Introduction to Protein 2D IR Spectroscopy

Carlos R. Baiz, Mike Reppert, and Andrei Tokmakoff
Massachusetts Institute of Technology, Cambridge Massachusetts 02139
June 1, 2012

INTRODUCTION

Proteins are molecules that behave in beautiful and astounding ways in the course of their biological function, and all biological processes involve protein conformational changes. These processes might be enzyme catalysis, transport and signaling, dynamic scaffolding for structures, charge transfer, or mechanical or electrical energy transduction. Our view of such processes is colored by the methods we use to study them, and most of what we know about proteins is based on structural studies and biochemical assays. Biochemists and biologists think of the processes in which proteins engage in terms of directed motion, often illustrating them through movies, but in current experiments, one rarely actually watches the conformational changes that occur directly. The method of two-dimensional infrared (2D IR) spectroscopy(5-7) is providing new approaches that can be used to characterize features that are masked by traditional methods, particularly for visualizing conformational dynamics on picosecond to millisecond time scales, and characterizing conformational variation and structural disorder. Furthermore, it also is proving useful for samples that are difficult to study by traditional techniques, such as protein aggregates and amyloid fibers, (9, 10) intrinsically disordered peptides(11) and membrane proteins. (12, 13)

Two-dimensional infrared (2D IR) spectroscopy was developed as a tool to study transient molecular structure and dynamics in solution. As a vibrational spectroscopy, it directly interrogates the vibrations of chemical bonds and how the many vibrations of a molecule and its environment interact with one another. Inspired by the two-dimensional techniques first developed in the field of nuclear magnetic resonance, 2D IR spectroscopy spreads a vibrational spectrum over two frequency axes which report on how excitation of a vibration with a given frequency influences all other vibrations within a detection window, following a waiting time. Spectral features, in the form of frequencies, amplitudes, and lineshapes, and the evolution of these features with time are used to understand structural connectivity in space and time and leads to new avenues for studying molecular structure, dynamics, and dynamical heterogeneity. Given enough information on the mechanism of vibrational interactions, one can model a spectrum to reveal structural information on the vibrations being observed.

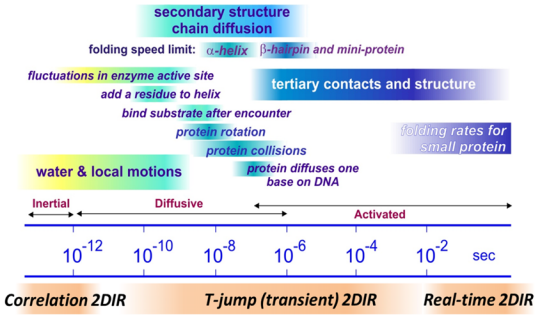

Figure 1 – Timescales of protein dynamics

From subpicosecond water fluctuations to hour-long aggregation processes, biophysical processes vary over many orders of magnitude in time, and therefore require methods that can span a wide range of timescales. Figure 1 shows a variety of processes along with their corresponding timescales.(17, 18) Since the 2D IR measurement is made with a picosecond or faster "shutter speed," it captures information on molecular structure in solution on a time scale fast compared to most dynamics, and is uniquely positioned to probe many of these processes. Correlation-2D IR, in which the waiting time is varied, characterizes dynamics on the picosecond timescale. Non-equilibrium variants of 2D IR, such as temperature-jump 2D IR, expand the dynamic range of the technique from picoseconds to milliseconds and enable the study of transient processes such as protein folding or association. (21) Longer timescales can be reached by rapid-acquisition continuous-scanning 2D IR methods.

Careful consideration of the experimental objectives facilitates the design of 2D IR experiments. For example, short range structure and dynamics, such as hydrogen bonding environments, are visualized through 2D lineshape analysis(22) or waiting time dynamics experiments(24) on localized vibrations, whereas delocalized amide I spectra give a global

view of the protein architecture.(26) Triggered experiments are used to study non-equilibrium conformational dynamics. Finally, interpretation of experimental spectra can be done on multiple levels: empirical rules offer basic structural information, whereas a more sophisticated structural view involves spectral modeling. Analogous to NMR spectroscopy, 2D IR provides a set of structural constraints that can be used with structure-based modeling methods to provide structures that are consistent with the experimental constraints. (22) One approach describes 2D IR spectra in terms of semi-classical electrostatic maps that describe the structure-dependent vibrational couplings and frequency shifts within the protein. (14, 15, 23, 27) These techniques can be interfaced with molecular dynamics (MD) simulations and Markov-state models, or with alternative structure-based computational biophysics simulation models.

This chapter offers an introduction of 2D IR methods aimed at new users who wish to use the technique in order to shed light on questions related to structure and dynamics of proteins. Sections 2 and 3 introduce the methods, theory, and applications of backbone 2D IR spectroscopy of as a probe of protein structure, protein solvation, folding and binding, and protein-protein interactions. To date the majority of 2D IR spectroscopy has been focused on amide I vibrations, largely composed by C=O stretching and N-H wag vibrations of the amide moiety. Delocalized over the entire protein backbone, amide I modes are sensitive to the global structure of the protein. There are, however, many alternative vibrational probes, which can be tailored to probe localized structure, solvent exposure and hydrogen bonding. Cross-correlations between different vibrational modes compound the information contained within individual modes in order to provide new structural insights that are not available from single vibrations. The merits of localized vibrational probes and other backbone modes are discussed in Section 3. Vibrational modeling provides the crucial link between structure and spectra. Section 4 describes the framework for modeling protein 2D IR spectra and the methods available for protein 2D IR analysis. Section 4 is geared towards theoreticians with experience in molecular dynamics simulations and a basic background in quantum mechanics. Section 5 introduces recent examples of 2D IR spectroscopy in order to illustrate applications of 2D IR to study protein structure, conformational heterogeneity and solvent exposure, and transient temperature-jump induced denaturation.

BACKGROUND

2.1 Amide I modes

Amide vibrations of the polypeptide backbone are the most commonly used vibrations in infrared studies of proteins. (32-35) Of these, the amide I vibrations observed between 1600-1700 cm^{-1} are of particular interest since they provide distinct spectroscopic signatures of secondary structure and hydrogen bonding contacts in peptides and proteins. As illustrated in Figure 2, amide I vibrations are combinations of C=O stretching and in-plane N-H bending vibrations of the backbone amide units which have a strong IR transition dipole moment. The physical interactions or *couplings* between amide I vibrations of the protein's multiple peptide units give rise to delocalized vibrations of the protein backbone. With the exception of proline, the side chain vibrations do not interact strongly with the amide I vibration. Thus, amide units are chemically identical throughout the entire backbone and have similar vibrational frequencies, allowing for efficient coupling of the local modes on each amide to form delocalized vibrations. We refer to the vibrations of the peptide unit as local amide I modes (or sites) which serve as a basis to describe the delocalized vibrations—called *normal modes* for harmonic systems or more generally *excitons*, a term borrowed from solid state physics(36)—as linear combinations of the sites.(35) Figure 2 shows a typical β-sheet normal mode encoded onto the ribbon diagram of ubiquitin. Similar to vibrations in small molecules, the character of the normal mode is dictated by the local symmetry of the protein backbone: α-helices and β sheets exhibit modes characteristic of the local arrangement of the residues within the structure.(37-41)

As discussed in more detail in Section 4, structural models for the coupling interactions between peptide units provide a means of connecting the exciton band information contained in IR spectra to protein structure. Unfortunately, broad absorption lineshapes and the large number of delocalized modes obscures much of the information present in amide I spectra, making the task of extracting structure from absorption spectra particularly difficult. Two-dimensional spectroscopy unpacks

structural information by spreading the spectral contents onto two frequency axes and measuring frequency correlations between multiple vibrations.

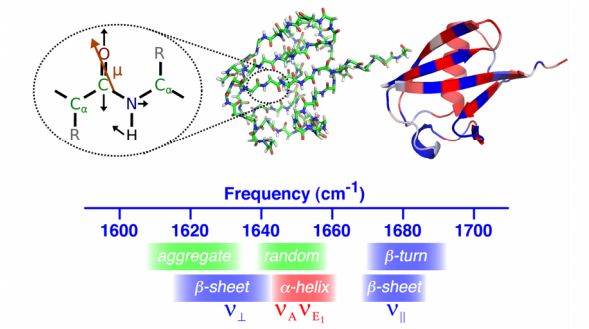

Figure 3 – Amide I spectroscopy of proteins: (top) atom displacements and transition dipole moment associated with amide I vibrations in a single amide unit, along with the contributions of different units to a normal mode in Ubiquitin. The amplitude and phase of the vibration is encoded on the grayscale intensity: dark, light are 180 degrees out of phase. (bottom) Representation of experimentally-observed IR bands corresponding to different structural motifs.

2.2 Structural interpretation of amide I infrared spectra

Figure 2 shows an empirical relationship between common secondary structures and spectral features observed in amide I spectra. α-Helices exhibit a single peak centered near 1650 cm^{-1}, while β-sheets exhibit two peaks centered near 1630 and 1680 cm^{-1} whose center frequencies and intensities depend on the length and number of β strands. (26, 38, 42) Vibrations associated with unstructured regions also appear as a single broad peak near 1650 cm^{-1}. Figure 3 shows FTIR and 2D IR spectra of proteins with different secondary structures: Myoglobin (α-helix), Ubiquitin (mixed α/β), and Concanavalin A (antiparallel β-sheet). Here we provide a brief interpretation of amide I 2D IR spectra, more detailed descriptions of 2D IR spectroscopy and a step-by-step interpretation of a 2D IR spectrum are provided in the next section.

Two-dimensional spectra can be interpreted as a two-dimensional map that links a set of excitation frequencies (ω_1), to a set of detection frequencies (ω_3). The features observed along the diagonal axis correspond to excitation and detection of the same frequencies, and can be related approximately to the absorption spectrum. Off-diagonal peaks correspond to exciting one particular peak and detecting emission from a different one, which reflects

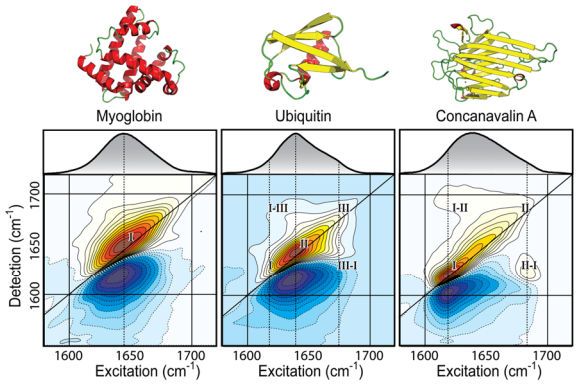

Figure 2 – Amide I absorption and 2D IR spectra of Myoglobin, Ubiquitin, and Concanavalin A. Dashed lines indicate the approximate peak positions. Cartoon structures are shown for reference (PDB codes: 1MBO, 1UBQ, 1JBC). Non-linear scaling of the contours emphasizes the low-amplitude features in the spectrum.

coupling or energy transfer between two modes. As a result of vibrational anharmonicity each peak in the spectrum appears as a positive/negative doublet. When these peaks are in a congested spectrum, complex lineshapes can arise. Myoglobin exhibits a round diagonal peak centered near 1650 cm^{-1} (denoted as *I*). Concanavalin A shows two peaks centered near 1620 cm^{-1} and 1670 cm^{-1}, labeled *I* and *II* in the 2D spectrum. Since the two peaks are broad, the spectrum appears elongated along the diagonal. The presence of narrow off-diagonal peaks (*I-II* and *II-I*) however, indicates the presence of two distinct transitions underneath the broad diagonal lineshape. Beta-sheet amide I spectra provide a clear illustration of the information content gained by two-dimensional spectroscopy over one-dimensional techniques.(43) In the case of Ubiquitin, contributions from the α-helix and β-sheet residues show a feature centered near 1640 cm^{-1} (*II*) with a strong diagonal elongation towards the low frequency (*I*) high-frequency regions (*III*). Cross peaks (*I-III* and *III-I*) are observed near [1640, 1680] cm^{-1}, a clear signature of β-sheet contribution. Similarly, two broad peaks centered near 1640 and 1680 $cm^{-1}$are observed in the corresponding FTIR spectrum. Mixed α/β proteins, such as ubiquitin tend to exhibit lineshapes resembling the combination of purely α-helix and purely β-sheet spectra. The interpretation of these features has been aided greatly by 2D IR studies,(26, 43, 44) as described below.

2.2.1 Spectral signatures of secondary structure

Similar to most spectroscopic techniques, structural

assignments are facilitated by theoretical models and simulations. (37, 45, 46) Standard modeling approaches are described in section 4. This section provides a qualitative interpretation of the spectral features associated with different secondary structures.

α helices

Simulations for idealized α-helices have revealed the presence of two main IR-active modes with A and E_1 symmetries: The most intense A mode, accounting for approximately 70% of the intensity, involves the in-phase oscillation of all the residues in the helix whereas the E1 mode has a periodic phase shift of approximately 3.6 residues per cycle. (27, 39, 47) The A mode red shifts from ~1660 cm^{-1} for helices of 5-10 residues in length to ~1650 cm^{-1} for helices of greater than 20 residues. In contrast the E_1 mode shows almost no frequency dependence with respect to the helix length. The A-E_1 frequency splitting and E_1 phase twist are attributed to the large positive coupling observed between adjacent peptide units and the negative coupling between hydrogen bonded ones. Since the A mode accounts for much of the main band intensity in α-helices, a frequency dependence on the helix length is observed for primarily α-helical proteins. The ~10 cm^{-1} shift, however, is relatively small compared to the ~60 cm^{-1} total width of the amide I band. Polarization-controlled 2D IR experiments have been able resolve the two modes experimentally in short 21-residue helical peptides the E_1 and A modes are centered at 1638 cm^{-1} and 1650 cm^{-1} respectively.(48)

3_{10} Helices

The spectral character of 3_{10} helices is very similar to those observed in α-helices: low-frequency A modes and higher frequency E modes are split by 10-15 cm^{-1} depending on the length of the helix. (49) Vibrational circular dichroism and FTIR, as well as computational studies suggest that the peaks of a 3_{10} helix are red-shifted by 5-10 cm^{-1} compared to α helices. (50, 51) Due to the broad lineshapes, it can be difficult to distinguish between α and 3-10 helices via IR absorption spectroscopy, but recent studies suggest that amide I 2D IR spectra contain specific spectral signatures for 3_{10} helices that can serve to spectrally separate them from α helices. (52-54)

β-sheets

Antiparallel β-sheets are constructed from a repeating rectangular four-peptide unit.(38, 43, 55) For idealized infinite sheets, the vibrations of a single unit describe the observed vibrations of the sheet. The unit cell for an antiparallel sheet consists of four oscillators therefore four distinct modes are observed, only two of which are IR active. The lowest IR active mode, which carries most of the intensity, appears as a narrow band near 1630-1640 cm^{-1} (see Concanavalin A spectrum in Figure 3), and involves the in-phase oscillation of sites on adjacent strands. In the four-peptide unit, this mode involves the in-phase oscillation of units that lie in opposite corners along the diagonal of the rectangle. Vibrations of this character are denoted as $\upsilon_{\perp}$ since the transition dipole lies perpendicular to the β-strands.(56-60) The high-frequency IR active mode, denoted $\upsilon_{//}$, is centered ~1670 cm^{-1} and has lower overall intensity since it involves the in-phase oscillation of adjacent residues along the β-strands, and are out-of-phase with respect to the hydrogen-bonded neighboring residue on the adjacent strand. Figure 4 shows simulated spectra for idealized antiparallel β-sheets. As the size of the sheet increases, $\upsilon_{\perp}$ red-shifts by ~20 cm^{-1} and gains additional intensity with respect to $\upsilon_{//}$. The high-frequency $\upsilon_{//}$ peak position remains nearly unchanged, and therefore serves as an internal reference marker of β-sheet structure. Thus, the red shift of the $\upsilon_{\perp}$ band in β-sheets serves as an indicator of the β-sheet size.(38)

Parallel β-sheets can be described as a repeating set

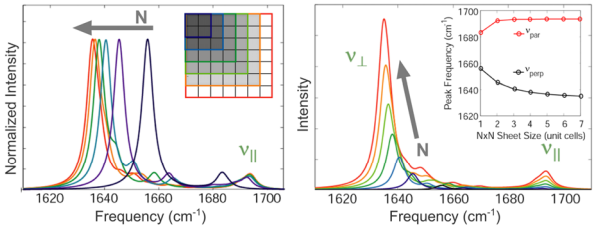

Figure 4 – Simulated spectra of idealized N×N-unit antiparallel β-sheet structures ranging from 2×2 to 7×7 units. Each unit corresponds contains 4 residues arranged in a 2×2 square configuration. (right) Size-dependence of the $\upsilon_{//}$ and $\upsilon_{\perp}$ peak positions. (right) Size dependence of the peak intensities. The inset shows the center frequencies of the two peaks as a function of the sheet size.

of two-residue unit cells. (41) Therefore the vibrational spectrum of an infinite parallel β-sheet will show two main vibrational bands. In a 4x4 residue idealized sheet, the high-frequency band,

appears near 1660 cm^{-1} but carries virtually no intensity. Most of the oscillator strength is shifted to the low-frequency band, centered around 1635 cm^{-1}, described as out-of-phase oscillations of residues within the same strand. Transition dipole modes for vibrations of this character lie perpendicular to the β-strands.

