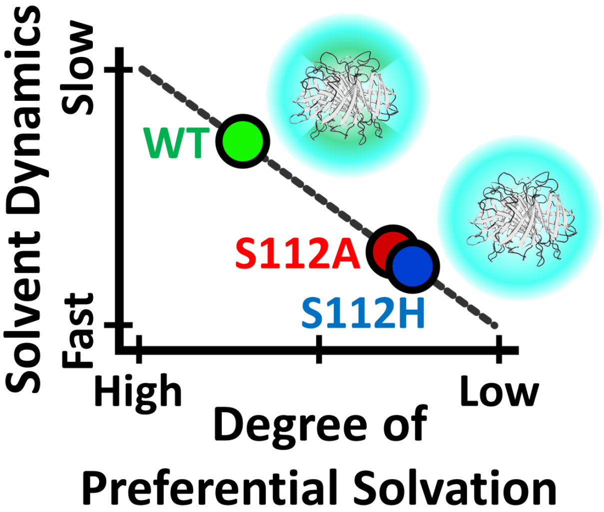

Figure 5.4 Initial studies of BTNN-BCT in a DMF-Water cosolvent system highlighted a striking dependency of the observed solvent dynamics on the degree of probe preferential solvation. This same trend is observed for the streptavidin protein systems offering insight into how local dynamics around the protein can be modulated and how these dynamics can be studied by 2D-IR

be attributed to the preferential interaction of the hydrophobic side chains amongst themselves compared to in aqueous solution. Similar interactions can be attributed to the insertion of some molecules into the lipid environment of the cell membrane. Preferential solvation and interactions restrict the entropy of a system by creating a heterogeneous environment.[22] We have shown 2D-IR to be a sensitive probe for monitoring this heterogeneity though the introduction of a slow timescale in the spectral diffusion measurements, detailed in chapter 3.

This was first realized in the solvent mixture DMF-D_2O, which was studied at multiple concentrations of DMF. In either pure solvent the spectral diffusion is known to decay on the picosecond timescale. This rapid decay in spectral diffusion was lost however in mixtures of these solvents where we instead found that at low compositions of DMF, the spectral diffusion slowed down three-fold. Utilizing explicit molecular dynamics simulations, we were able to show a strong correlation between the degree of the spectral diffusion slow down and preferential interaction of the DMF and our probe. In the framework of preferential solvation, this heterogeneous structuring of the solvent around our probe slows down the system dynamics by introducing an additional, slow, time component to the spectral diffusion. This slower term arises from the solvent exchange near the probe. From modeling our system we found the additional decay terms to quantitatively agree with the timescale for diffusional exchange between DMF and D_2O. This trend is also observed in the streptavidin protein system, Figure 5.4, and is linked to the mutation of a single amino acid near the metal carbonyl probe. Together these measurements highlight 2D-IR as a sensitive measurement of the local dynamics of this biological probe. The observations for BTNN-BCT are expected to be applicable to additional spectroscopic probes and will aid in the mapping and understanding of the dynamical consequences of solvation heterogeneity in biological systems. Through mapping how solvation dynamics of enzymes are modulated by sequence mutations, 2D-IR offers insight into how enzymes can be engineered to increase the diffusion rate of substrates and products in their active sites. This information will prove invaluable in the development of enzymes that are diffusion-controlled, allowing for a deeper understanding of the effects of sequence mutations on enzyme function.

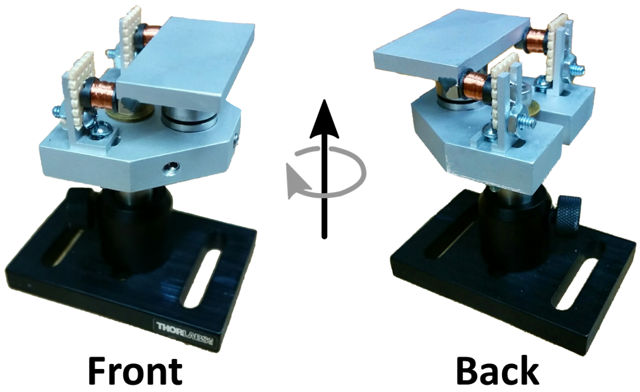

Figure 5.5 Front and back views of phase wobblers built to resonate a Brewster window at 250 Hz to modulate the phases of the pulses. The optical window (not shown) is mounted to the upper platform of the wobbler which is connected to the lower platform utilizing a flex bearing. The two inductors serve as a driver and monitor of the upper platform's resonance.

5.5 Increasing Experimental Capabilities: Phase Wobblers

The setup and implementation of the current 2D-IR data acquisition system has benefited greatly from the adoption of a more modular control system. This has allowed for the implementation of two new data acquisition methods, CS-2D-IR and

RASD, and offers the ability to easily implement novel methods in the future. One such implementation involves realizing the benefit offered by phase control of the pulses used for excitation and detection of the 2D-IR spectra.[23-24] By controlling the phases of these pulses we would be able to operate our traditional 2D-IR setup in such a way as to remove the contributions of the signal arising from the scattering of the IR pulses with the sample. This method has been shown by pulse shaping method to allow for the study of fibroid systems by 2D-IR[24] and would allow for an over ten-fold reduction in the noise for regular 2D-IR spectra.[23] For our traditional pulse scanning 2D-IR setup, phase control is operated in the far field, where by periodically displacing the pulses, we can control their phase when they interact with the sample. This is accomplished by designing 'phase wobblers' as described by Hamm et. al[23], Figure 5.5, to resonate Brewster windows at 250 Hz. These wobblers, when introduced to the delay arms separating the two excitation pulses and the detection pulse, enable for the decoupling of the phases of the three pulses and cancelation of the scattering terms from the signal.

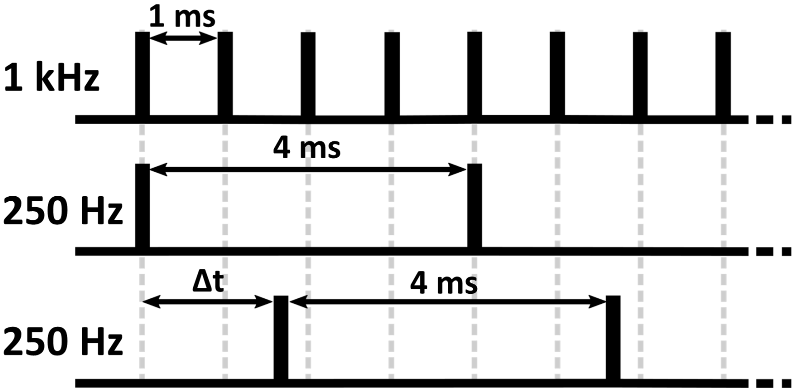

Figure 5.6 Timing diagram highlighting the function of the designed pulse delay generator. The 1 kHz laser pulse train is down sampled to generate a 250 Hz pulse train. An arbitrary delay between the triggering pulse and the corresponding 250 Hz pulse, Δt, can also be controlled to vary the phase between the two pulse trains.

Complementary to the design and building of the phase wobblers, significant work has been accomplished in the electronics required to drive their resonance at 250 Hz. The pulse delay generators have been implemented on an Arduino Due utilizing serial communications over USB to allow for the down-sampling ratio of the triggering pulse as well as a phase offset for the pulse, Figure 5.6. A simple

LabVIEW program has also been written to communicate with the pulse delay generator allowing for simple control of its parameters. The final realization of these phase wobblers requires their placement in the optical path, the development of further triggering electronics to allow for controlling the amplitude of their displacement, characterization of the optical delays, and adaptation of the analysis methods to account for the phase correction. Proof of concept analysis code has been written demonstrating the compatibility of phase wobbling with the continuous acquisition of the coherence time as is common in our experiments. This method will allow for a greatly improved signal to noise in our spectra without the need to average over additional scans. By also allowing the investigation of strongly scattering samples such as aggregates, phase control will broaden the application of 2D-IR to biological samples.

5.6 Increasing Experimental Capabilities: Syringe Pump

It is often advantageous to flow samples in spectroscopy due to the build-up of photoproducts in the sample. For biological samples this can prove difficult to achieve due to the complication of procuring enough material to allow for samples on the milliliter scale needed for traditional peristaltic pumps. To address this large sample requirement work has been done to utilize a syringe pump, allowing for control over volume flow rate and allowing the solvent to be pumped in both

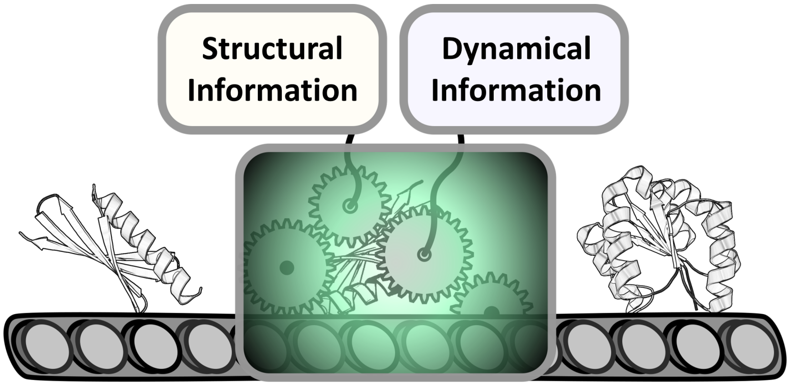

Figure 5.7 To fully realize the ability to manufacture designed proteins structural information will be supplemented by dynamical information as well as other information. These powerful nano-machines will continue serving an important role across the sciences.

directions. A LabVIEW interface for the syringe pump has been implemented and is easily incorporated into the acquisition software of the experiment. This new pump setup will allow for the controlled flow of the sample as well as sample recapture in day-to-day use. Syringe pumps also are well suited to studying systems in microfluidic sample cells, allowing precise control over the flow rate and allowing the study of dynamical systems such as protein unfolding using specialty microfluidic channels.[25]

5.7 Looking Forward

Over the last five years substantial progress has been made, in the 2D-IR community and in our group, on the application and development of 2D-IR to biological systems. We have demonstrated some of the first novel pulse scanning methods in the field, allowing for significant improvements in the acquisition time.[14-15] We have also demonstrated a modular framework for the development and implementation of biologically compatible molecular 2D-IR probe.[2] These probes have demonstrated their ability to monitor the ultrafast dynamics in biological systems as well as provide insight into the manifestation of slower dynamics in these systems. The continued development and application of these probes promises to offer great insight into the dynamics of proteins and their amino acids. Site-specific dynamical information, as is made possible with 2D-IR, will provide a valuable framework to aid in the rational design process of *de novo* enzymes, Figure 5.7. This site-specific dynamical information will complement the

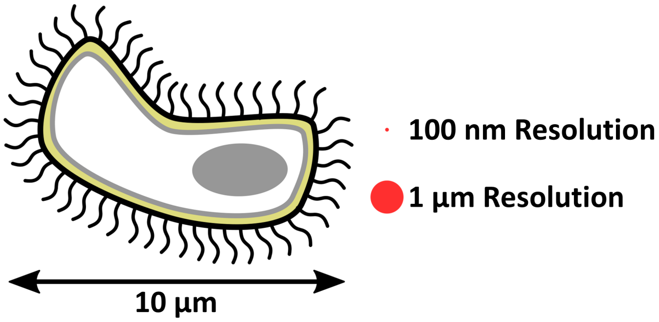

Figure 5.8 The far-filed resolution limit of infrared approaches, ca. 5 µm, is not able appreciable resolve the featured of a cell. By adopting newer near-filed imaging method, which routinely show <100 nm resolution in the IR, 2D-IR will be able to monitor the dynamics across different aspects of a cell.

vast structural knowledge utilized in protein structure prediction tools, helping the protein design community realize their full potential.

Concurrently with the development of new pulse sequences and molecular probes in our group, other groups have realized the ability to use 2D-IR as a microscopic approach.[17] This technique, still in its infancy, promises to offer a significant improvement in the spatial resolution of 2D-IR. Coupled with near field imaging methods, 2D-IR microscopy will enable the discerning of dynamics across different areas of the cell, Figure 5.8. In these biological applications the ability to decrease the acquisition time utilizing our developed methods, and future methods, will greatly aid in the detection of weak signals. 2D-IR is expected to continue evolving as a powerful analytical method as the field matures and the light sources used for illumination continue being developed. By broadening the spectral composition of the laser pulses, one is also able to additionally resolve even higher temporal resolutions. These broader light sources have already been shown to be useful in the measurement and characterization of small molecule structures[16] and promise a detailed look into the tertiary structure of proteins in their native states.

5.8 References:

1. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Crowding Induced Collective Hydration of Biological Macromolecules over Extended Distances. *J. Am. Chem. Soc.* **2014,** *136* (1), 188-194.

2. Osborne, D. G.; Dunbar, J. A.; Lapping, J. G.; White, A. M.; Kubarych, K. J., Site-Specific Measurements of Lipid Membrane Interfacial Water Dynamics with Multidimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2013,** *117* (49), 15407-15414.