β-turns

Contributions from residues in β-turn configurations are difficult to isolate experimentally. For this reason the spectral assignment of β-turns, is largely derived from simulations. In general β-turn peaks appear near 1680 cm^{-1}. (61) Since the $\upsilon_{//}$ band of β-sheets also appears in the same frequency range, and both peaks are expected to have low amplitudes, the first due to the relatively few oscillators in β-turn configurations, the second due to the low oscillator strength of $\upsilon_{//}$, peaks in the 1680 cm^{-1} and above can be easily mis-assigned. One must therefore be particularly careful when assigning spectral features associated with β-turns.

Coils

Unstructured coils appear as a broad, featureless peak centered near 1650-1660 cm^{-1}. Lacking long-range order, coils exhibit random coupling patterns which give rise to broadened peaks. In addition the increased solvent-exposure of the backbone contributes to the variation in site frequencies and further broadens the peaks. The difficulty in resolving α-helices from random coils is one of the shortcomings of amide I IR absorption spectroscopy. Since random-coils are more solvent exposed, hydrogen/deuterium exchange experiments(62) and waiting-time experiments(63) combined with isotope labeling can aid in assigning individual residues to secondary structures.

2.2.2 Doorway Mode Analysis

Intuitive interpretation of individual amide I modes is difficult for proteins for two main reasons: Firstly, a large number of normal modes is packed into a small spectral region, and modes have partially mixed character, for example, individual modes are often delocalized over α-helices or β-sheets. Secondly, the character of the normal modes is highly dynamic, and slight changes in structure or solvent environment can remix the modes. Therefore, while the normal mode picture provides a convenient framework for describing the models, individual normal modes have limited relevance for interpreting spectra. Since normal modes within the same frequency region share similar overall characteristics, instead of focusing on individual modes, a more intuitive interpretation is obtained via a doorway mode analysis. In brief, the doorway mode analysis described here relies on singular value decomposition to project out the shared vibrational features within a set of normal modes in a small frequency window.(37, 40, 64) Intensity-weighted components are referred to as *doorway modes* and provide an intuitive visualization of the normal modes' character within the selected frequency region. The character of the principal component modes is less affected by small changes in structure allowing for comparison of modes among different conformations of a protein or even different proteins.

Figure 5 shows simulated absorption spectra of four proteins with varying α-helix/β-sheet conformation: from 50% β-sheet (Concanavalin A) to 75% α-helix (Myoglobin). In order to interpret amide I spectral signatures, each residue is assigned to one of four structures: α-helix, β-sheet, β-turn, and coil. The structural character of the doorway state is calculated by amplitude weighting the contribution of each oscillator to an individual secondary structure. For example, a β-sheet doorway mode will have a large number of amplitude-weighted residues in a β-sheet conformation contributing to the mode, in comparison to the α-helix. Doorway mode analyses (Figure 5, c) show that β-sheets contribute mostly in the regions below 1630 cm^{-1} and above 1670 cm^{-1}, whereas α-helices contribute intensity mainly towards center of the amide I band. Random coils also contribute in the 1650-1670 cm^{-1} region, and overlap some of the helix peaks. These observations are consistent with spectral assignments shown in Figure 3.

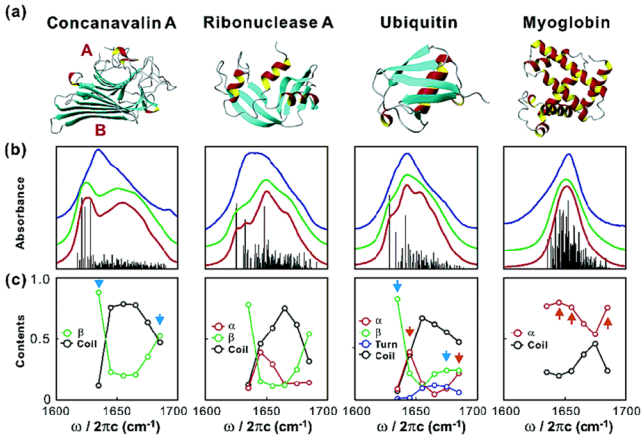

Figure 5 – (a) Crystal structures of the four proteins selected for doorway mode analysis. (b) (top) experimental absorption spectra of proteins in D2O; (middle) spectra calculated with site disorder; (bottom) spectra calculated without site disorder. Black lines represent individual mode frequencies and intensities. (c) amplitude-weighted structure of the doorway-modes. Adapted from [Chung, H. S.; Tokmakoff, A. J Phys Chem B 2006, 110, 2888-2898.]

2.2.3 Beyond secondary structure

The basic examples described above are not sufficient to explain all the features observed in 2D IR, suggesting that spectra contain more detailed information related to the specific architecture of the protein. For example, 2D IR spectra of proteins with similar percentage of residues in α-helix and β-sheet conformations exhibit very distinct features, suggesting that the three-dimensional architecture of proteins is encoded in the spectrum.(26) To date, the spectral signatures of super-secondary motifs remain largely unexplored. Further investigations will be required in order to understand questions such as: How does the spectrum of a twisted β-sheet differ from a flat, idealized, β sheet? How does one distinguish parallel and anti-parallel β-sheet contacts in real systems? Do multiple secondary structures in well-defined registry, for instance a coiled-coil or a protein oligomer, have distinct intermolecular couplings? These are areas of active research.

Amyloids and aggregates

Misfolded proteins, such as the extensively-studied β-amyloid peptide, aggregate into insoluble fibrils with well-defined stacked parallel cross-β-sheet structures. A high-degree of order and limited water penetration into the fibrils contributes to the sharp β-sheet peaks observed in absorption spectra.(9, 10, 65, 66) The kinetics of amyloid aggregation can be easily monitored with infrared spectroscopy, and molecular insights into the aggregation mechanism are aided by isotope labeling experiments. (9) (67) Other insoluble protein aggregates appear as a single peak centered near 1620 cm^{-1}. Narrow lineshapes and red-shifted frequencies (~1620-1630 cm^{-1}) make IR spectroscopy particularly sensitive to the presence of aggregates in the sample. Non-linear methods further enhance the spectral features, such that even a small amount of aggregate present in the sample produces intense features which dominate the amide I spectrum.

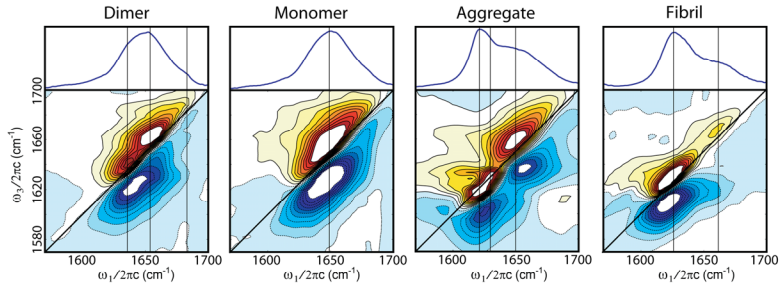

Figure 6 – Two dimensional spectra of insulin in different forms. Solid and dashed contours represent positive and negative features respectively.

Spectra of insulin (Figure 6) illustrate how multidimensional spectroscopy has the sensitivity to probe the inter-protein contacts and interactions responsible for dimerization, oligomerization, and aggregation. In comparison with monomer spectra, dimers show features associated with β-sheets: a two-peak structure and a more pronounced β-ridge, indicative of an inter-protein β-sheet present in the dimer and suggesting that the interfacial residues of the monomer remain in a disordered configuration. The aggregate and fibril samples show a two-peak pattern but with different center frequencies and peak intensity ratios, indicating that the aggregate has a significant percentage of residues in α-helix or disordered regions, whereas the fibril exhibits features characteristic of ordered β-sheets. Though qualitative, the simple interpretation of the data helps characterize the structural changes associated with protein association and fibril formation. Recent studies on the amyloid-β peptide illustrate how quantitative modeling, and incorporation of isotope labels, provides detailed insights into the mechanisms of protein aggregation and fibril formation.(67)

2.3 Spectrally Isolated Sites: Hydrogen Bonding and Isotope labels

2.3.1 Hydrogen Bonding

Hydrogen bonds stabilize protein structure. IR spectroscopy is one of the few experimental techniques that is sensitive to hydrogen bonding: amide I vibrations shift depending on the number of hydrogen bonds accepted by the oxygen atom or donated by the hydrogen atom in the amide unit.(68) A single hydrogen bond between the amide oxygen atom and a water molecule red shifts the amide I frequency by approximately 16 cm^{-1}.(19) The ability of the oxygen atom to accept multiple hydrogen bonds causes solvent-exposed residues to have strongly red-shifted transitions compared to residues locked into stable protein-protein hydrogen bonds. Understanding the origin of the H-bond induced shift is straightforward; vibrational frequencies are proportional to the square-root of the bond stretching constant and inversely proportional to the reduced mass of the oscillator. From a chemical-bonding perspective an H-bond reduces the π-orbital overlap CO imparting partial single bond C-O-H character to the C=O···H double bond. Similarly, the hydrogen atom can be thought of as effectively increasing the mass of the oxygen atom along with the reduced mass of the C=O bond, thus lowering the vibrational frequency. The relationship between H-bond strength and vibrational frequency has been measured for many systems. In general, the observed shift in vibrational frequency is proportional to the strength of the hydrogen bond. The empirical relationship between hydrogen bond distance and C=O frequency shift, $\delta\nu$ in cm^{-1}, for dipeptides in water is given by

$$\delta\nu = 30(r_{OH} - 2.6) \qquad 1$$

where r_{OH} is the distance between the carbonyl oxygen and the water hydrogen atom (in $\AA$).(68, 69) This simple relationship illustrates the sub-Angstrom sensitivity of IR spectroscopy.

2.3.2 Isotope labeling

A unique advantage of vibrational spectroscopy is its ability to spectroscopically isolate a single residue or small region of interest with a non-invasive isotope label. (2, 35, 66, 70) Isotope labeling aids in determining the local structure and dynamics of individual residues within a peptide or protein.(71) A single ^{13}C isotope label red shifts the local vibration by 35-40 cm^{-1}, largely decoupling the site from other residues. Since amide I peaks are 80-100 cm^{-1} broad, although the spectral shift afforded by a single ^{13}C is not necessarily sufficient to isolate the residue peak from the main band, a ^{18}O label gives the additional 25-35 cm^{-1} needed to resolve single residue peaks.(72)

Figure 7 provides a straightforward illustration of how the 2D IR spectrum of a 12-residue β-hairpin peptide TrpZip2 with an isotope label at the K8 position reveals conformational heterogeneity.(22) The K8 label exhibits two-peaks separated by the difference of a single hydrogen bond (16 cm^{-1}). The two peaks can be attributed to different β-turn conformations: K8-1 corresponds to a bulged-loop configuration in which the solvent-exposed carbonyl accepts two hydrogen bonds on average, whereas K8-2 is attributed to type-I' turn with a single K8-W4 internal hydrogen bond. Assignments are based on spectral simulations from molecular dynamics simulations and Markov state modeling. The minor structural difference between the two structures illustrates the exquisite structural resolution of vibrational spectroscopy. Conformational dynamics can be extracted from the 2D lineshapes, the

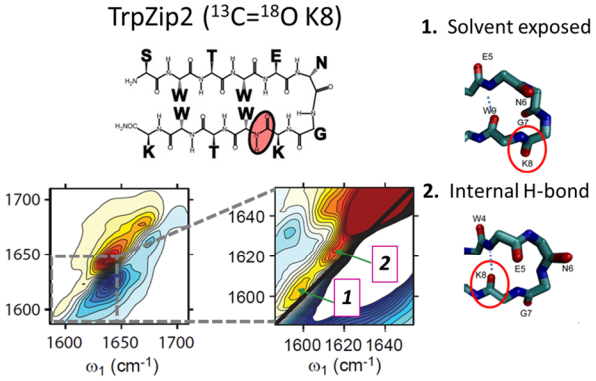

Figure 7 – Structure and 2D IR spectra of 12C=18O K8 TrpZip2. The two peaks observed in the 1600-1630 cm-1 region correspond to two different conformations of the β-turn as shown in the figure.

diagonal linewidth is related to the conformational disorder, and the anti-diagonal describes the timescale in which molecules sample the states (see Section 3). K8-2 has a broader diagonal lineshape, indicating that there is an increased amount of disorder in the solvent-exposed conformation compared to the more rigid K8-W4 hydrogen bond. The two β-turn conformations are predicted to interconvert on the microsecond timescale. This example highlights the structural and inherent ultrafast time resolution of nonlinear IR spectroscopy. In contrast, multidimensional NMR's millisecond time resolution results in an averaged spectrum which makes the two conformations indistinguishable.

2.5 Additional probes of protein structure: Amide II and side-chain vibrations

Here we provide a description of vibrations in the 1400-1800 cm^{-1}, which contain structural and dynamical information (33, 73).

2.5.1 Amide II vibrations

Amide II modes are characterized by large-amplitude C-N bond stretching combined with N-H bending motion. The large motion of the H atom causes the mode to red-shift from ~1550 cm^{-1} (amide II) to ~1450 cm^{-1} (amide II') upon deuteration. In comparison, amide I, which also includes some N-H bending motion, only red-shifts by ~10 cm^{-1} upon deuteration. Hydrogen/deuterium (H/D) exchange rates are dictated by the solvent-exposure and conformational flexibility of the protein. The ratio of amide II to amide II' peak intensities is proportional to the ratio of protonated to deuterated residues. However, amide II modes are thought to be less sensitive to structure. In addition, a variety of side-chain absorptions overlap the amide II band. Therefore, to date, relatively little experimental and theoretical work has been directed towards exploring the structural sensitivity of amide II modes. Section 5 describes a recent 2D IR method whereby combining amide I with amide II spectroscopy enables the direct measurement of the solvent exposure of individual secondary structures within proteins.

2.5.2 Other amide modes

In addition to amide I and amide II modes, there are a number of other backbone vibrations associated with the amide moiety: Amide III (1200-1400 cm^{-1}) consists of an in-phase combination of the C-C and C-N bond vibrations with contributions from the C=O in-plane bend. Amide IV (~630 cm^{-1}) consists mainly of C=O in-plane bending mode with small contributions from a C-C stretch and a CCN deformation. Amide A and amide B modes (~3170 and ~3300 cm^{-1}) are characterized by N-H stretches.(73) For various reasons these modes have not received much attention as spectroscopic markers of protein structure.