3. King, J. T.; Kubarych, K. J., Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy. *J. Am. Chem. Soc.* **2012,** *134* (45), 18705-12.

4. King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J., Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B* **2012,** *116* (19), 5604-5611.

5. Woys, A. M.; Mukherjee, S. S.; Skoff, D. R.; Moran, S. D.; Zanni, M. T., A Strongly Absorbing Class of Non-Natural Labels for Probing Protein Electrostatics and Solvation with FTIR and 2D IR Spectroscopies. *J. Phys. Chem. B* **2013,** *117* (17), 5009-5018.

6. Wang, L.; Middleton, C. T.; Singh, S.; Reddy, A. S.; Woys, A. M.; Strasfeld, D. B.; Marek, P.; Raleigh, D. P.; de Pablo, J. J.; Zanni, M. T.; Skinner, J. L., 2DIR Spectroscopy of Human Amylin Fibrils Reflects Stable beta-Sheet Structure. *J. Am. Chem. Soc.* **2011,** *133* (40), 16062-16071.

7. Reddy, A. S.; Wang, L.; Singh, S.; Ling, Y. L.; Buchanan, L.; Zanni, M. T.; Skinner, J. L.; de Pablo, J. J., Stable and Metastable States of Human Amylin in Solution. *Biophys. J.* **2010,** *99* (7), 2208-2216.

8. Reddy, A. S.; Wang, L.; Lin, Y. S.; Ling, Y.; Chopra, M.; Zanni, M. T.; Skinner, J. L.; De Pablo, J. J., Solution Structures of Rat Amylin Peptide: Simulation, Theory, and Experiment. *Biophys. J.* **2010,** *98* (3), 443-451.

9. Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D., Two-Dimensional IR Spectroscopy of Protein Dynamics Using Two Vibrational Labels: A Site-Specific Genetically Encoded Unnatural Amino Acid and an Active Site Ligand. *J. Phys. Chem. B* **2011,** *115* (38), 11294-11304.

10. Fang, C.; Hochstrasser, R., Two-dimensional infrared spectra of the C-13=O-18 isotopomers of alanine residues in an alpha-helix. *J. Phys. Chem. B* **2005,** *109* (39), 18652-18663.

11. Wang, J.; Chen, J.; Hochstrasser, R., Local structure of beta-hairpin isotopomers by FTIR, 2D IR, and ab initio theory. *J. Phys. Chem. B* **2006,** *110* (14), 7545-7555.

12. Park, J.; Hochstrasser, R., Multidimensional infrared spectroscopy of a peptide intramolecular hydrogen bond. *Chem. Phys.* **2006,** *323* (1), 78-86.

13. Londergan, C.; Wang, J.; Axelsen, P.; Hochstrasser, R., Two-dimensional infrared spectroscopy displays signatures of structural ordering in peptide aggregates. *Biophys. J.* **2006,** *90* (12), 4672-4685.

14. Dunbar, J. A.; Osborne, D. G.; Anna, J. M.; Kubarych, K. J., Accelerated 2D-IR Using Compressed Sensing. *J. Phys. Chem. Lett.* **2013,** *4* (15), 2489-2492.

15. Osborne, D. G.; Kubarych, K. J., Rapid and accurate measurement of the frequency-frequency correlation function *J. Phys. Chem. A* **2013,** *117* (29), 5891-8.

16. Bian, H. T.; Li, J. B.; Wen, X. W.; Sun, Z. G.; Song, J. A.; Zhuang, W.; Zheng, J. R., Mapping Molecular Conformations with Multiple-Mode Two-Dimensional Infrared Spectroscopy. *J. Phys. Chem. A* **2011,** *115* (15), 3357-3365.

17. Baiz, C. R.; Schach, D.; Tokmakoff, A., Ultrafast 2D IR microscopy. *Opt Express* **2014,** *22* (15), 18724-35.

18. Donoho, D. L., Compressed sensing. *Ieee Transactions on Information Theory* **2006,** *52* (4), 1289-1306.

19. Lustig, M.; Donoho, D.; Pauly, J. M., Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* **2007,** *58* (6), 1182-1195.

20. Andrade, X.; Sanders, J. N.; Aspuru-Guzik, A., Application of compressed sensing to the simulation of atomic systems. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (35), 13928-13933.

21. Sanders, J. N.; Saikin, S. K.; Mostame, S.; Andrade, X.; Widom, J. R.; Marcus, A. H.; Aspuru-Guzik, A., Compressed Sensing for Multidimensional Spectroscopy Experiments. *J. Phys. Chem. Lett.* **2012,** *3* (18), 2697-2702.

22. Agmon, N., The dynamics of preferential solvation. *J. Phys. Chem. A* **2002,** *106* (32), 7256-7260.

23. Bloem, R.; Garrett-Roe, S.; Strzalka, H.; Hamm, P.; Donaldson, P., Enhancing signal detection and completely eliminating scattering using quasi-phase-cycling in 2D IR experiments. *OPTICS EXPRESS* **2010,** *18* (26), 27067-27078.

24. Asplund, M.; Zanni, M.; Hochstrasser, R., Two-dimensional infrared spectroscopy of peptides by phase- controlled femtosecond vibrational photon echoes. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (15), 8219-8224.

25. Hertzog, D. E.; Ivorra, B.; Mohammadi, B.; Bakajin, O.; Santiago, J. G., Optimization of a microfluidic mixer for studying protein folding kinetics. *Anal. Chem.* **2006,** *78* (13), 4299-306.

Appendix A
Further details of Compressed Sensing

A.1 Direct Comparison of CS and FFT Analysis

In order to validate the use of Compressed Sensing for the analysis of 2D-IR spectra, we analyze a comprehensive set of data recorded using the traditional scanning approach. Data was collected on an 8 mM solution of rhodium dicarbonyl (RDC) in hexane. Unless otherwise stated the Fourier transformation of the data used the whole t_1 delay range of 10 ps and the CS analysis is done using a range on the order of 600 fs of t_1 delay with $\varepsilon = 5\text{x}10^{-3}$. A Matching Pursuit (MP) algorithm was used (see below for code), for the implementation of CS. The MP algorithm requires four inputs; a tolerance value (ε), the number of iterations to run the algorithm, coherence time data composed of the complex signal field values, and a discrete Fourier transform basis matrix. Before reconstructing spectra, we

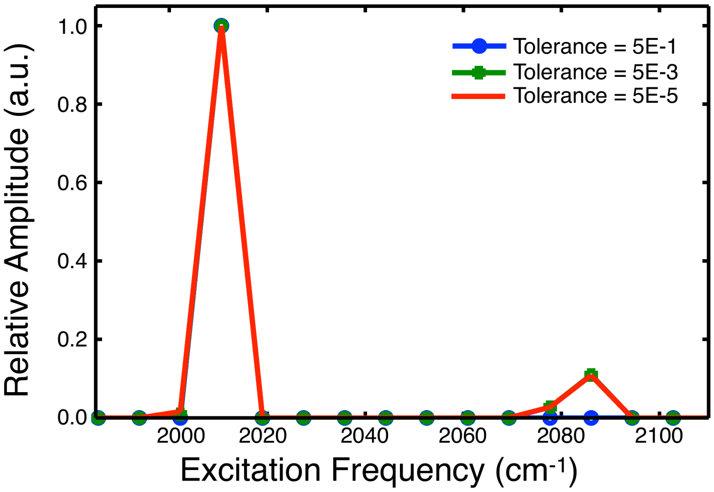

Figure A.1 The ability to resolve peaks of different amplitudes was found to depend on the value of ε. Large values of ε were found to cause the algorithm to miss peaks. Decreasing ε resulted in quick convergence in the ability to reconstruct peaks. Further lowering of ε resulted in similar peak recovery though required more iterations of the function. Data are shown for a single detection frequency and normalized for clarity.

characterized the algorithm by testing the effects of different tolerance values and different coherence times. In the algorithm, ε is the maximum residual, which triggers the end of the algorithm. Smaller values of ε require more iterations of the MP algorithm in order to reduce the fit residual. By using the MP algorithm on coherence time data from a single detection frequency, we found (Figure A.1) that values of ε that are too large resulted in missed frequencies during reconstruction, whereas values of ε smaller than a threshold resulted in diminishing returns.

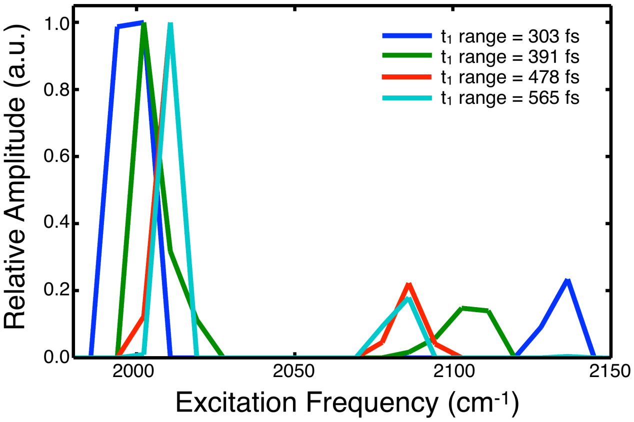

Figure A.2 Dependence of reconstruction on the length of coherence time used. As the coherence time window used is decreased the central frequency and the peak splitting is found to change. Spectra are normalized by the amplitude of the low frequency peak.

To assess the influence of choosing different maximum coherence time windows on the reconstruction, we analyzed a single detection frequency. Smaller maximum coherence time windows decrease the peak position and alter the splitting between the two IR bands (Figure A.2). Although the peaks found using a 303 fs coherence window are shifted, it is possible to correct the peak locations by referencing the excitation frequency axis to a full linear absorption spectrum. A dependence between the window location and the ability to reconstruct the lower amplitude cross peak was noticed for the smaller windows, suggesting a tradeoff between coherence time data and dynamic range in the reconstruction. Convergence of the reconstructed spectra is seen using larger window sizes.

Once characterized, the MP algorithm was used with a coherence time window of 565 fs and an ε value of $5x10^{-3}$ to construct the rephasing 2D-IR spectra

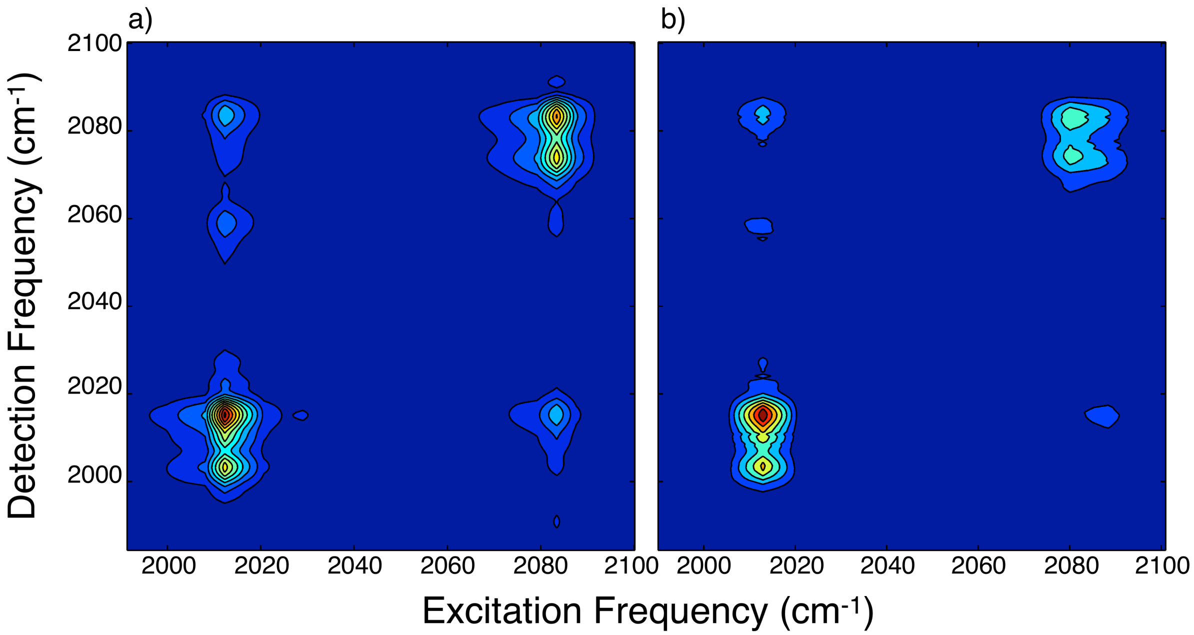

Figure A.3 Comparison of rephasing 2D spectra from analyzing a full, 10 ps, coherence data set with the Fourier transform (a) and analyzing 565 fs of the same data set using the MP algorithm (b). The main peaks along the diagonal are nicely resolved as well as the cross peaks. The excitation frequency of the cross peaks is found to be shifted away from the actual value though their location remains constant as a function of waiting time.

in Figure A.3b. Compared to the spectra obtained using the Fourier transform (FT) of the full 10 ps data set (Figure A.3a), we see that MP reproduces the main features of the spectrum. The cross peaks of the spectra, although slightly shifted on the excitation frequency axis, remain fixed in location at all waiting times (t_2). To follow the waiting-time dependent peak volume dynamics, which contain the majority of information relevant to chemical dynamics, we integrated the rephasing peak volumes for a series of waiting times. Figure A.4 shows the waiting time dependent rephasing peak volumes for the low frequency (2015 cm^{-1}) diagonal peak as well as the cross peak at excitation frequency 2084 cm^{-1} and detection frequency 2015 cm^{-1}. We compare the full FT results with two different choices of the coherence time window width used with the MP algorithm. Figure A.4a,b compare the full FT peak volume analysis with a coherence window width of 522 fs, and Figure A.4c,d compare the full FT peak volume analysis with a coherence window width of 260 fs.