2.5.3 Side-chain absorptions

Side chains with carboxyl groups, aromatic rings, or the guanidinium moiety have absorptions near the amide I region.(74) These include arginine, asparagine, aspartic acid, glutamic acid, glutamine, tyrosine and tryptophan. Side chain vibrations can provide a useful probe of protein structure: carboxylic acids are sensitive to metal coordination and protonation state, and thus are sensitive to pH and solvent exposure. Some modes, particularly those involving large motions of hydrogen atoms are also sensitive to deuteration, and as such, the frequencies report on the solvent-exposure in hydrogen/deuterium exchange experiments. In contrast, aromatic ring modes are characterized by narrow lineshapes and center frequencies that remain insensitive to the environment. Non-covalent interactions, such as electrostatic interactions (i.e. salt bridges), or non-polar aromatic interactions are crucial to stabilizing protein structure, as well as providing surface contacts that drive protein recognition and binding. Side-chains inside binding pockets determine substrate selectivity and provide the essential interactions needed for catalysis. Side-chain spectra may provide useful insights into

structure, heterogeneity, folding or catalytic mechanisms, especially when combined with amide I or amide II spectroscopy.(62) The broadband pulses generated by ultrafast IR sources facilitate the simultaneous measurement of the backbone amide I and amide II bands as well as side-chains vibrations in the 3-7 um region (see Figure 9). Probing the differences between backbone and side-chain kinetics can be valuable in pinpointing the essential interactions that drive the backbone and side-chain ordering in protein folding.(75-78)

Table 1 - Common absorption frequencies observed near the amide I region for deuterated side-chains.

AA	Freq (cm-1)	Mode	pKa
Arg (NH2+)	1605	$\nu_{as}CN_3D5^+$	11.6-12.6
	1586	$\nu_sCN_3D5^+$	
Asp (COOH)	1713	vC=O	4.0-4.8
Asp (COO-)	1584	$\nu_{as}COO^-$	
	1404	ν_sCOO^-	
Asn	1648	vC=O	
Cys	1849	vSD	9.0-9.5
Glu (COOH)	1706	vC=O	4.4-4.6
Glu (COO-)	1567	$\nu_{as}COO^-$	
	1407	ν_sCOO^-	
Gln	1640	vC=O	
	1409	vC-N	
His	1600	vC=C (D_2^+)	6.0-7.0
	1569	vC=C (D)	
	1439	δCD_3, vCN	
Trp	1618	vC=C, vC-C	
	1455	δC-D, vC=C, vC=N	
	1382	δC-D, vC=C, vN-D	
Tyr (OH)	1615	vC=C, vC-D	9.8-10.4
	1590	vC=C	
	1515	vC=C, vC-D	
Tyr (O-)	1630	vC=C	
	1499	vC=C, δC-H	

Table 1 provides a reference of common side-chain absorptions in D_2O, along with the pKa of the individual side-chains. Note that the modes associated with C=O vibrations in carboxylic acids, Asp and Glu, blueshift by >100 cm^{-1} upon protonation.

2.5.4 Vibrational probes and unnatural amino acids

A relatively recent experiment strategy involves incorporating unnatural side-chains as localized probes of structure and dynamics.(79) Vibrational probes provide a minimally perturbative way of obtaining fine control over the spatial and vibrational localization of selected protein sites and take advantage of the quiet spectral region between amide I and the 3-um C-H, N-H or O-H stretching regions. Common probes include nitrile (-C≡N), azides (-N=N=N) thiocyanate (-S-C≡N) and C-D bonds.(80-82)

In the case of synthetic peptides, vibrational probes can be easily incorporated during solid-phase synthesis. Larger proteins can be produced by recombinant expression. One common method is to substitute an amino acid for a close structural analog in bacterial growing media.(83) However, substitution at multiple sites throughout the protein limits the site selectivity of this technique. New techniques incorporate unnatural amino acids at specific sites by creating a unique tRNA-codon that does not encode for any of the natural amino acids, along with a corresponding aminoacyl-tRNA synthetase.(84) To date, over thirty unnatural amino acids have been incorporated for spectroscopic and reactivity studies. It is also worth mentioning that the first ultrafast IR spectroscopy experiments were carried out on metal-bound C≡O ligands in heme proteins.(85) Narrow lineshapes, long vibrational lifetimes, and large absorption coefficients make metal carbonyls excellent vibrational probes for nonlinear IR spectroscopy. More specifically, vibrational dephasing, and lifetime, and 2D IR measurements of C≡O bound myoglobin have provided a localized view of the fluctuations at the active site of this heme protein.(86) Recently, ruthenium carbonyls, known as CO-releasing molecules, bound to the aromatic ring of solvent-exposed histidines have provided useful insight on interfacial water dynamics in globular proteins.(87)

Small vibrational probes have the advantage of being amenable to high-level theoretical modeling. (25) Spectroscopic observables, such as absorption lineshapes, frequency-frequency time correlation functions, and vibrational relaxation rates can be extracted from semi-classical simulations. (88) Computational simulations serve a twofold purpose: Firstly, simulated spectra can be compared to experiment. If there is good agreement, the simulations are used to extract an atomistic interpretation of the data. Secondly, simulations can be used to aid in the strategic development of experiments in order to maximize the amount of structural information extracted from the data. For example, simulations of label spectra can be

particularly informative for selecting unique labeling sites. Electrostatic maps, similar to those described in section 4, have been developed for a number of probes, including nitriles, azide and thiocyanate probes.(80, 81) Within the semi-classical fluctuating frequency approximation, the effects of the environment are captured by the fluctuating frequency of the site, which is calculated through molecular dynamics simulations combined with spectral maps to translate electrostatic potential trajectories into frequency trajectories. Once this trajectory is obtained, calculating linear and nonlinear spectra is relatively straightforward. Though short vibrational lifetimes limit 2D IR spectroscopy to picosecond timescales, it is becoming clear that fast fluctuations are important for biological function.

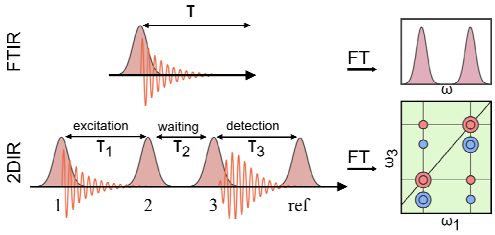

Figure 9 – Pulse sequences for FTIR and 2D IR spectroscopy

EXPERIMENTAL METHODS

Here we present an overview of the experimental infrared spectroscopy from a practical perspective. We introduce the basic background knowledge required to understand the experimental implementations of 2D IR spectroscopy, data processing methods, and interpretation of spectra. General sample preparation procedures are described, and the main experimental limitations and other practical considerations are outlined.

3.1 Infrared spectroscopy in one and two dimensions

Infrared absorption is a result of the interaction between an electromagnetic field and the oscillating molecular dipole moment. Within a classical picture, the oscillating electric field interacts with the atomic charges in the molecule and amplifies the vibrations that are in resonance with the frequency of the incoming light. In turn the oscillating charges emit an electromagnetic field out-of-phase with the incoming light, which cause a destructive interference that gives rise to the observed absorption peaks. Large atomic charges, give rise to strong oscillating dipole moments, which in turn produce strong absorption peaks.

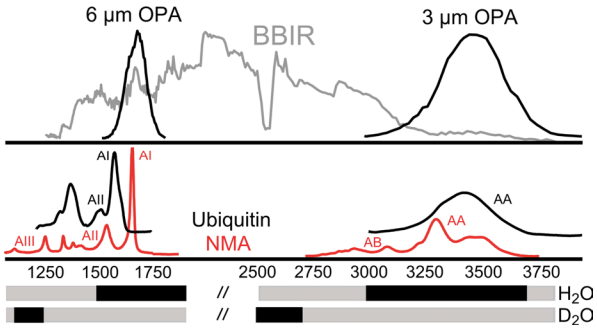

Figure 8 – Absorption spectra of ubiquitin in D2O and N-methylacetamide (NMA) in DMSO. The backbone amide bands (Amide I/II/III/A/AB) are indicated in the spectrum. The transparency windows for H2O and D2O are shown in grey for reference. The top panel shows typical output spectra for mid-IR optical parametric amplifiers in the 3 um region (~35 fs pulses) and 6 um regions (~90 fs pulses) along with a spectrum of a plasma-based broadband IR (BBIR) source.(3, 4)

Infrared absorption spectroscopy is generally carried out using incoherent light. If short pulses of coherent light are used, the oscillating *polarization* persists after the IR pulse passes through the sample, much like a bell ringing after being knocked by a hammer. One can measure the field radiated by this polarization directly by interfering it with a second pulse which arrives at varying time-delays. A Fourier transformation of this time-domain oscillation results in the absorption spectrum. The use of pairs of electric fields generated by an interferometer to measure absorption spectra is Fourier transform spectroscopy, the most commonly used method. This spectrum is the same as what one observes in a frequency-domain experiment, where one measures the change of light intensity through a sample as a function of the frequency of a monochromatic field. These are both one-dimensional experiments since there is one independent time or frequency variable (see Figure 8).

In the case of two-dimensional spectroscopy, one measures the changes to the absorption spectrum induced by excitation at an independent frequency. This also can be performed in the frequency-domain as a double resonance experiment with independent excitation and detection beams; however, here also Fourier transform methods are more commonly applied in one or both dimensions of the 2D spectrum. Following the initial short pulse excitation,

subsequent pulses can interact with the sample polarization in order to generate a *nonlinear* polarization.(5, 89, 90) Since multiple pulses can excite/de-excite the sample multiple times, the final nonlinear polarization radiates light carrying frequencies with new information about the molecular potential not present in linear absorption spectrum.

The pulse sequences used for 2D IR spectroscopy are shown in Figure 8. An interferogram is recorded at each τ_1 delay by overlapping the emitted signal with a reference beam, and the amplitude and phase of the emitted electric field are recovered through spectral interferometry. The frequency components of the emitted signal are represented directly in the detection axis of the spectrum, and a Fourier-transformation of the signal along the time delay between the first two pulses (τ_1) produces the excitation frequency (ω_1). The delay between the second and third pulses (τ_2) is referred to as the *waiting* or *population* time, and corresponds to the time between the excitation and detection events. The excitation and detection times are often referred to as *coherence* times, since the measured signal involves a coherent superposition of different states during these time periods.

The vibrational modes available for measurement in a 2D IR spectrum are dictated by the carrier frequency, duration, and spectral bandwidth of femtosecond infrared pulses used in the experiment. Commercially available laser sources currently generate pulses between 3-8 um (3600-1200 cm^{-1}) with pulse lengths between 50-100 femtoseconds with a corresponding bandwidth of 300 to 150 cm^{-1}. To illustrate the capabilities and limitations of these light sources, Figure 9 shows the IR absorption spectrum of ubiquitin and N-methylacetamide (NMA) along with spectra of typical OPA sources. New plasma-based broadband IR (BBIR) generation methods have been recently demonstrated which generate IR pulses that cover the entire vibrational spectrum from terahertz up to ~4000 cm^{-1}.(3, 91) Currently, low pulse energies prevent their use as sources for IR excitation, but provide a new avenue for simultaneously probing multiple transitions, following the interaction with narrowband pump.
Two-dimensional spectra are the most complete representation of the complex third-order IR response of the sample. One-dimensional projections of the response function are measured in

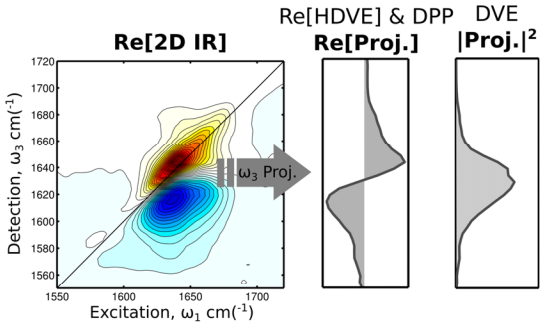

Figure 10 – Absorptive 2D IR, DPP, Real[HDVE] and DVE spectra of ubiquitin in D2O collected in the ZZYY polarization condition.

the form of *dispersed pump-probe* (DPP), *dispersed vibrational echo* (DVE), and *heterodyne-detected vibrational echo* (HDVE) signals. These signals contain much of the information represented in the 2D IR spectrum, however, projected onto a single frequency axis. These can be collected significantly faster than full 2D IR spectrum. Therefore some experiments, such as the T-jump experiments described below, utilize DPP and HDVE as a probe of kinetics, whereas 2D IR spectra are collected at few selected time delays in order to obtain a structural view of the transient response. Two-dimensional spectra are complex quantities, containing *absorptive* and *dispersive* features, however it is common to plot only the absorptive component the spectrum. Dispersed pump-probe, or simply referred to as *pump-probe*, spectra are absorptive and correspond to the projection of the real-part of the 2D IR spectrum onto the ω_3 (detection) axis (Figure 10). Complex HDVE spectra correspond to the projection of the complex 2D IR spectrum onto the detection axis, although it is common to plot only the real part of the HDVE spectrum (equivalent to the DPP spectrum). DPP and HDVE signals are measured interferometrically by combining the signal with a reference pulse. DVE spectra represent the power spectrum (absolute-value squared) of the complex HDVE spectrum and is measured without a reference.

In addition to excitation and detection frequencies, an important spectroscopic degree of freedom is the polarization of the four IR pulses—as it refers to the orientation of the pulse electric field vector in space rather than the collective oscillating dipole moment of the sample discussed above. Particular polarization conditions, in which one controls the polarization of the four excitation and detection fields can be used to suppress or enhance certain spectral features, and can be used to determine the

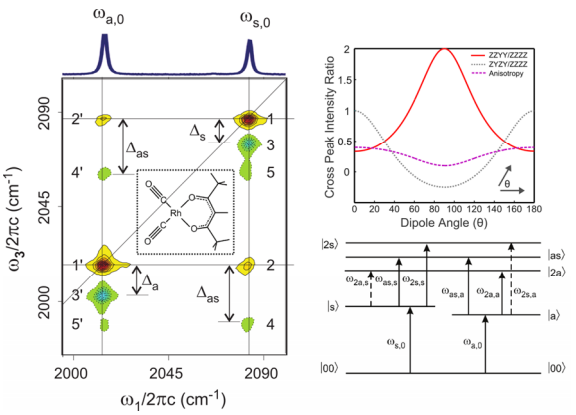

Figure 11 – (left) Infrared absorption and 2D IR spectrum of rhodium acetylacetonato dicarbonyl (RDC, structure shown for reference) in the terminal carbonyl stretching region. Solid and dashed contours represent positive and negative peaks respectively. (right) Simulated cross-peak intensity ratio between two polarization conditions as a function of the dipole angle between two coupled vibrations.

orientation of the different transitions within the molecule, providing additional structural information. The two main polarization conditions are parallel ZZZZ (all four pulses polarized parallel with respect to each other) and perpendicular ZZYY (first pair perpendicular to the others).(92) The parallel and perpendicular polarization conditions enhance cross peaks between vibrations with relative transitions dipoles oriented parallel and perpendicular respectively. The angle between two transition dipoles can be directly derived from the intensities of the peaks under different polarization conditions.

To illustrate the information content in a 2D IR spectrum, Figure 11 shows absorption and 2D IR spectra for the carbonyl vibrations of a model compound RDC in n-hexane.(93) The narrow peaks of the two carbonyl stretches provide a clean illustration of peak patterns typically observed in 2D IR spectra. The absorption spectrum (top) exhibits two narrow peaks at 2014 and 2084 cm^{-1} corresponding to the asymmetric (ω_a) and symmetric (ω_s) stretching vibrations respectively. In this context, it is important to clarify the distinction between *local mode*, *normal mode*, and *eigenstate*. Local modes are typically associated with a single-bond vibration, and, importantly, they form a basis onto which normal modes are described. In RDC, the two local modes correspond to the stretch of each individual terminal C≡O bond. In proteins the amide I local mode basis we use corresponds to a combination of C=O stretching and N-H bending vibrations of a peptide group. The local modes are often referred to as *oscillators* or *sites*. Normal modes involve linear combinations of local modes, in-and-out of phase oscillations of individual sites give each normal mode a unique character, and, since the two representations are connected by a linear transformation, the number of normal modes equals the number of local modes. In the case of RDC the two normal modes correspond to the in-phase and out-of-phase vibration of the C≡O oscillators: symmetric and asymmetric stretches respectively. Finally *eigenstates* represent the fully orthogonal set of anharmonic vibrations that are derived from the Hamiltonian for the vibrational system, and are commonly described as perturbative mixtures of the normal modes. For example, the symmetric, asymmetric, and combination state between the asymmetric and symmetric stretches are all eigenstates. Within the normal mode representation eigenstates are denoted by the number of quanta in each normal mode. For example the singly and doubly-excited symmetric stretches are denoted as $|s\rangle$ and $|2s\rangle$ respectively whereas the combination state is denoted as $|as\rangle$. Distinguishing between normal modes and eigenstates is important in nonlinear spectroscopy these distinctions arise from the very vibrational couplings that we hope to measure, and as spectra are sensitive to frequency differences and dipole amplitudes for transitions between eigenstates, not normal modes.