The traces show that with the 522 fs coherence time window, the MP algorithm is able to reproduce the data from the FT spectra well. With the narrower window, the MP algorithm loses some of its ability to follow the FT data, particular the low amplitude excursions of the coherent vibrational quantum beats. By fitting the cross peak rephasing amplitudes to a double exponential and looking at the residuals of both the FT and MP data we see that the data are nearly identical (Figure A.5a). The residuals oscillate with the same beat frequency, as shown by their Fourier transforms (Figure A.5b), further validating the ability of the MP algorithm to recover the peak amplitudes.

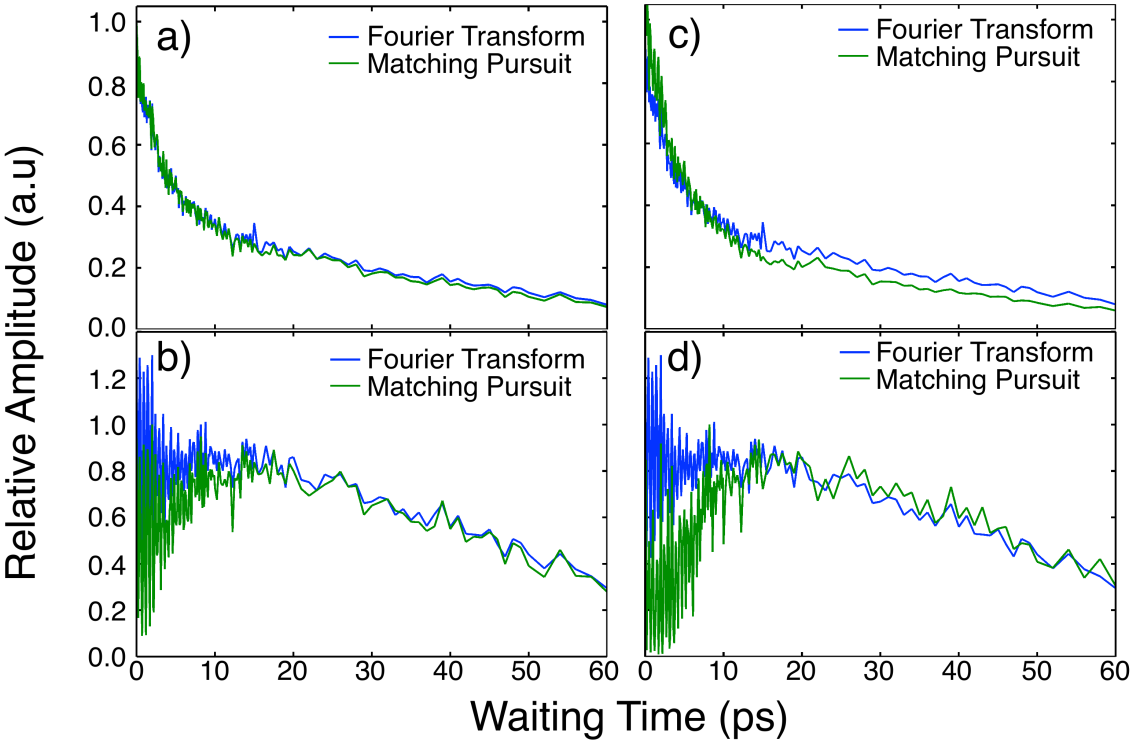

Figure A.4 Rephasing peak volume trace comparison between spectra calculated using the Fourier transform with the full 10 ps data set and spectra calculated using the MP algorithm with 522 fs (a and b) and 260 fs (c and d) coherence time data. Traces comparing the low frequency diagonal peak (a and c) and the cross peak at excitation 2084 cm^{-1}, detection 2015 cm^{-1} (b and d) are shown. The ability of the MP spectra to follow the same dynamics as the FT spectra is apparent when using 522 fs of coherence time. With the shorter data set the ability of the MP spectra to follow the FT spectra is reduced.

By comparing MP and FT analysis of the same data set, we find that it is possible to extract waiting time dependent peak amplitude information from MP

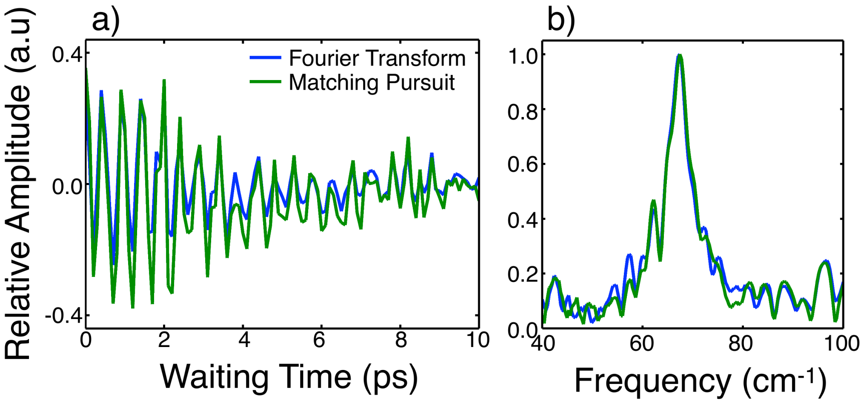

Figure A.5 Further validation of the MP algorithm comes from looking at the residuals of fitting a double exponential function to the cross peak trace, Figure S4 b. The beating pattern seen in the FT spectra is also seen in the MP spectra (a). Fourier transforming the beating pattern recovered by each method reveals the same underlying frequency components in the spectra.

that is nearly identical to that obtained using exhaustive FT analysis on a much larger portion of the data. By first characterizing the MP algorithm to determine proper ranges for ε and coherence time windows needed, we have shown a remarkable reduction in coherence time data required to generate a 2D-IR spectrum in conjunction with compressed sensing strategies. This enhancement can be visualized in Figure A.6 comparing the excitation frequency at a single detection frequency obtained by FT of the full data set, MP of a 565 fs window, and FT of the same 556 fs used for the MP case. Unsurprisingly the peak resolution and position of

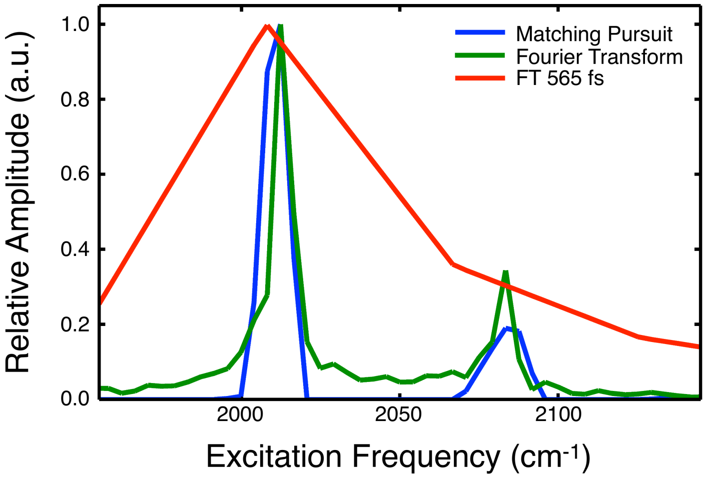

Figure A.6 Comparison of the spectra obtained by looking at a single excitation frequency and using: the full 10 ps data set with the Fourier transform (blue line), 565 fs of the data set with the MP algorithm (green line), and the same 565 fs with the Fourier transform (black line).

the data obtained using the FT on the shorter data set is imperfect. MP along with other compressed sensing reconstruction algorithms, however, are able to recover the main features of the data from these incomplete data sets and show great promise in their application to 2D-IR.

A.2 Matching Pursuit Algorithm

```
Initialize Variables
    $IterationNumber \leftarrow 1$
    $ConstructedSignal \leftarrow 0$
    $Signal \leftarrow InputSignal$

While $(||Signal||_{2,j-1} - ||Signal||_{2,j}) \geq Epsilon$ & $IterationNumber < IterationMax$
        $RecoveredFreq \leftarrow \langle BasisMatrix, Signal \rangle$
        $Index \leftarrow \arg(\max(RecoveredFreq))$
        $Amplitude \leftarrow \max(RecoveredFreq)$
        $Signal \leftarrow Signal - Amplitude * BasisMatrix(Index, :)$
        $ConstructedSignal(Index) \leftarrow ConstructedSignal(Index) + Amplitude$
        $IterationNumber \leftarrow IterationNumber + 1$

Return $ConstructedSignal$
```

A.3 Matching Pursuit Code

```
% Matlab interpretation of cuda code for compressive sampling
% as done by M. Andrecut Engineering Letters, 17:3, EL_17_3_01.
%
% Written by Josef Dunbar, University of Michigan 2013.
%
% Function should be passed:
```

```
%   basisMat = DFT matrix for us, can be changed as needed.
%               Normalized by sqrt(# of rows).
%   timeData = Sparse-in-time data vector to reconstruct
%               frequencies with
%   numItr = Number of iterations to try for convergence
%  epsilon = Convergence threshold for solution
%
% This code solves the basis pursuit denoising problem:
%   [A]*[B] - [C] <= sigma
% using the Matching Pursuit algorithm.
%
% output maxHldr contains the reconstructed signal
% output time contains the runtime of algorithm
% output failed is Boolean array containing non-converged
% pixels
%


function [maxHldr, time, failed] = CPUcs3_fxn(basisMat,...
timeData, numItr, epsilon)
%% Check input for correct dimensions
if(size(basisMat,2) ~= size(timeData,2));
    fprintf('Number of rows in basis matix %d != ...
      number of columns in timeData %d \n', ...
      size(basisMat,2), size(timeData,2));
    return
else
    fprintf('***Data Dimensions Consistent***\n');
end

%% Initialize Program Variables
maxHldr   = zeros(size(basisMat,1),size(timeData,1));
```

```
                        % allocate space for solution
vector
signal    = timeData.'; % convert signal to row vector
normi     = abs(sqrt(sum(signal.^2,1)));
                        % calculate norm-2 of vector
normf1    = normi;      % initialize normf for loop
normf2    = normi;
delNormf  = abs(normf2 - normf1);
converged = delNormf <= epsilon;  % initialize Boolean
array
failed    = 0;
fprintf('***Variables Initialized***\n');

%% Matching Pursuit Algorithm
fprintf('***Starting MP Algorithm***\n');
t = 1;
x = tic;
while (sum(converged) ~= 0 && t < numItr);
 % stop loop when all pix < epsilon or after t iterations
    recovered        = basisMat*signal;
                    % remaining frequencies in signal
    [maximum, index] = max(recovered);
                    % find frequency with largest
contribution
    idx              = ...
       sub2ind(size(maxHldr), index, 1:size(index,2));
                    % get linearized indices of matrix
    maxHldr(idx)     = maximum*diag(converged) +
maxHldr(idx);
             % update frequency coefficients, ignore
converged
    signal           = ...
```

```
-(basisMat(index,:)'*diag(maximum*diag(converged))) +
signal;
                   % remove freq component from signal
    normf1          = abs(sqrt(sum((signal).^2,1)));
    delNormf        = abs(normf2 - normf1);
%   dNorm(:,t)      = delNormf; % look at how values
converge
    normf2          = normf1;
                % determine length of remaining signal
vector
    converged       = delNormf >= epsilon;
      % Boolean vector holding pixels that have not
converged
    t = t + 1;    % increment loop counter
    if t == numItr;
        failed = 1;
    end
%     plot(converged);
end
time    = toc(x);
maxHldr = conj(maxHldr);

%% Print Results
[max2, ind2] = max(delNormf);
fprintf(strcat('***MP Algorithm
Stopped***\n\niterations’,...  ‘= %d\nmax epsilon = %d\n'),
t, max2);
fprintf('***Done***\n');
end
```

Appendix B
Ultrafast 2D-IR and Simulation Investigations of Preferential Solvation and Co-solvent Exchange Dynamics

B.1 Molecular Dynamics Details

Explicit all atom simulations were first equilibrated using a steepest-decent algorithm with a step-size of 1 fs for 100,000 steps and an energy tolerance of 10 kJ mol^{-1} nm^{-1}. 100 ps isothermal and subsequent isothermal-isobaric equilibrium simulations, utilizing the Verlet (leap-frog) algorithm and time-steps of 0.5 fs followed the initial minimization step. Equilibration and production simulations were simulated for each system at 1 atm and 300 K using the Parrinello-Rahman barostat and V-rescale thermostat, with respective time constants of 2 ps and 0.1 ps. System compressibility was held at 4.5 X 10^{-5} bar for all simulations. Particle Mesh Ewald summation was used to calculate electrostatic energies and used a Fourier grid spacing of 1.6 Å with real-space Coulomb and Lennard-Jones cutoffs of 10 Å. Production simulations were run for 20 ns using step-sizes of 1 fs and saving coordinates every picosecond, this generated 20,000 structures for each system.