Two-dimensional infrared spectra can be interpreted as a two-dimensional map with one excitation (ω_1) and one detection axes (ω_3), which are represented on the abscissa and ordinate respectively. Diagonal peaks (1 and 1', Figure 11) correspond to excitation and detection at the same frequency. Peaks which appear immediately below the diagonal (3 and 3') correspond to excitation of $|0\rangle \rightarrow |1\rangle$ ($|0\rangle \rightarrow |s\rangle$, $|0\rangle \rightarrow |a\rangle$) transition and further excitation of the $|1\rangle \rightarrow |2\rangle$ ($|s\rangle \rightarrow |2s\rangle$ or $|a\rangle \rightarrow |2a\rangle$) transition during the detection time. The cross peak (2) represents $|0\rangle \rightarrow |s\rangle$ excitation vibration and stimulated emission from the $|a\rangle \rightarrow |0\rangle$. Analogously, peak 2' corresponds to excitation of the asymmetric stretch and detection of the symmetric stretch. Peaks 4 and 4' arise from transitions involving the two quanta combination

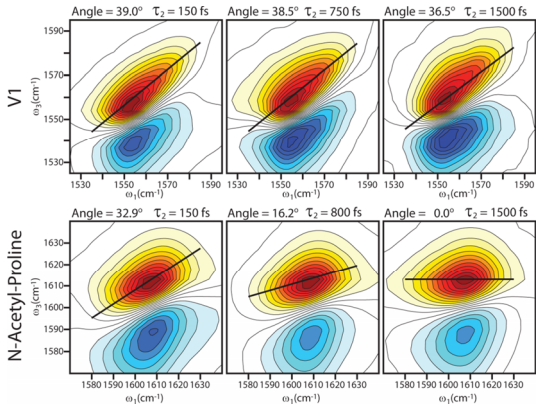

Figure 13 – Waiting time 2D IR spectra for a β-turn peptide GVGVP*GVG and the single amino acid N-Acetyl-Proline in D2O. The star denotes a 13C=18O label.

state. Finally, peaks 5 and 5' arise from excitation of the symmetric (asymmetric) stretch $|0\rangle \rightarrow |s\rangle$ ($|0\rangle \rightarrow |a\rangle$) and stimulated emission of the asymmetric $|2a\rangle \rightarrow |a\rangle$ (symmetric) stretch ($|2s\rangle \rightarrow |s\rangle$). One and two quanta energy levels of the molecule can be simply read-out from the peak positions. Anharmonicities and coupling constants can be determined by spectral modeling. Changes in the intensities of the cross peaks with respect to the polarization of the excitation and detection pulses is determined by the relative angle of the transition dipoles between the two vibrations. To illustrate this example, Figure 11 shows the calculated ratio of cross peaks using three different polarization geometries (ZZYY/ZZZZ, ZYZY/ZZZZ) as a function of the transition dipole angle between two coupled oscillators.

Structural information is extracted from the position and intensities of the peaks, whereas dynamics are extracted from the peak shapes, more specifically, the ellipticity of the peaks as well as the change in ellipticity with waiting time. (94) Figure 12 shows an example of changes in lineshape along the waiting time: a diagonal elongation at early waiting times, as each conformation within the ensemble has a different transition frequency such that there is a large correlation between the excitation and detection frequencies. Node lines rotate as a function of waiting time, as dynamical effects cause the correlation to decrease. The rate at which frequency correlation is lost can be directly related to the frequency fluctuations of the molecules in the sample, referred to as *spectral diffusion*. Elongation along the diagonal is referred to as *inhomogeneous* broadening while elongation along the antidiagonal is called *homogeneous* broadening. The example shows how residues that are locked into rigid intramolecular hydrogen-bonding configurations have slower dynamics (V1 peak), which lead to smaller frequency fluctuations, and a slower rotation of the nodal line, whereas residues exposed to the solvent (N-Acetyl-Proline) quickly lose the correlation between excitation and detection.

3.2. Transient temperature-jump 2D IR spectroscopy

While waiting-time 2D IR spectroscopy offers structural information with sub-picosecond time-resolution, a powerful experimental toolbox is developed by combining 2D IR spectroscopy with non-equilibrium methods in order to study longer timescale protein folding, dynamics, and function. Since IR spectroscopy produces ensemble measurements, a rapid trigger event is required to synchronize the time evolution of the ensemble. The time-resolution of the pump along with the timescales accessible by the probe often dictate the type of dynamics that can be studied. The "probe" method, such as 2D IR, must have high time resolution, high structure-sensitivity, but most importantly, the reaction coordinate must project favorably onto the selected spectroscopic coordinate. In the case of proteins, isotope labels often provide a method to select a set of spectroscopic variables of interest.

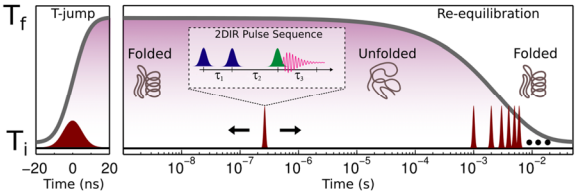

Figure 12 – Schematic representation of the t-jump pulse sequences and temperature profile. The delay between the T-jump pulse and the 2D IR pulse sequence is electronically controlled. The 1 ms repetition rate of the 2D IR pulses enables us to use subsequent pulses in the train to collect time delays from 1 ms to 50 ms.

Common triggers include the following(95) (approximate time resolution indicated in parenthesis): 1. Electronic excitation (10-100 fs) to probe ultrafast excited-state dynamics and photoinduced reactions, or to gate photorelease of caged compounds.(96) 2. Temperature-jump spectroscopy (1-10 ns) to study thermal

denaturation, protein folding, association, non-equilibrium dynamics.(97, 98) 3. Pressure-jump (1 us) to study pressure-induced conformational changes of proteins such as partial denaturation.(99) 4. pH-jump methods (100 us) to study pH-induced conformational dynamics or rapid protonation of different protein sites.(100) 5. Stopped-flow rapid mixing techniques (1-3 ms) to probe protein unfolding and conformational changes induced by denaturants or mixed solvents.(101) Of these triggered methods, only electronic excitation and temperature-jump triggering have been demonstrated within the context of ultrafast 2D IR spectroscopy.(96, 102, 103) In principle any of these techniques can be combined with a 2D IR probe; one practical limitation, however, is related the weak nonlinear signals observed in biological samples which inevitably lead to long data collection times. Triggering methods must therefore cause large changes in the signal and have repetition rates commensurate with the timescales of interest so as to maximize the duty-time of the measurement.
Here we focus on nonequilibrium 2D IR spectroscopy initiated by a temperature jump. Figure 13 shows a schematic representation of the temperature profile of the sample following a T-jump. A 2D IR pulse sequence is able to probe structural rearrangements from the nanosecond to millisecond timescales. Protein dynamics occur on various timescales, but in general nanosecond to microsecond response times correspond to barrierless rearrangements whereas transitions occurring from 100us - 1ms are attributed to barrier-crossing events. Delay times are limited by the repetition rate of the T-jump laser, and the relaxation of the sample. Within our current implementation the maximum delay is 50 ms.(97)

A T-jump pulse must carry sufficient energy to raise the temperature of the solvent by a few degrees. In our implementation 20 mJ pulses centered around 2 μm are generated by a frequency-doubled Nd:YAG laser with a repetition rate of 20 Hz, coupled to an optical parametric oscillator (OPO). The T-jump laser is electronically timed to the 1 kHz femtosecond laser source. The shortest delay is given by the T-jump pulse width (10 ns) and longest delay by the repetition rate of the T-jump laser to (50 ms). T-jump pulses are resonant with the overtone of the OD stretch in the solvent. Following the temperature jump the solvent returns to equilibrium as heat diffuses away from the focus region. The temperature relaxation profile is described by a stretched exponential with a time constant of approximately 3 ms. In order to achieve uniform heating of the focus region, the T-jump pulses are focused to a ~1 mm spot at the sample and only about 10% of the light is absorbed. Interferometric measurements are particularly demanding since the phase of all beams must be kept constant before and after the trigger event. Within our implementation, all pulses are focused onto the interaction region, and are thus subject to the same changes in solvent absorption and index of refraction induced by the T-jump. Depending on the initial temperature, a 10 degree temperature jump causes transmission changes of approximately 5% in the amide I region which magnifies the nonlinear signal. The spectrum of the reference pulse is collected along with the signal with every laser shot, and the signal is corrected to account for transmission changes. Corrections are necessary in order to interpret difference 2D IR spectra as well as to obtain undistorted sample kinetics.(104) In contrast to absorption spectroscopy, an important advantage of nonlinear infrared spectroscopy is that solvent signals are greatly suppressed, thus it is not necessary to separately measure a solvent background in order to isolate the solute response.

A 2D IR spectrum represents the most complete characterization of the third-order vibrational response, but for a variety of reasons, acquisition of a 2D IR spectrum is relatively slow. Techniques that reduce data collection times are better suited to finely-sample the time delay following the T-jump in order to measure kinetics. Three frequency-resolved nonlinear measurements can be carried out with the 2D IR spectrometer: dispersed pump-probe (DPP), dispersed vibrational echo (DVE), and heterodyne-detected dispersed vibrational echo (HDVE) spectroscopy.(105) DPP measures transmission changes of a probe pulse following an IR excitation by a "pump" pulse. The DPP signal measured is equivalent to the 2D IR signal integrated over the excitation axis, therefore DPP cannot clearly separate contributions from diagonal and off-diagonal features of a 2D IR spectrum. Similarly, DVE and HDVE represent the four-wave mixing signal emitted by the sample following the interaction with three IR pulses at fixed time delays. Similar information can be extracted from DVE and HDVE measurements, but HDVE, a more demanding phase-sensitive measurement, is sensitive

to changes in the amplitude and phase of the reference induced by the T-jump pulse. Unlike DPP, DVE and HDVE are background-free measurements and thus offer improved signal to noise and sensitivity, and DPP signals can be recovered from an HDVE measurement. Fourier-transformation of a series of HDVE spectra at various delays between pulses 1 and 2 produces a 2D IR spectrum. Since HDVE spectra can be collected ~250-500 times faster than equivalent 2D IR spectra, HDVE is used for collecting kinetic data and 2D IR for extracting structural information at selected time delays.

3.3 Sample preparation

Compared to standard biophysical techniques, the sample preparation procedures for 2D IR spectroscopy are relatively straightforward. Here we describe the typical sample preparation for peptides and proteins. For amide I spectroscopy, protein samples are often dissolved in D_2O and gently heated for a few hours in order to partially denature the protein and exchange the labile protons with deuterons. The sample is then lyophilized and re-dissolved in pure D_2O. Solutions are placed between two CaF_2 windows with a spacer. The transparency of D_2O limits the pathlength to 50 μm. In principle the minimum volume needed to collect a spectrum is 10 pL, but in practice a volume of at least 200 nL is required for each sample. Since the D_2O bend vibration absorbs strongly below 1500 cm^{-1} (see Figure 9), the path length must be kept to a minimum, therefore sample concentrations must be relatively high. The residual H_2O bend vibration overlaps with the amide I band, therefore the residual H_2O contents should be kept to less than 3%. In proteins the amide I absorbance is proportional to the number of residues, if we consider the average mass of a residue to be 120 g/mol, the approximate concentration needed for 2D IR is 80 mM/residue (~10 mg/mL) for most proteins. The concentration can be lower when absorption bands are sharp, such as in the case of aggregates or membrane proteins. This sample concentration regime is comparable to that used in NMR spectroscopy, except that sample volumes used in 2D IR are significantly smaller. Carboxylic acid buffers, such as acetate, or compounds that contain a carbonyl group, such as urea, must be avoided due to the absorption overlap with the amide I band. In the case of synthetic peptides, trifluoroacetic acid left from solid-state synthesis and subsequent purification steps must be thoroughly removed by multiple lyophilizations from acidic solution.

3.4 Isotope Labeling

Isotope labels spectroscopically isolate single residues by red shifting their transition frequencies from the main band. A single ^{13}C replacement is not always sufficient to vibrationally isolate a residue, particularly in larger proteins where the ~1% ^{13}C natural abundance increases the probability of finding other naturally-occurring ^{13}C sites within the backbone. In addition, broad lineshapes will hide the residue peak underneath the main band. Proline residues also contribute intensity in the 1630 cm^{-1} region.(11) Because of these contributions, often a $^{13}C=^{18}O$ double-label is necessary to spectrally isolate a region of interest. ^{13}C-labeled amino-acids are commercially available but ^{18}O variants have to be synthesized. Recently a new method was developed for synthesizing $^{13}C,^{18}O$-double-labeled N-(9-fluorenylmethoxycarbonyl) (FMOC)-amino acids with high levels of ^{18}O incorporation for amino acids without acid-labile side-chain protecting groups (Gly, Ala, Val, Ile, Leu, Phe, Trp, and Pro).(106) The acid hydrolysis reaction is carried out by refluxing the FMOC-amino acid for periods between 3-30 h in acidic mixtures of $H_2{}^{18}O$ and organic solvents. The organic solvent must dissolve the FMOC-amino acid but not act as a source of ^{16}O. The final product is recovered by lyophilization of the reaction mixture. Enrichment ratios are typically >90%, thus the recovered product may be used without further purification.

THEORY

An important feature of the IR spectroscopy of biological systems is that the field rests on a well-developed theoretical basis, allowing in many cases for comparison of experimental data with molecular-level models such as molecular dynamics (MD) simulations. In this way, infrared spectra can be used as a direct probe into the structure and dynamics of proteins. In this section, we outline the basic features of the theory developed for amide I vibrational spectroscopy applied to IR absorption spectroscopy and 2D IR.

From an experimental point of view, the information content of protein 2D IR spectra comes in two forms: local structural information such as hydrogen

bonding and solvent exposure is provided by site-specific isotope labels, while global secondary structure content is encoded in delocalized excitonic bands such as the β-sheet $\upsilon_{//}$ and $\upsilon_{\perp}$ peaks. Theoretical models for amide I spectroscopy must likewise account both for local effects, particularly the influence of an oscillator's local electrostatic environment on its vibrational frequency, and global effects such as the delocalization of amide excitations—excitons—due to site-to-site coupling. We begin our discussion by focusing on local parameters, particularly relevant to isotope labeling experiments, before introducing the exciton treatment of globally delocalized vibrational modes and computational methods for estimating the relevant parameters for amide I vibrations. Finally, after providing a brief explanation of the theory behind 2D IR spectroscopy itself, we close with a short description of practical approaches to simple 2D IR modeling for amide I spectra.

4.1 Oscillator Site Energies

At the simplest level, a theoretical description of amide I spectra must describe the frequency of each vibrational oscillator due to its local environment, also known as the site energy. One can understand these effects intuitively in terms of the influence of local interactions on bond strengths within the amide unit. For example, donation of hydrogen bonds from water molecules to the oxygen of the amide unit stabilizes the C=O bond, lowering its vibrational frequency and causing a red-shift of the corresponding amide I absorption peak. Therefore this site energy can report on the local hydrogen bonding configuration in which the peptide resides. On a computational level, the correlation of local structure and vibrational frequency can be expressed in terms of electrostatic variables (potential, field, gradient, etc.), which is the basis for most of the site energy maps which will be discussed in more detail below.(15, 19, 20, 23, 25, 27-30, 107, 108) Although the details of these methods vary, all seek to predict IR spectroscopic features based on the local electrostatic environment of each oscillator in a given structure (for example from an MD simulation).