Table 1 Solvent box compositions for each simulation

System Name	# Water Molecules	# DMF Molecules	Calculated x_{DMF}
x_{DMF} = 1.00	0	733	1.00
x_{DMF} = 0.41	864	607	0.41
x_{DMF} = 0.19	1657	391	0.19
x_{DMF} = 0.11	2195	273	0.11
x_{DMF} = 0.07	2569	194	0.07

B.2 Solution Composition Dependent Conformational Distributions

Monitoring the angle distribution of the biotin solute's linker arm provides additional insight into the conformational space. Previous simulation studies have shown the ability of the ureido group to form hydrogen bonds with the solvent.[1] These hydrogen bonds limit the intramolecular interaction between the ureido and the carboxyl group of the biotin, leading to a predominately open structure of biotin in pure aqueous environments. Functionalizing the carboxyl moiety with BCT and removing the intramolecular interaction of the biotin, the open biotin structure should be further enhanced.

To monitor the conformational distribution of the biotin probe in each solvent we utilize the structures provided by the simulations. For each frame the angle between the center of the benzene ring, alpha carbon to the biotin carbonyl, and the epsilon carbon to the carbonyl (Figure B.1), is calculated. A histogram is then constructed showing the population of the angles for each system simulated. It is found that each distribution is described well by the sum of three Gaussians centered at 82°, 102°, and 148° (Figure B.2).

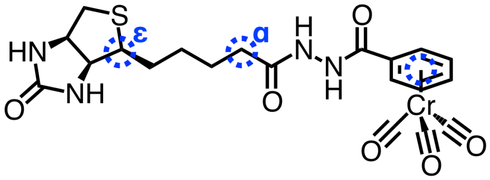

Figure B.1 Structure of BTNN-BCT with sites used for the conformational determination highlighted (blue circles). From left to right: epsilon carbon to biotin carbonyl group, alpha carbon to biotin carbonyl group, and center of benzene ring.

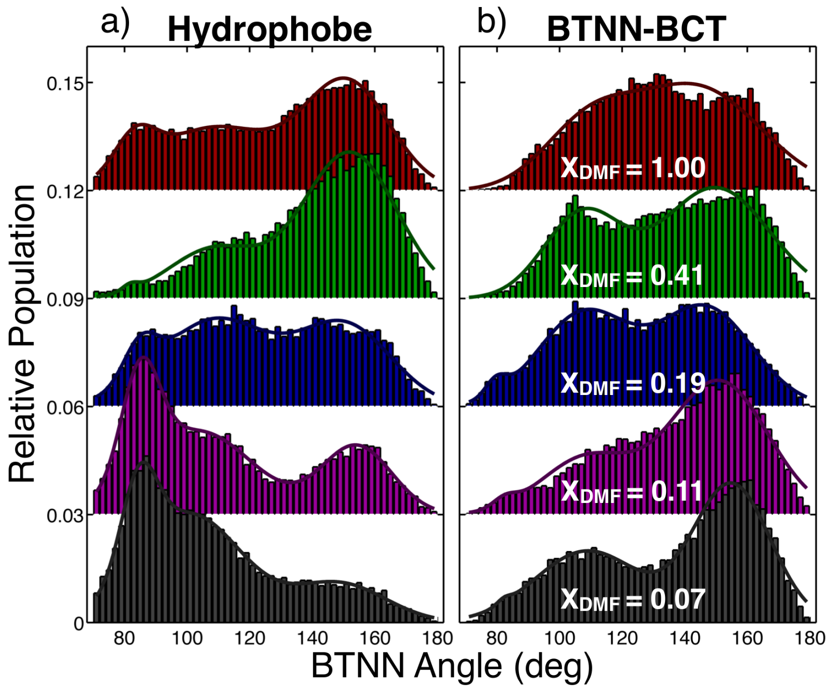

Figure B.2 Angle populations shown with 3-Gaussian fit for the model hydrophobe and BTNN-BCT for each solvent composition. As xDMF increases the model hydrophobe (a) transitions from a compact structure to a more extended state. This is in contrast to BTNN-BCT (b) which exists in a less collapsed state at low xDMF . BTNN-BCT is still found to extend as DMF is added however samples a narrower distribution of angles than the hydrophobe.

For the model hydrophobe the collapsed, small angle, structure of the ligand at low x_{DMF} is found to open with increased DMF. This conformational change contrasts with the BTNN-BCT case where BTNN-BCT is found in a more extended conformation than the model hydrophobe at low x_{DMF}. Both systems show a broadening of the angle distributions as x_{DMF} increases, showing the loss of constrained conformations of the biotin solute in pure DMF. This large variability between the hydrophobe and the BTNN-BCT structures does not manifest in the solvation structure. From calculating the partial RDFs (see main text) for these two models a remarkable agreement in solvent structure is found. This similarity in the RDFs for all values of x_{DMF}, considering the solutes structural conformation, demonstrates that this probe does not significantly perturb the solvent energy and can be thought of as being below the Lum-chandler-Weeks transition length for hydrophobic solvation, ~1 nm.[2]

B.3 Comparison between RASD and traditional 2D-IR

RASD has been previously shown to allow for the collection of Inhomogeneity Index (*I.I.)* data, similar to traditional 2D-IR.[3-4] As can be clearly seen in Figure B.3 the decay in the *I.I.* for the RASD and traditional data set are the same for BTNN-BCT in pure DMF (RASD data has been normalized to the traditional method's *I.I.* initial and final values). Due to the averaging of multiple data sets and the applied 100 fs window RASD shows a significant reduction in noise while also decreasing the experimental acquisition time. Also shown in Figure B.3 is a 2D-IR rephasing spectrum of the symmetric mode of BTNN-BCT in DMF taken at a waiting time of 500 fs. The circular shape of the mode highlights the rapid loss of frequency correlation seen for this system.

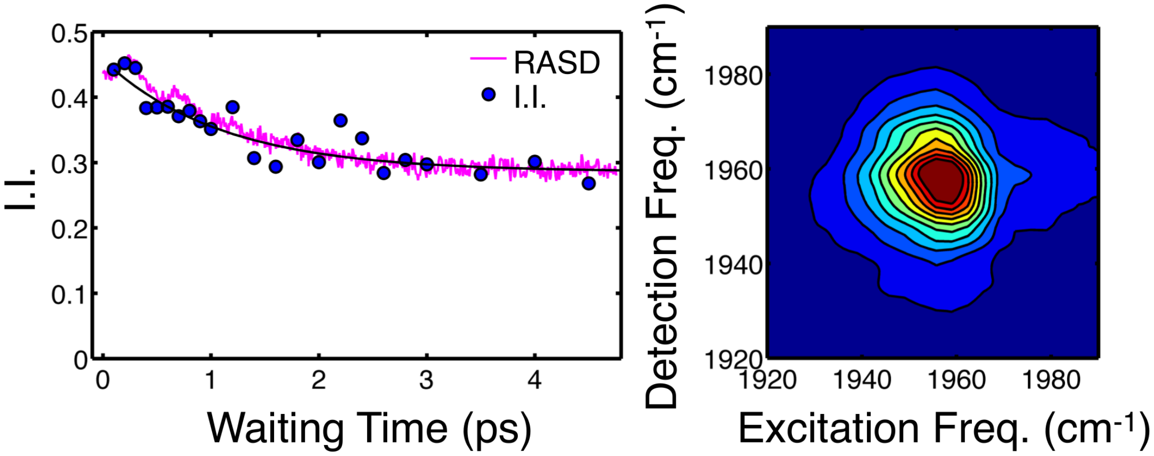

Figure B.3 Comparison between traditionally measured *I.I.* and RASD method (left) for BTNN-BCT in pure DMF. RASD shows identical decay characteristics as the traditional method with greatly improved signal to noise. On the right a 2D-IR rephasing spectrum of the symmetric mode of BTNN-BCT is shown in pure DMF for 500 fs. The apparent circular line shape is indicative of the rapid spectral diffusion experienced by the probe in this system.

B.4 Additional details from preferential solvation fit

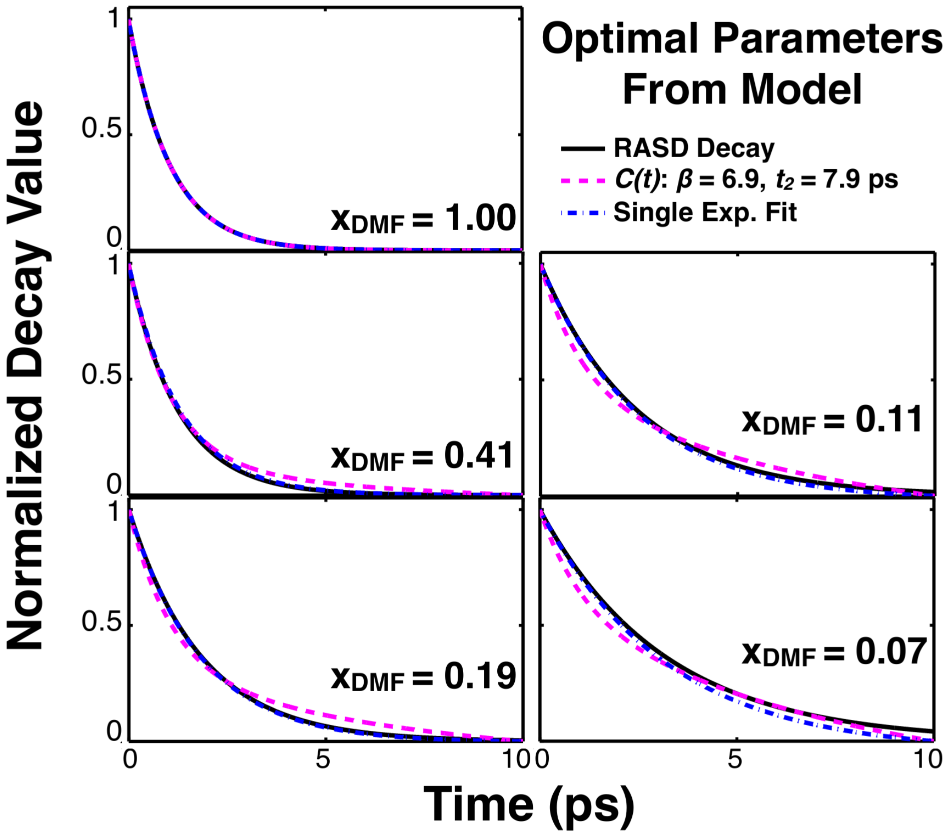

Figure B.4 **Experimentally obtained RASD decay in spectral diffusion (black), *C(t)* calculated from the optimal parameters (magenta), and single exponential fit to the normalized *C(t)* function. By removing the offset and normalizing the double exponential expression for *C(t)* we find excellent agreement between the experimental decay constant and the single exponential fit to the *C(t)*.**

The measured spectral diffusion was fit to a double exponential, *C(t)*, as described in the main text. Due to the difficulties in obtaining accurate decay offsets by the RASD method the data is normalized to remove any contribution from the offset before analysis. Applying this same normalization method to *C(t)* and fitting the model with a single exponential, we find excellent agreement between the model and the experimental data as shown in Figure B.4. Shown are the optimal parameters of t_2 and β. These were obtained from a binary search to minimize the sum of the squares between the model fits and the experimental data. The high degree of over-sampling offered by the RASD method provides robust statistics to distinguish the decay rate between the systems studied.

B.5 References:

1. Lei, Y.; Li, H. R.; Zhang, R.; Han, S. J., Molecular dynamics simulations of biotin in aqueous solution. *J. Phys. Chem. B* 2004, *108* (28), 10131-10137.

2. Lum, K.; Chandler, D.; Weeks, J. D., Hydrophobicity at small and large length scales. *J. Phys. Chem. B* 1999, *103* (22), 4570-4577.

3. Nilsen, I. A.; Osborne, D. G.; White, A. M.; Anna, J. M.; Kubarych, K. J., Monitoring equilibrium reaction dynamics of a nearly barrierless molecular rotor using ultrafast vibrational echoes. *J. Chem. Phys.* 2014, *141* (13).

4. Osborne, D. G.; Kubarych, K. J., Rapid and accurate measurement of the frequency-frequency correlation function *J. Phys. Chem. A* 2013, *117* (29), 5891-8.