Within a protein structure, these electrostatic factors produce a broad distribution of site energies for individual amide oscillators in the protein backbone, which reports on protein-protein and protein-solvent interactions. The frequency shifts of an amide I vibration relative to an isolated oscillator varies from by 40 cm^{-1} depending on the nature of these interactions. Experimentally, the shift of a particular oscillator can be accessed most easily by introducing ^{13}C and/or ^{18}O isotope labels into specific amide I carbonyl groups on the protein backbone. The higher reduced mass of isotope labeled units results in a red-shift of the corresponding amide I vibration of ~40-75 cm^{-1}, decoupling the selected oscillator from other amide I vibrations and moving its absorption into a spectral window that allows for direct observation of the site energy shift. These isotope labeled samples are particularly useful in 2D IR spectroscopy since peak shapes provide insight not only into the total distribution of transition frequencies, but also into the timescales for spectral diffusion and dephasing. A clear understanding of the relationship between local electrostatic effects and oscillator frequency thus provides a direct link between spectral characteristics and molecular structure and dynamics, such as hydrogen bond configurations and lifetimes.

4.2 Delocalized Vibrations and Excitons

Although local electrostatic effects are sufficient to describe the vibrational frequencies of individual sites, interactions between neighboring residues generally complicate the picture by inducing delocalization of vibrational modes across multiple residues. For this reason, beyond electrostatics, the central feature of most theoretical descriptions of amide I and II vibrations is the concept of the *exciton*—a delocalized excited state produced by the interaction of many individual sites. The concept of the exciton was originally introduced in the context of solid state physics,(36) but has found applications in modeling a wide variety of biophysical processes, including the visible absorption spectra of photosynthetic systems(109) and ultraviolet circular dichroism of protein backbones.(110)

The vibrational excitons observed in amide I and II spectra behave in many respects like a network of oscillators—or sites—coupled together into a single system—much as in a bedspring a large number of individual springs are linked together to produce a single mattress. Just as deformation of one spring in a mattress affects neighboring springs, vibrational coupling between individual amide units in a protein causes excitation of any given amide vibration to induce vibrations in adjacent units, a process known

as *delocalization*. The extent to which vibrational motion is delocalized over a system depends on the number of sites, the site-to-site coupling strength, and the variation of site energies. For two coupled vibrations with site energy ε_1 and ε_2 and coupling strength J, the resulting exciton states have energies $E_{\pm} = \frac{1}{2}(\epsilon_1+\epsilon_2)^2 \pm \frac{1}{2}\sqrt{(\epsilon_1-\epsilon_2)^2+4J^2}$. Similar site energies produce the greatest delocalization since it allows neighboring sites to oscillate in-phase with each other without interference. The case of amide I vibrations proves to be complicated because the variation in vibrational coupling, typically between -10 and +10 cm^{-1}, is of a similar scale to the variation in site energy, 30-40 cm^{-1}.

These concepts were first applied to amide I and II vibrations by Miyazawa who demonstrated that the characteristic amide I IR absorption spectra observed for α-helix and β-sheet structures could be well explained by a normal mode analysis in which each carbonyl group acts as a single oscillator vibrationally coupled to its nearest neighbors and hydrogen-bonding partners.(111) As described by Miyazawa, the similarity of carbonyl group site energies across the protein backbone, together with the regularly repeating patterns of α-helices and β-sheets, leads to strongly delocalized excited states. These ideas were developed by many over the years, notably Krimm(112) who introduced a transition dipole coupling (TDC) model for predicting inter-site coupling constants from molecular structures, and Torii and Tasumi(64) who more clearly expressed the concept of an amide I subspace adiabatically separated from the remaining protein vibrations.

4.3 Amide I Hamiltonian

These methods were adapted for nonlinear spectroscopy by Hamm, Lim, and Hochstrasser who recast the earlier force constant matrix calculations into a quantum Hamiltonian including doubly excited states.(69) They described the exciton states as eigenstates of a Hamiltonian operator that describes the interactions between *N* sites

$$\hat{H} = \sum_{n=1}^{N} \varepsilon_n |n\rangle\langle n| + \sum_{m,n=1}^{N} J_{mn} |m\rangle\langle n| + \sum_{m,n=1}^{N} (\varepsilon_m + \varepsilon_n - \Delta\delta_{mn}) |mn\rangle\langle mn| + \sum_{m,n=1}^{N} \sum_{\substack{j,k=1 \\ (m,n)\neq(j,k)}}^{N} J_{mn,jk} |mn\rangle\langle jk| \quad (2)$$

In this expression, ε_n and J_{mn} represent the site energies and coupling constants between singly-excited states. The second set of terms represent doubly-excited states in which a single oscillator is excited twice ($m = n$) or two different oscillators are each excited ($m \neq n$). The *anharmonicity* value Δ (~16 cm^{-1} for amide I) is the difference in the absorption frequency for the fundamental ($0 \rightarrow 1$) transition compared with its overtone ($1 \rightarrow 2$). This value is particularly important for 2D IR spectroscopy since in a perfectly harmonic system (with $\Delta = 0$), the 2D IR signal vanishes due to interference between the fundamental and overtone transitions. Note that coupling between one- and two-quantum states is neglected, so that the states are separated into zero-, one-, and two-quantum subspaces. The resulting block-diagonal amide I Hamiltonian is illustrated graphically in Figure 14 in which site energies occur along the diagonal and site-to-site couplings appear off-diagonal within the one- and two-exciton blocks. The eigenstates of the system are obtained by numerically diagonalizing the resulting matrix, providing absorption frequency values, dipole moments, and oscillator strengths.

The necessary input for the calculation is the site energies and coupling constants for each state. In structure-based modeling, the site energies are dependent on the local structure about a site, and the coupling will depend on the configuration of two sites with respect to one another, for instance a dipole-dipole interaction. These assignments are aided by the structure/spectroscopy maps described below. Usually, a harmonic approximation is assumed to obtain the two-exciton coupling constants from the one-exciton energies ε_n and coupling constants J_{mn}.

Having outlined the principles of this mixed quantum-classical approach, it is worth stating the assumptions that go into such modeling. The principle assumption is an adiabatic separation of the amide I vibrational motions and the dynamics of the protein. The semi-classical approximation then states that essentially all large amplitude protein motions such as conformational fluctuations can be treated classically via static protein structures or MD simulations, while the higher frequency vibrations such as amide I which are interrogated by the infrared radiation must be treated quantum-mechanically. The interaction between the amide I vibrations and protein structure is handled with the mapping procedures described below. For dynamic

calculations, the classical motion of the protein then determines the parameters of a quantum excitonic Hamiltonian which describes the high-frequency amide I dynamics and the interaction of the system with the perturbing electromagnetic fields.

4.4 Structure-Based Calculations

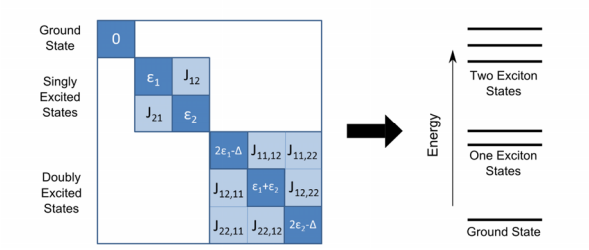

Figure 14 – Schematic representation of the excitonic Hamiltonian matrix and energy levels for a 2-oscillator system.

A number of schemes have been constructed to obtain the one-exciton site energies and coupling constants directly from molecular structures, e.g. MD simulations. For small systems, electronic structure methods such as density functional theory can be applied directly to obtain energies, couplings, and transition dipole moments. However, such calculations are limited, at least for the present, to systems no larger than a few amino acids and including at most a few solvent molecules. A variety of parameterized *maps* have been constructed to predict frequencies and couplings in larger systems without the need for electronic structure calculations. The maps take advantage of the correlation that exists between the local structure and electrostatic environment of the amide bond and the frequencies of the associated vibrational modes in order to greatly reduce the computational expense of calculating site frequencies. As an example, the left frame of Figure 15 shows the calculated site energy shift induced by a +0.5 point charge located around the amide unit using one such map.(14) The right frame shows for comparison the correlation between hydrogen bond donor/acceptor distance and site energy shift for a for a small β-turn peptide using another map.(15) Although the maps differ in their details, both reflect the observation that amide I site energies red-shift on hydrogen bond formation. These electrostatic maps are primarily parameterized through electronic structure calculations and, as illustrated in Figure 16, generally perform well in reproducing the influence of the local environment on the transition frequencies of small model systems.

Site energy maps may consist of two distinct contributions:

- *Electrostatic maps*: A non-specific electrostatic frequency shift induced by the electrostatic potential, field, or gradient evaluated at the amide bond. This contribution arises from the net electrostatic effects of *all* atoms in the simulation (excluding the nearest neighbors), regardless of their identity, and accounts for effects such as hydrogen-bonding and solvent-induced shifts. One such example is the map developed by Skinner and coworkers,(14) where the site frequency (ω in cm^{-1}) is given by

$$\omega = 1684 + 7729E_C - 3576E_N$$

  where E_C and E_N represent the electric field projection along the C=O bond (in atomic units) at the C and N positions. Other common map positions include the O, H and Cα atoms. The frequency shift from a point charge at different positions in the x-y, calculated using the above map, is depicted in Figure 15.
- *Nearest-neighbor shifts:* A specific through-bond frequency shift determined by the dihedral (ψ, φ) angle pair of adjacent amino acids in peptides. This contribution reflects the fact that amide I vibrations are in reality not completely decoupled from other vibrations. It is a function only of the backbone conformation of the polypeptide chain, and accounts for site energy shifts due to nearest-neighbor effects such as steric strain between adjacent residues. Figure 16 illustrates one such nearest-neighbor map.(108)

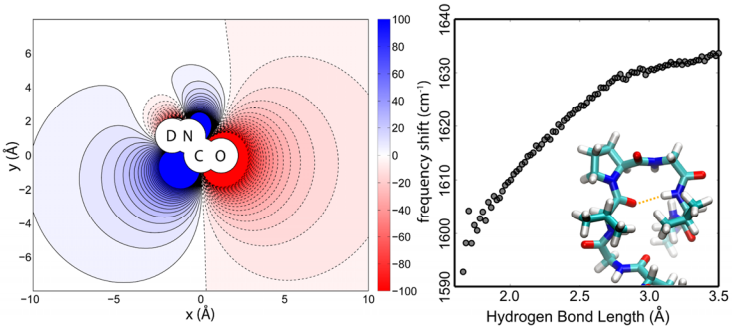

Figure 15 – (left) Frequency shift induced by a 0.5 a.u. test charge around the amide unit calculated using the map of Skinner et al.(14) (right) Correlation between V1-V4 hydrogen bond length and V1 frequency for the β-turn peptide GVGV1PGV4G. Frequencies were calculated using the parameterization of Jansen and Knoester(15, 16) (see Table 2).

Although existing site energy maps are for the most part all based on some combination of these two contributions, they differ considerably in their details. In particular, there is significant variation in the choice of electrostatic parameters considered (generally, the potential, field, or gradient) and the sampling points around the amide bond (usually evaluated at the N, H, C, and O atoms, possibly with the addition of other adjacent points). Various methods have likewise been proposed to predict the site-to-site coupling constants J_{mn}, including both through-bond (nearest-neighbor) and through-space effects.(8)

- *Transition dipole coupling* (TDC) treats through-space coupling by treating each oscillator as a simple dipole vector which interacts with adjacent dipoles through a coupling factor similar to the geometric factor encountered in FRET studies.
- *Transition charge coupling* (TCC) is an improvement on the TDC model in which a transition charge is assigned to each atom of the amide bond (or other suitably chosen nearby points in space) and an interaction energy is calculated between each atom. The TCC method is accurate to somewhat shorter distances than TDC.
- *Nearest-neighbor coupling* or through-bond coupling is often accounted for in much the same way as the nearest neighbor frequency shift used for site energy calculations (see Figure 16). A set of electronic structure calculations are used to parameterize a dihedral angle map for coupling between adjacent residues, accounting for steric effects and through-bond interaction.

Table 2 provides a comparison of the essential features of selected amide vibration maps published

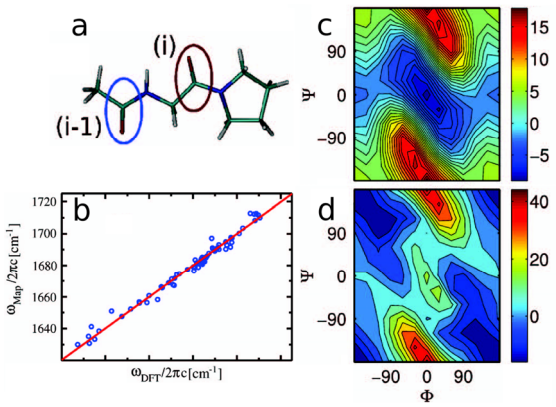

Figure 16 – Electrostatic maps corresponding to a two-residue fragment. (a) structure of the Pro-Gly fragment the two backbone C=O. (b) Correlation plot between frequencies computed through DFT calculations and the electrostatic map. (c) Coupling constants (cm-1) as a function of phi and psi angles (d) Frequency shift (cm-1) of site i-1 due to the perturbation by site i. Adapted from [Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. J Chem Phys 2011, 135.]

in recent years. A comparison of 2D IR experimental data with spectra calculated with various maps is presented in Figure 18.

4.5 Nonlinear Polarization and Signal

Thus far our discussion has focused on a microscopic picture of vibrational modes within an individual protein. Nonlinear spectroscopy, on the other hand, is usually described in terms of a macroscopic *nonlinear polarization* $\boldsymbol{P}(\boldsymbol{\mathcal{R}},t)$ induced in a sample by a series of perturbing electromagnetic fields (laser pulses).(89, 113, 114) This induced polarization in turn gives rise to an oscillating electromagnetic field which is emitted from the sample in the same wavevector-matched directions as the polarization itself and can be measured

experimentally. The connection between the two pictures is given by the relation

$$\mathbf{E}(\mathcal{R},t) \mu \mathbf{P}(\mathcal{R},t) = \langle \mu(\mathcal{R},t) \rangle \quad 3$$

i.e. the *macroscopic* polarization, and hence the emitted field, is proportional to the expectation value of the *microscopic* dipole moment evaluated at different points in the sample.[1](89, 113)

Under the semiclassical approximation that we use for amide I spectroscopy, the resulting expressions for the nonlinear signal from the excitonic system can be written as the sum of two terms. These are known as the *rephasing* and *non-rephasing* signals, and they differ by the phase of oscillating coherences in the excitation period τ_1.(115) If we assume that the perturbing field disturbs a system with a static Hamiltonian, we can express the signal in terms of the corresponding eigenstates as

$$\begin{aligned}
&E^{R}_{ijkl}(\tau_1,\tau_2,\tau_3) \mu \\
&Im\sum_{ab}\Big\{\langle \tilde{\mu}_i^{0a}\mu_j^{0b}\tilde{\mu}_k^{a0}\mu_l^{b0}\rangle e^{-i\omega_{0a}\tau_1}e^{-i\omega_{ba}\tau_2}e^{-i\omega_{b0}\tau_3} \\
&+\langle \tilde{\mu}_i^{0a}\tilde{\mu}_j^{a0}\mu_k^{0b}\mu_l^{b0}\rangle e^{-i\omega_{0a}\tau_1}e^{-i\omega_{b0}\tau_3}\Big\} \\
&-Im\sum_{abc}\Big\{\langle \tilde{\mu}_i^{0a}\mu_j^{0b}\mu_k^{bc}\mu_l^{ca}\rangle e^{-i\omega_{0a}\tau_1}e^{-i\omega_{ba}\tau_2}e^{-i\omega_{ca}\tau_3}\Big\} \quad 4 \\
&E^{NR}_{ijkl}(\tau_1,\tau_2,\tau_3) \mu \\
&Im\sum_{ab}\Big\{\langle \mu_i^{0a}\tilde{\mu}_j^{0b}\tilde{\mu}_k^{b0}\mu_l^{a0}\rangle e^{-i\omega_{a0}\tau_1}e^{-i\omega_{ab}\tau_2}e^{-i\omega_{a0}\tau_3} \\
&+\langle \mu_i^{0a}\mu_j^{a0}\mu_k^{0b}\mu_l^{b0}\rangle e^{-i\omega_{a0}\tau_1}e^{-i\omega_{b0}\tau_3}\Big\} \\
&-Im\sum_{abc}\Big\{\langle \mu_i^{0a}\tilde{\mu}_j^{0b}\mu_k^{ac}\mu_l^{cb}\rangle e^{-i\omega_{a0}\tau_1}e^{-i\omega_{ab}\tau_2}e^{-i\omega_{cb}\tau_3}\Big\}
\end{aligned}$$

In these expressions, the three sequential time intervals τ_1 τ_2 τ_3 are shown in Figure 8. The indices *a* and *b* run over all possible one-quantum eigenstates, while the third index *c* is summed over all possible two-quantum eigenstates. The tilde over some matrix elements indicates a complex conjugate. The subscript indices *i*, *j*, *k*, and *l* indicate the polarization direction of the three laser pulses (*i*, *j*, and *k*) or of the signal analyzer (*l*) in the laboratory frame; the angular brackets indicate an ensemble average over the orientation of the molecular reference frame.(113, 116) A practical guide for simulating 2D IR spectra is provided in section 4.7.

These expressions are easily rationalized in light of the transition dipole elements at the start of each term: the first three elements, μ_i, μ_j, and μ_k, indicate absorption or emission events induced by the three excitation pulses; the final element μ_l indicates the final emission event which is detected in the experiment. The oscillating exponentials corresponding to the three time intervals τ_1, τ_2, and τ_3, indicate the creation of *coherences* (oscillating expectation values arising from the superposition of two quantum states) or *populations* (static contributions to observable values reflecting the occupation of a single quantum state). In both the rephasing and non-rephasing response functions, coherences give rise to fast oscillations in the polarization as a function of the initial and final time delays τ_1 and τ_3, but much slower oscillations as a function of τ_2 since the difference frequency $\omega_{ab} \approx 0$ is for two nearly degenerate one-quantum states.

4.6 Frequency-Domain Spectra

The expressions we have discussed so far describe the polarization (or signal electric field) as a function of time. What we wish to obtain in the experiment is a two-dimensional frequency-frequency correlation spectrum linking the "absorption" frequency during τ_1 with the "emission" frequency during τ_3. Experimentally, one frequency-dimension can be obtained directly by dispersing the emitted field through a monochromator onto an array detector, essentially giving a Fourier transform of the field along τ_3 directly. This gives a frequency-domain spectrum as a function of the emission frequency (ω_3) for a single value of τ_1, but provides no information on the interaction events during τ_1. Referring back to Equation 4, note that although the oscillation frequencies along τ_1 do not directly affect the (integrated) intensity of the emitted signal, they do alter the phase. This dependence can be monitored experimentally by *heterodyne detection* in which the emitted signal is overlapped with a reference pulse (the local oscillator) with a fixed phase; the interference pattern between the two pulses is then recorded as a function of both delay time (τ_1) and emission frequency (ω_3). Finally, the collected data is post-processed with a numerical Fourier transformation along the τ_1 axis to obtain a two-dimensional frequency-frequency plot in the two

[1] Although formally, the polarization $\boldsymbol{P}(\boldsymbol{\mathcal{R}},t)$ acts as a source input in Maxwell's equations, to a good approximation, the electric field looks simply like a phase-shifted $\boldsymbol{P}(\boldsymbol{\mathcal{R}},t)$, so that for most purposes it suffices to treat the two as equivalent.

variables ω_1 and ω_3. Note that this gives peaks in the rephasing spectrum at negative frequencies in ω_1, so the spectrum must be inverted along the w1 axis as $-\omega_1 \rightarrow \omega_1$ for comparison with the non-rephasing spectrum.(117)

The expressions above (Equation 4) describe non-linear response for a system with a discrete number of energy levels and no population relaxation, giving rise to a discrete set of delta functions in the frequency-domain spectrum. For a real system, of course, neither of these assumptions is strictly appropriate. Population relaxation gives rise to an overall decay of the signal intensity, without directly affecting the phase, as excited oscillators return to the ground state, while coupling to low-frequency—phonon—modes gives rise to pure dephasing of the signal as the phase is gradually scrambled due to interference between closely-spaced modes. In the frequency domain, dephasing and relaxation processes together produce *phase twisted* two-dimensional spectra in which absorptive (amplitude modulation) features along ω_1 are mixed with dispersive (phase modulation) features, resulting in a complicated lineshape which is often difficult to interpret visually.(114, 115) These effects are illustrated in the upper panel of Figure 17 where experimental and calculated rephasing and non-rephasing spectra are plotted for the protein ubiquitin. Note the strong "smearing" of the spectrum along the diagonal in rephasing and against the diagonal in the non-rephasing spectrum. This distortion can be removed by adding the rephasing and non-rephasing signals together to obtain the *pure absorptive* spectrum plotted on the left. Note that only in the pure absorptive spectrum can the β-sheet cross peak ridge along $\omega_3 \approx 1680$ cm^{-1} region be clearly distinguished from the dispersive features of either rephasing or non-rephasing individually.

4.7 Static Simulations

Finally, by combining the 2D IR calculations just described with a parametrized excitonic Hamiltonian as described in section 4.3 above, one can directly link structural data (MD simulations or NMR/X-ray structures) to spectroscopic observables. In the simplest case, one can construct a 2D IR spectrum corresponding to a single protein structure by inserting excitonic eigenstate energies and transition dipole vectors obtained from the electrostatic map Hamiltonian into the complex exponentials of

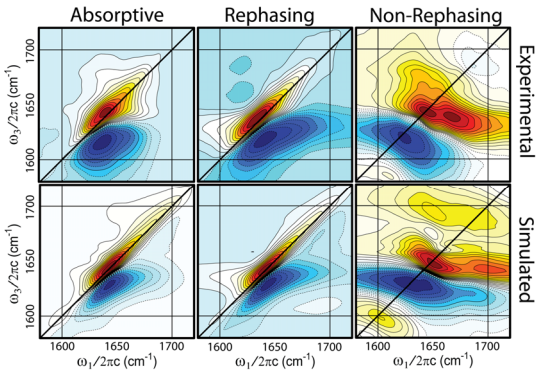

Figure 17 – Experimental and simulated absorptive, rephasing and non-rephasing spectra of Ubiquitin. Large rephasing amplitudes give rise to strongly inhomogeneously broadened absorptive spectra, as shown in this example. The amplitudes of the rephasing and non-rephasing signals normalized in the above contour plots. Positive and negative contours are represented by solid and dashed lines respectively. Non-linear contour spacing is used in order to emphasize the low-amplitude features.

Equation 4 or their Fourier transform representation as delta functions.(45) Dephasing is incorporated in the time-domain by adding exponential decays to each term or in the frequency domain by convolution with a complex Lorentzian (the Fourier transform of an exponential decay). As a reference, we summarize here the steps necessary to perform these calculations along with the necessary numerical factors.

1. *Solvated Structures*: To begin, a structural ensemble is generated, usually the result of fully solvated MD simulations. The inclusion of explicit solvent is important here to obtain the correct solvent-induced shifts of surface residues.
2. *One-Quantum Hamiltonian*: These starting structures are then used as input to generate a Hamiltonian trajectory/ensemble using one of the electrostatic and/or nearest-neighbor maps discussed above. The parameters generated consist of a set of one-quantum site energies ε_n , coupling constants J_{mn}, and zero- to one-quantum transition dipole moments $\boldsymbol{\mu}^{0,m}$.
3. *Two-Quantum Hamiltonian*: Using the assumption of a weakly anharmonic system described above, a two-quantum Hamiltonian is generated from the one-quantum Hamiltonian via the relations

Table 2 - Summary of selected electrostatic maps developed for Amide vibrational frequencies. The columns indicate the model system used to parameterize the map, the electrostatic parameters used to describe the vibrations, the number of sites (usually the amide bond N, H, C, and O atom locations), and the vibrational modes considered.

Model	Electrostatic Variable	Sites	Amide Modes	Site Energies	Coupling	System
Torii and Tasumi(1, 2)	N/A	N/A	I	no	yes	Gly_2 and Gly_3
Hamm and Woutersen(8)	N/A	N/A	I	no	yes	Gly_2
Cho(19)	Potential	4	I	yes	yes	NMA and Gly_2
Bouř and Keiderling(20)	Potential	4	I	yes	no	NMA and penta-peptide
Skinner(23)	Field	4/2	I	yes	no	NMA and Small Peptides
Mukamel(25)	Field, Gradient, 2nd Derivative	19	10 lowest modes	yes	no	NMA
Hirst(28)	Potential	4/7	I	yes	no	NMA and [Leu]-enkephalin
Jansen and Knoester(15, 29)	Field and Gradient	4	I	yes	yes	NMA
Wang(30)	Potential	4	I	yes	yes	Ala_2 and Gly_2
Ge(31)	Potential	4	I and II	yes	yes	3-10 peptide

$$\langle nn|\hat{H}|nn\rangle = 2\varepsilon_n - \Delta_n$$

$$\langle mn|\hat{H}|mn\rangle = \varepsilon_m + \varepsilon_n$$

with Δ = 16 cm^{-1} for diagonal entries (site energies) and for off-diagonal elements (couplings)

$$\langle mm|\hat{H}|mp\rangle = \sqrt{2}J_{mp}$$

$$\langle mn|\hat{H}|mp\rangle = J_{np}$$

$$\langle mn|\hat{H}|pq\rangle = 0$$

with *n*, *m*, and *p*, and *q* distinct site indices. The dipole moments are likewise obtained via harmonic scaling.

$$\mu^{m,mm} = \sqrt{2}\mu^{0,m}$$

$$\mu^{m,mn} = \mu^{0,n}$$

$$\mu^{m,np} = 0$$

4. *Eigenstates and Stick Spectra*: For each frame in the trajectory, the Hamiltonian generated above is diagonalized and the eigenstates and eigenvectors are used to calculate transition frequencies and dipole moments. These values then act as input for stick spectra, [see Eq 5] i.e. the Fourier transform of the static field expressions above

As above, the indices *a* and *b* indicate one-quantum states, while the sum over *c* is over two-quantum states. Note that the rephasing profile here gives signals at negative frequency in ω_1; the spectrum will be flipped (after convolution) before comparison with the rephasing signal. To obtain the correct peak intensities, the lab frame dipole moment components $\mu_i^A \mu_j^B \mu_k^C \mu_l^D$ must be ensemble averaged to account for the random orientation of molecules in the sample with respect to the laser polarization. For a spherically symmetric distribution of molecular orientations, the results are given analytically as (Equation 6 below)(113, 116) where the indices *Z* and *Y* refer to laboratory frame Cartesian axes, the subscripts α and β run over the *x*, *y*, and *z* axes of the molecular frame coordinate system and the quantities M_α^A are molecular frame transition dipole moment components obtained directly from the MD

trajectories above. For spherically symmetric systems, only the four laboratory frame polarizations *ZZZZ*, *ZZYY*, *ZYZY*, and *ZYYZ* give non-vanishing signals, with the last condition $\langle \mu_Z^A \mu_Y^B \mu_Y^C \mu_Z^D \rangle$ obtained via the relation

$$\left\langle \mu_Z^A \mu_Z^B \mu_Z^C \mu_Z^D \right\rangle = \left\langle \mu_Z^A \mu_Z^B \mu_Y^C \mu_Y^D \right\rangle + \left\langle \mu_Z^A \mu_Y^B \mu_Y^C \mu_Z^D \right\rangle + \left\langle \mu_Z^A \mu_Y^B \mu_Z^C \mu_Y^D \right\rangle .$$

5. *Convolution*: To simulate dephasing and population decay the stick spectra are convolved with complex profiles of the form

$$\frac{1}{(i\omega_1 + \gamma)(i\omega_3 + \gamma)} = \frac{\gamma^2 - \omega_1\omega_3}{(\omega_1^2 + \gamma^2)(\omega_3^2 + \gamma^2)} - i\frac{\gamma(\omega_1 + \omega_3)}{(\omega_1^2 + \gamma^2)(\omega_3^2 + \gamma^2)}$$

i.e. the half-sided Fourier transform of a two-dimensional exponential decay.

6. *Combined Spectra*: Finally, the rephasing spectrum is flipped horizontally by exchanging ω_1 with - ω_1. The complex spectra can then be plotted in a variety of ways for comparison to experiment. The real part of the sum of rephasing and non-rephasing signals gives the so-called "purely absorptive" (or correlation) surface, while the real part of the difference gives the dispersive contribution. Absolute value surfaces (of the complex signal—not the real part) are sometimes plotted instead, avoiding the appearance of negative peaks in the spectrum, but significantly increasing the width of the peaks.

A comparison of static ensemble spectra simulated using three different maps shows how the features observed in experimental spectra are all qualitatively reproduced by the different maps (see Figure 18).(15, 20, 28, 45) The two-peak spectra associated with β-sheets, and the single symmetric peak of α-helices is reproduced by the three maps; however some of the details, such as overall shift of the spectra, the ratio of intensities corresponding to the $\upsilon_{//}$ and $\upsilon_{\perp}$ peaks along with the ridge-structure associated with β-sheet spectra not well captured by the simulations. The parameterization does not include effects from other protein residues, which may or may not be captured in the simulations. Finally, the broadness of the calculated peaks in comparison with experiment are, in part, due to motional narrowing effects that are not captured by the static simulations.(15)

$$\langle \mu_Z^A \mu_Z^B \mu_Z^C \mu_Z^D \rangle = \sum_\alpha \left[\frac{1}{5} M_\alpha^A M_\alpha^B M_\alpha^C M_\alpha^D + \frac{1}{15} \sum_{\beta \neq \alpha} \left(M_\alpha^A M_\alpha^B M_\beta^C M_\beta^D + M_\alpha^A M_\beta^B M_\alpha^C M_\beta^D + M_\alpha^A M_\beta^B M_\beta^C M_\alpha^D \right) \right]$$

$$\langle \mu_Z^A \mu_Z^B \mu_Y^C \mu_Y^D \rangle = \sum_\alpha \left[\frac{1}{15} M_\alpha^A M_\alpha^B M_\alpha^C M_\alpha^D + \frac{1}{30} \sum_{\beta \neq \alpha} \left(4 M_\alpha^A M_\alpha^B M_\beta^C M_\beta^D - M_\alpha^A M_\beta^B M_\alpha^C M_\beta^D - M_\alpha^A M_\beta^B M_\beta^C M_\alpha^D \right) \right] \quad 5$$

$$\langle \mu_Z^A \mu_Y^B \mu_Y^C \mu_Z^D \rangle = \sum_\alpha \left[\frac{1}{15} M_\alpha^A M_\alpha^B M_\alpha^C M_\alpha^D - \frac{1}{30} \sum_{\beta \neq \alpha} \left(M_\alpha^A M_\alpha^B M_\beta^C M_\beta^D - 4 M_\alpha^A M_\beta^B M_\alpha^C M_\beta^D + M_\alpha^A M_\beta^B M_\beta^C M_\alpha^D \right) \right]$$

$$S_{ijkl}^R(\omega_1, \omega_3) \mu \sum_{a,b=1}^{N} \delta(\omega_1 + \omega_{a0}) \left\{ \left(\langle \mu_i^{0a} \mu_j^{0b} \mu_k^{a0} \mu_l^{b0} \rangle + \langle \mu_i^{0a} \mu_j^{a0} \mu_k^{0b} \mu_l^{b0} \rangle \right) \delta(\omega_3 - \omega_{b0}) - \sum_{c=1}^{N(N+1)/2} \langle \mu_i^{0a} \mu_j^{0b} \mu_k^{bc} \mu_l^{ca} \rangle \delta(\omega_3 - \omega_{ca}) \right\} \quad 6$$

$$S_{ijkl}^{NR}(\omega_1, \omega_3) \mu \sum_{a,b=1}^{N} \delta(\omega_1 - \omega_{a0}) \left\{ \left(\langle \mu_i^{0a} \mu_j^{0b} \mu_k^{b0} \mu_l^{a0} \rangle + \langle \mu_i^{0a} \mu_j^{a0} \mu_k^{0b} \mu_l^{b0} \rangle \right) \delta(\omega_3 - \omega_{a0}) - \sum_{c=1}^{N(N+1)/2} \langle \mu_i^{0a} \mu_j^{0b} \mu_k^{ac} \mu_l^{cb} \rangle \delta(\omega_3 - \omega_{cb}) \right\}$$

4.8 Dynamic Simulations

The methods described so far model 2D IR spectra by combining an ensemble of "snapshot" 2D spectra calculated for a single static structure. In reality, although global protein structure is static on the picosecond time scale relevant to 2D IR measurements, local solvent environments do fluctuate on a sub-picosecond timescale giving rise to complex dynamic effects in the 2D spectrum, in particular as a function of the *waiting time*. Such dephasing and relaxation processes can be modeled at a higher level using a numerical integration approach originally described for amide I vibrations by Torii(118) and applied to a wide variety of systems by Jansen and Knoester.(88) In this approach, the time-dependent excitonic Hamiltonian generated from an MD trajectory (step 3 in the list above), is broken down into small time intervals over which it can be assumed to be static. An exponential time-evolution operator is then calculated for each individual time step, allowing the time-dependent Schrödinger equation for the unperturbed system to be integrated numerically. This approach is particularly useful for providing realistic lineshapes for 2D surfaces as well as providing a numerically straightforward description of energy transfer between sites or bands, such as amide I and II.(27, 119) The accuracy of this approach depends strongly on the accuracy of the mapping used to parameterize the excitonic Hamiltonian.

More recently, a variety of flavors of mixed QM/MM type calculations have begun to be applied to linear and 2D amide I spectra.(120-122) These approaches are appealing in that they in principle allow spectroscopic features to be obtained directly from *ab initio* calculations, although at least for the present, computational requirements limit both the accuracy of the calculations and the size of system to which they can be applied. In the coming years, an important challenge for all of these approaches will be to bridge the gap between qualitative descriptions of amide I spectroscopy and a quantitative map between IR spectroscopy and protein structure and dynamics in solution.

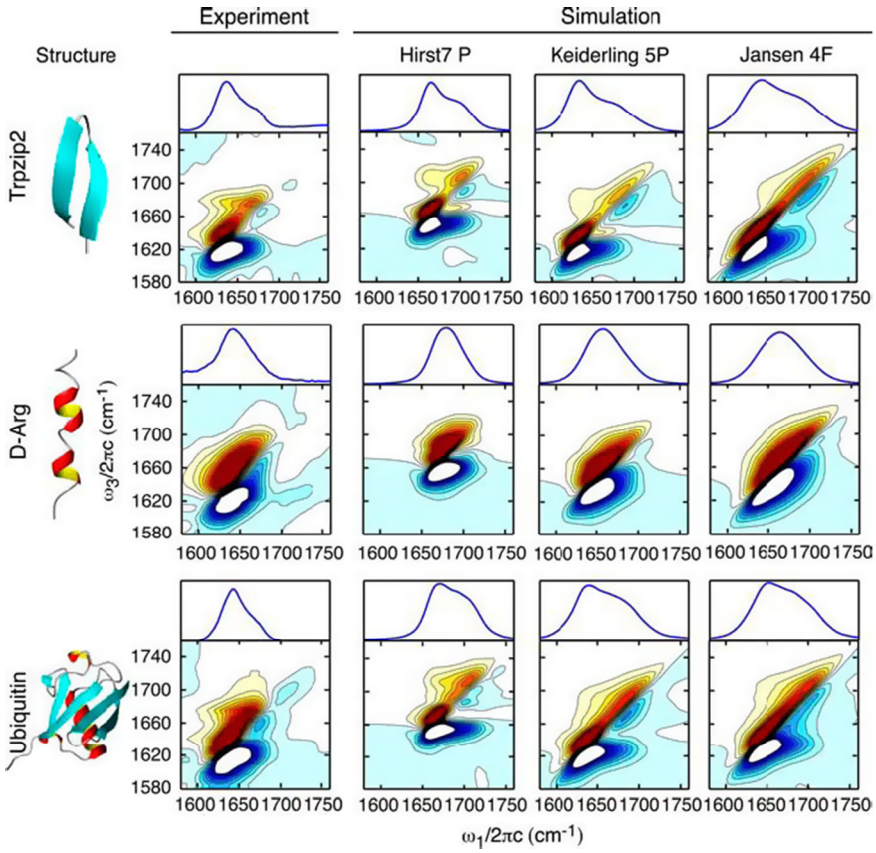

Figure 18 – Comparison between calculated and experimental spectra for three different site maps. Adapted from [Ganim, Z.; Tokmakoff, A. Biophys J 2006, 91, 2636-2646.]

4.9 Summary and Outlook

Structure-spectroscopy maps provide the essential connection between structure and spectra. Though the agreement is for the most part qualitative, the simulations provide an intuitive interpretation of the spectral features in order to help interpret the experimental data and inform the design of new experiments. For example, simulations can be used to explore the effects of isotope labels at different positions in order to strategically select labeling sites. It is important to emphasize that experimental spectra provide structural constraints, and the quality of the map determines the degree to which the experimental data can be interpreted and how much structural information can be extracted from the measured spectra. Since the simulations are relatively inexpensive and trivially parallelizable, hundreds or thousands of trial structures can be simulated in order to extract structures that best describe the experimental data. Importantly, the maps can be easily interfaced with Markov state models in order to explore the heterogeneity of folding mechanisms in small proteins.(22)

It is worth noting a number of possibilities for future work which may prove useful in broadening the scope of amide I spectral modeling capabilities. Although as illustrated above, existing models provide a good qualitative description of experimental amide I spectra, quantitative precision is often lacking. Several opportunities for improvement stem from protein sidechain considerations. For example, although several sidechain groups absorb in the amide I spectral region (especially Asp, Glu, Asn, Gln, and Arg), these moieties are usually neglected for spectral calculations. An accurately parameterized site energy map for these groups would be of interest for many applications, particularly since their vibrations are expected to be largely localized, making them potentially useful as probes of local structure. Likewise, a sidechain-specific study of amide I and II vibrations would be useful to more closely examine the effects of individual side chains on the site energies of adjacent peptide groups.(108, 123) More generally, it should be noted that the existing spectral maps have been parameterized and tested against a relatively small number of model compounds (notably NMA and Alanine and Glycine dipeptides), raising questions about their quantitative accuracy when transferred to new systems such as large proteins. The basic assumption involved in these calculations is that site energy shifts are essentially non-specific, so that, for example, the electrostatic influence of solvating water can be treated identically to that of adjacent side chains and backbone atoms. It would be of great interest to examine these assumptions on a case by case basis, potentially allowing for greater accuracy through individual parameterizations for different amino acids or functional groups. On the experimental front, residue-specific isotope labels could assist this process, since they provide non-perturbative, spectrally isolated local probes of the site energy shifts associated with individual residues.

EXAMPLES OF PROTEIN 2D IR SPECTROSCOPY

The capabilities of multidimensional spectroscopy as a technique for measuring protein structure, conformational flexibility, and transient peptide unfolding mechanisms are demonstrated by the examples in this section. The first example shows how protein secondary structure is extracted from amide I 2D IR spectra. The second example highlights how cross-peaks provide a direct structural view of the solvent exposure of proteins. The third example describes the unfolding mechanism of a small peptide, TrpZip2, investigated using non-equilibrium temperature-jump 2D IR spectroscopy.

5.1 Determination of protein structure composition with 2D IR spectroscopy

This example illustrates the analytical aspect of 2D IR spectroscopy and its application to the quantitative measurement of secondary structure composition based on amide I two-dimensional spectra. As described in previous sections, secondary protein structures exhibit particular signatures in the amide I spectrum. However, to date most infrared studies of protein structure have provided qualitative structural information, or have been carried out on systems with known equilibrium structures. Previous attempts to quantitate secondary structure from FTIR spectra have been based on spectral deconvolution combined with lineshape fitting methods which rely on a number of assumptions which limit the usefulness and applicability of the technique.(124) Here we provide an example of how the enhanced structural sensitivity of 2D IR spectroscopy can be harnessed to develop a quantitative assay of protein secondary structure

composition without *a priori* spectral assignments.(26)

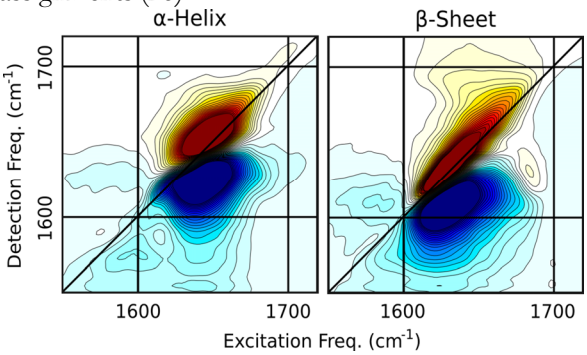

Figure 19 – Singular value components extracted from a library containing sixteen protein spectra with mixed alpha/β components. Adapted from [Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. Analyst 2012, 137, 1793-1799.]

In brief, a library is constructed by collecting spectra of proteins with known crystal structures, and the signatures of individual secondary structures are extracted through singular value decomposition (SVD), a form of principal component analysis. For the purpose of the analysis, proteins are treated as mixtures of α-helices, β-sheets and unassigned conformations. The analysis makes two important assumptions: 1. all α-helices, or β-sheets, have the same spectral signatures regardless of the structural details. For example, parallel and antiparallel β-sheets are assumed to have the same spectrum. As described previously, peak center frequencies and intensity ratios depend on the size and number of strands in a β-sheet. 2. Spectra are not sensitive to super-secondary structure, namely, the coupling between secondary structures is negligible.

Figure 19 shows component 2D IR spectra decomposed from the protein set. The helix spectrum shows a typical "figure-8 shape" represented by a single peak centered near 1650 cm^{-1} and the β-sheet shows a diagonally elongated band, corresponding to the $\upsilon_{\perp}$, and $\upsilon_{//}$ bands centered around 1630 cm^{-1} and 1670 cm^{-1} respectively, along with the cross-ridges that give the spectrum its characteristic "Z-shape". The projection of an "unknown" spectrum onto the α-helix and β-sheet component spectra is proportional to the percentage of residues in α-helix and β-sheet conformations in the unknown protein respectively. Figure 20 shows the fraction of residues in different secondary structures predicted from 2D IR spectra along with the respective fractions extracted from analysis of the crystal structures. The root-mean-squared errors are: 6.7 % for α-helices, 7.7 % for β-sheets, and 8.2% for unstructured conformations respectively. These results represent a significant improvement over linear IR absorption spectroscopy, for which the same singular value decomposition analysis produces an average error of 19.5 % for α-helices, 8.3% for β-sheets and 21.5% for unassigned conformations. The comparison highlights the enhanced structural-sensitivity afforded by nonlinear spectroscopies, which are inherently sensitive to the anharmonic site energies and coupling contributions to the amide I vibrations. Although the data analysis overlooks the finer spectral details associated with different protein structures, the accuracy of the results justifies the assumptions. In principle, the conformational selectivity can be further refined by expanding the protein set and including more varied secondary and supersecondary structural motifs. Finally, the example demonstrated here highlights particular strengths of infrared spectroscopy: 1. The analysis is performed on stock proteins without the need for complex sample preparation techniques. 2. Protein spectra can be collected in solution under biological conditions. 3. The analysis is not limited to fully-solvated globular proteins: membrane proteins, protein-complexes, aggregates or other samples which remain difficult to characterize with traditional techniques, can be analyzed with amide I 2D IR spectroscopy.

5.2 Measuring solvent exposure of secondary structures through multimode amide I/II 2D IR spectroscopy

While the amide I band has received a great deal of attention as a structural probe, amide II vibrations are excellent markers of conformational flexibility and solvent exposure. The predominant N-H bending character is responsible for the large shift observed upon hydrogen/deuterium (H/D) exchange. Two-dimensional spectroscopy has the advantage of being able to measure the couplings between amide I and amide II modes, combining the structural sensitivity of amide I with the solvent-exposure sensitivity of amide II.(62, 125) The two bands are coupled primarily through shared motions of the backbone atoms such that modes that share common residues are more strongly coupled, than modes that are delocalized over different sections of the protein backbone. The number of shared residues determines the strength of the coupling. During an H/D exchange experiment, residues that

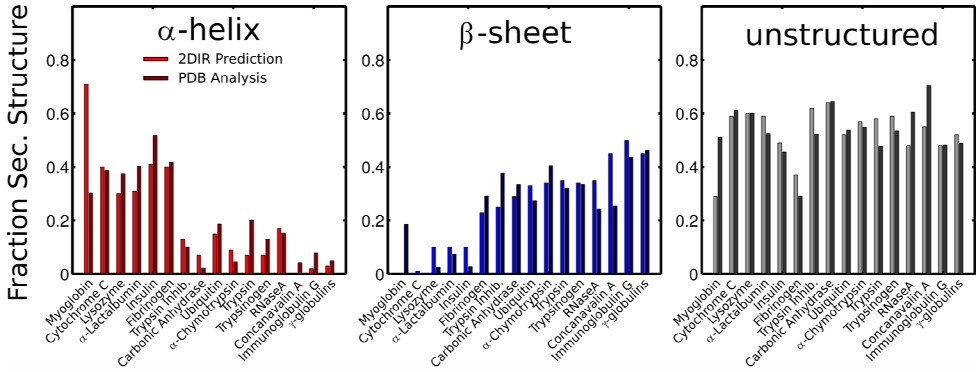

Figure 20 – Percent of residues in alpha-helix, β-sheet and unstructured configurations for sixteen proteins with known crystal structures. The dark bars indicate the values extracted from the crystal structures and the light bars indicate the values predicted from analysis of experimental 2D IR spectra. Adapted from [Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. Analyst 2012, 137, 1793-1799.]

are solvent exposed exchange faster that those buried within the protein core. The 100 cm^{-1} frequency difference between amide II and II' modes allows for selective excitation of deuterated—solvent exposed—or protonated –buried—residues. On-site couplings allow for efficient energy transfer to the amide I band. As described previously, amide I peaks can be related to protein secondary structure, thus the 2D IR cross peaks between amide II/II' and amide I/I' contain a *structural* view of solvent-exposure. In simpler terms, the amide II excitation selectively "tags" either buried or exposed residues, and the amide I detection extracts the spectral information related to secondary structure.

Figure 21 shows a series of multimode 2D IR spectra of Concanavalin A, Myoglobin, RNase A and Ubiquitin, together with projections onto the detection axis of the main amide I band and amide I/II cross peaks. The difference between the features observed in the diagonal peak projection and the cross peaks provides a qualitative view of the secondary structures. The absence of cross peaks indicates rapid H/D exchange for that protein. First, in the case of Concanavalin A and Ubiquitin, the amide II/I slices show significantly less intensity in the center of the band (1660-1680 cm^{-1}), compared to the amide I/I diagonal peak slices, indicating that turn and coil regions exchange faster than α-helices or β-sheets amide I 2D IR spectra and crystal structures are shown in Figure 3. Similarly, the cross-peak projection of Myoglobin is red-shifted with respect to the diagonal slice, indicating exchange primarily in the coil regions. RNase A contains features corresponding to β-sheet and α-helices, indicating a high degree of structural stability under the experimental conditions. In the case of Ubiquitin, the lack of intensity around 1680 cm^{-1} suggests that residues in β-sheet conformations are less stable and more prone to undergo H/D exchange.

In summary, multimode H/D exchange spectroscopy gives a direct view of structural stability and solvent exposure of secondary structures within the protein. The technique is general, does not require labeling, and has high time resolution, which in principle allows for the combination of H/D exchange with temperature-jump in order to probe structural changes during protein unfolding. Millisecond time resolution can be attained through rapid mixing experiments. Combined with isotope labeling, multimode 2D IR spectroscopy has the ability to measure the exchange kinetics of individual residues and thus provide a much more detailed view of the backbone solvent exposure.

5.3 Transient T-jump spectroscopy as a probe of unfolding dynamics of TrpZip2

Understanding protein folding is an outstanding challenge in modern biophysics and biology. The three-dimensional structure of proteins is encoded in the sequence, but it is not understood how proteins adopt a stable conformation, how sequence determines structure, or what microscopic interactions, among the thousands of degrees of freedom, are relevant to the folding process. Conformational dynamics, dictated by the subtle energetic interactions between contact formation and protein/solvent enthalpy and entropy remain extremely difficult to study with current

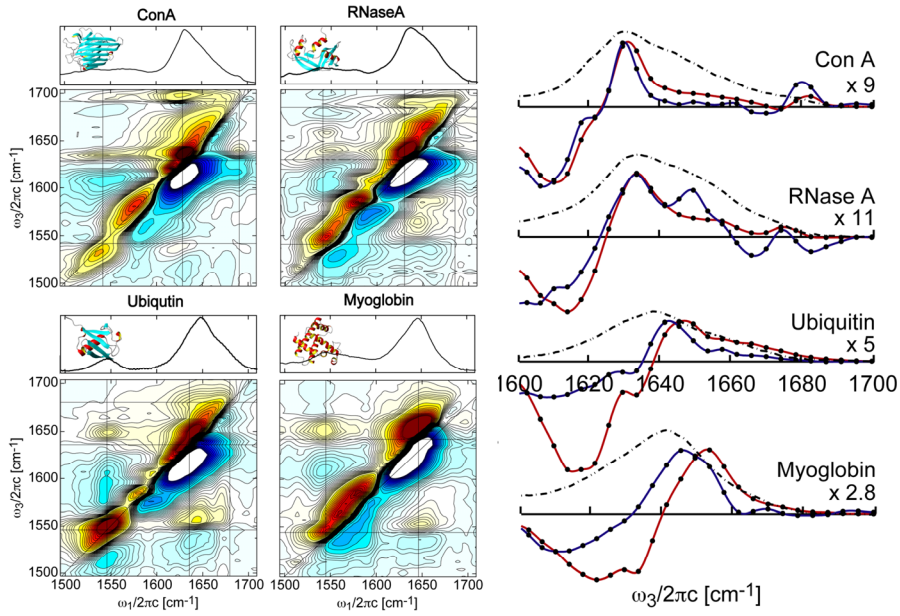

Figure 21 – Amide I/II 2D IR spectra of Concanavalin A (Con A), RNaseA, Ubiquitin and Myoglobin. The spectra show two main diagonal bands corresponding to Amide I/I' (~1650 cm-1), Amide II (1550 cm-1). Adapted from [DeFlores, L. P.; Ganim, Z.; Nicodemus, R. A.; Tokmakoff, A. J Am Chem Soc 2009, 131, 3385-3391]

experimental techniques. Advances in computer technologies have recently enabled atomistic molecular dynamics simulations of protein folding processes.(126-128) Although the simulations provide very detailed mechanisms for protein folding, there has been virtually no experimental data to validate the predictions. Temperature-jump (T-jump) 2D IR spectroscopy is a promising new technique that leverages the structural resolution power of 2D IR spectroscopy and applies it to the study of thermally-induced changes in peptides and proteins. (97, 104, 129) As illustrated in section 2, isotope-edited amide I 2D IR spectroscopy provides a localized view of structural disorder and heterogeneity of peptides undergoing denaturation. In addition to ultrafast picosecond dynamics of protein-protein or protein-solvent interactions, ultrafast measurements enable differentiation of conformations interconverting on the micro- to nanosecond timescales.

The tryptophan zipper 2 (TrpZip2), shown in Figure 22, forms a type-I' β-turn stabilized by cross-strand hydrophobic interactions involving the tryptophan indole rings.(130) Fast folding rates and small number of residues make it an attractive target for experimental and computational folding studies. As described in section 2, temperature-dependent 2D IR spectroscopy has been used to characterize the residual structure and disorder of TrpZip2.(22) Spectra suggest the presence of at least two stable turn conformations: a native type-I β-turn, and a disordered bulged loop.

T-jump spectroscopy is able to assess the rate of interconversion between the states by examining the kinetics of the isotope peaks as a function of delay following the T-jump. Here we present a combination of T-jump HDVE and 2D IR spectroscopy. Since HDVE spectra can be collected ~300x faster than 2D IR, the standard approach involves obtaining kinetic rates by finely sampling the T-jump delays using HDVE spectroscopy, and extracting structural information at selected times using T-jump 2D IR spectroscopy. Figure 23 shows the measured sample response times. The curves are extracted from a singular value decomposition (SVD) analysis of the kinetic traces. Three distinct timescales are observed: a small-amplitude nanosecond rise, a stronger microsecond rise, and a millisecond decay. The millisecond decay is due to re-equilibration as heat is dissipated from the laser focus region. The first two components are attributed to the response of the peptide. In order to interpret the rates from a structural perspective, 2D

IR spectra are collected at nanosecond and microsecond delays following the T-jump. T-jump difference (transient - equilibrium) spectra are shown in Figure 23. Peaks above the diagonal are positive whereas peaks below the diagonal are negative in equilibrium spectra. In difference spectra negative peaks above the diagonal represent loss of signal, whereas negative peaks below the diagonal represent gain of signal and vice versa.

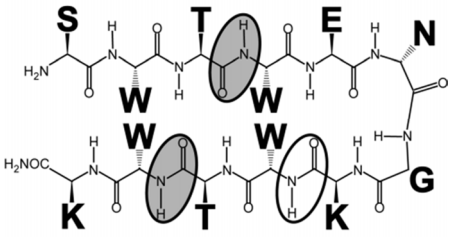

Figure 22 – Tryptophan Zipper 2 (TrpZip2). Residues selected for isotope labels T3, K8, and T10 are highlighted for reference.

Transient-2D IR: nanosecond response

The nanosecond component shows positive/negative doublet peaks with narrow anti-diagonal widths spanning the entire diagonal region. The bleach is most intense in the low and high frequency regions ~1635 cm^{-1} and 1670 corresponding to the equilibrium loss of the parallel and antiparallel β-sheet modes. The off-diagonal ridges are attributed to a broadening of the peak due to increased conformational flexibility and weaker interstrand hydrogen bonds. The K8 and TT double labels show features consistent with the loss of hydrogen bonding: loss of the equilibrium peaks, indicated by diagonal arrows, along with appearance of blue-shifted peaks, indicated by horizontal arrows. Spectral shifts can result from changes in hydrogen bond strengths and do not necessarily reflect large conformational changes within the system.

Transient-2D IR: microsecond response

Changes observed in the microsecond spectra (Figure 23, third column) are indicative of disordering and potential fraying of the β-strands. Two loss features at 1640 cm^{-1} and 1680 cm^{-1} along with the corresponding off-diagonal ridges, are associated with the loss of β-sheet structure. The gain features in the 1660 cm^{-1} region, indicated by the arrows, are attributed to frayed or partially disordered peptide conformations. The interpretation is consistent with results from Markov state modeling, which indicate that disordered conformations show an intensity gain in the 1650 cm^{-1} region. The double-label, TT, spectrum shows loss features above and below the diagonal indicative of an overall loss of intensity due to weakening of contacts in the mid-strand region. A ~15 cm^{-1} blue-shift is expected for the TT peak upon loss of coupling, which has the effect of hiding the nascent peak within the main amide I bleach feature.

Assignment of Markov states

Interpretation of the T-jump 2D IR spectra in the context of Markov state simulations (see Ref (22)), suggest the presence of three main states: a *folded* state characterized by a type-I' turn stabilized by inter-strand hydrogen bonds and side chain packing, a *misfolded* conformation exhibiting a bulged turn conformation with a solvent-exposed K8 and mis-registered hydrogen bond contacts, a *disordered* state representing conformations that lack the standard interstrand hydrogen bonds. Equilibrium spectra indicate that a significant portion of the misfolded state exists in equilibrium with the folded state. The pulse-width limited response (<5ns) is attributed to loosening of the interstrand contacts; the microsecond response indicates a partial shift of population from the folded and misfolded into the disordered state. Since the refolding kinetics are significantly faster than the temperature relaxation, there is no trigger event to synchronize the re-folding process, thus the peptide response follows the solvent relaxation kinetics, making it difficult to separate the spectral changes associated with refolding from the solvent response.

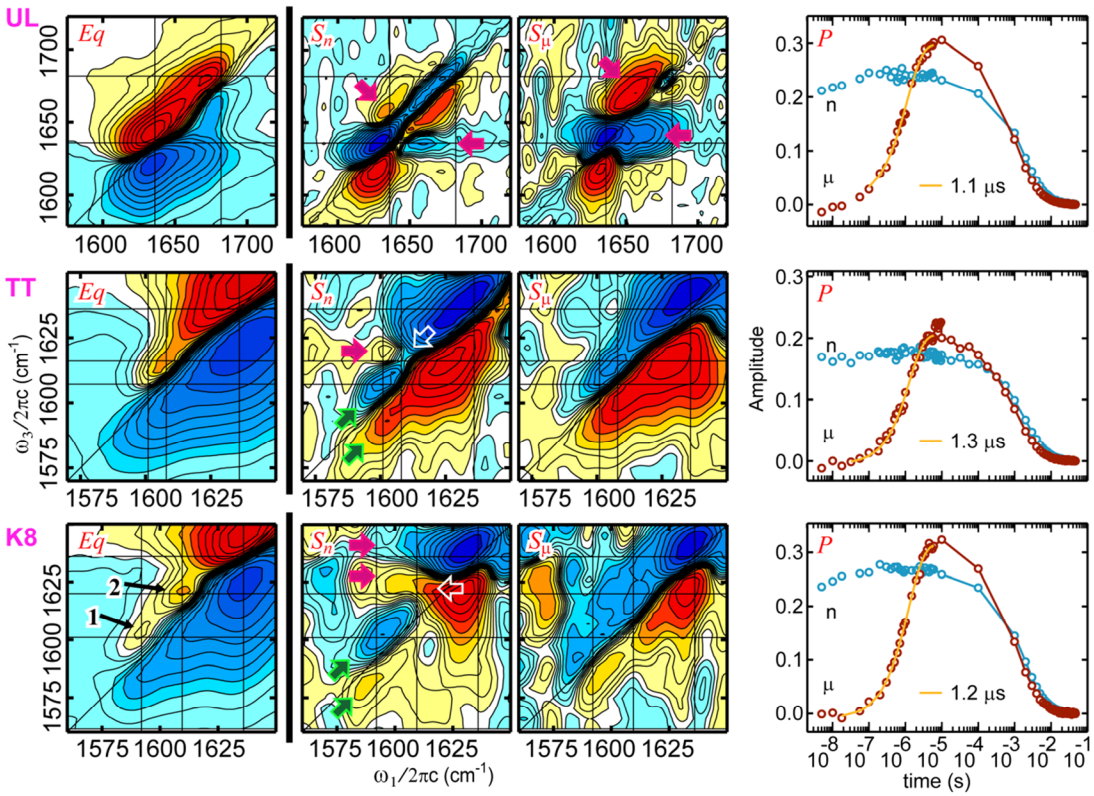

Figure 23 – TrpZip2 transient and equilibrium 2D IR spectra and HDVE kinetic curves. The left column shows equilibrium 2D IR spectra of unlabeled (UL), T3T10 (TT) and K8 13C=18O isotope labeled samples. The center columns show difference spectra associated with the nanosecond (right) and microsecond (left) responses. Solid and dashed contours represent positive and negative peaks respectively. The right column shows the response curves measured with HDVE spectroscopy.

5.4. Additional examples

The examples above are vignettes that illustrate some capabilities that are enabled by amide 2D IR spectroscopy of proteins. Recent applications of 2D IR spectroscopy in the study of the structure and dynamics of proteins are numerous, and we here briefly point the reader to some of the many recent examples of this method. Recent studies of amyloid formation by amylin have been used to explain the mechanism of fibril nucleation and growth (67) and the binding of a fibril inhibitor.(9) Structural studies using isotope labeling have described conformationally heterogeneous states of peptides.(11, 22) A variety of implementations have been used to investigate structure, conformational changes, and fast fluctuations in membrane proteins including the CD3ζ transmembrane peptide,(131) the influenza virus M2 proton channel,(12, 13) and transmembrane helix dimers.(132) Transient 2D IR has been used in protein folding studies aimed at understanding photoinitiated structural rearrangements in small peptides,(133) and analogs,(134) and temperature-jump induced unfolding dynamics of ubiquitin.(135, 136) Transient and steady state methods have revealed fast time-scale fluctuations and site-specific protein dynamics in the active sites of enzymes, such as myoglobin,(137) horseradish peroxidase,(138) and formate dehydrogenase,(139) and bulk solution, villin headpeice,(140, 141) HIV reverse transcriptase.(142) Protein-protein interactions and the folding associated with binding have been observed in insulin dimer.(143) A broader perspective can be gained from a number of recent reviews on protein 2D IR. (2, 7, 44, 90)

CONCLUSION AND OUTLOOK

Over the past decade protein 2D IR spectroscopy has evolved from a novel optical technique into a more mature research tool which provides the structural sensitivity and ultrafast time resolution

required to shed light on problems related to protein structure and conformational dynamics. With the ability to probe structural information in systems which are not readily accessed by traditional methods, we imagine that the discipline will continue to find a variety of applications. It will find increasing use in the study of membrane proteins, where it can be applied to structure and folding of single proteins and complexes, the mechanism of transduction of ions and small molecules through channels,(144) and signal transduction processes in receptors. The combination of isotope-edited 2D IR and structural modeling will not only provide restraints to solve structures for systems not amenable to crystallography or NMR, but also enable studies of disordered proteins and unstructured states. We imagine that these studies will provide structural ensembles that characterize intrinsically disordered chains, encountered in coupled folding and binding problems in protein-protein and protein-DNA interactions. The level of structural detail needed for various protein structural problems varies from atomistic to nanometer-scale. 2D IR can be fashioned to reveal contacts in binding sites, gross features in protein complexes, and oligomerization processes. To enable such studies, isotope labeling need not only at specific sites, but it can be applied to an entire secondary structure, domain, or protein to facilitate studies of mesoscopic structures.

The truly unique nature of 2D IR lies in its combined structural and temporal resolution, which enables kinetics and dynamics studies and opens the door to direct observation of the function of proteins. Conformational dynamics in folding, binding, signal transduction are processes which can potentially be visualized at high time-resolution with triggered transient 2D IR methods. Similarly proton and ion transport processes in enzymes and membrane proteins can also be triggered and followed with ultrafast time resolution.

Technological advances have improved and simplified the experimental methods, so that 2D IR instruments can be commercialized. Nonetheless, the need for innovation and development of the method remains higher than ever. Further developments in instrumentation, theory, and synthesis are needed to make them more widely available. An increased library of protein vibrations available for detailed structural interpretation and the new pulsed IR light sources that can simultaneously excite and probe many vibrational modes are both avenues that will greatly increase the number of applications. Significant efforts have to be made in order to develop structure based models for 2D IR constraints that are quantitative and have accuracy to meaningfully distinguish conformational variation and solvent environment on the angstrom scale. Close ties with the protein simulation community are needed for experimentalists to atomistically interpret data, and will also provide the new data that can be used to validate the force fields used in simulation. Similarly methods for incorporating isotope and molecular labels into proteins will become increasingly important for these studies. Fortunately, this area of research is of wide interest to many spectroscopies, and is closely related to problems which the protein NMR community has worked on for many years.

In conclusion, the ultrafast time resolution and structural sensitivity of 2D IR spectroscopy offers a broadly applicable approach to study protein structure and fast conformational motions which remain virtually inaccessible by conventional techniques. Given the possibilities and the applicability to complex samples such as membrane proteins, complexes, fibers, oligomers and aggregates, and disordered systems, we believe there will be extraordinary future growth in information content and applications of protein 2D IR.

Acknowledgements

We acknowledge members of the Tokmakoff group whose accomplishments over the past decade are summarized in this chapter: Hoi Sung Chung, Matthew Decamp, Lauren DeFlores, Nuri Demirdöven, Ziad Ganim, Munira Khalil, Poul Petersen, and Adam Smith. We acknowledge Adam Squires and Ziad Ganim for providing the unpublished insulin data, Krupa Ramasesha and Luigi DeMarco for providing the BBIR spectrum in Fig. 9, Kevin Jones for the TrpZip2 T-jump data, and Joshua Lessing for the waiting time data on the elastin-like peptide. We thank Kevin Jones and Chunte Sam Peng for helpful comments on the manuscript. The work is supported by the National Science Foundation (CHE-0911107) and Agilent Technologies. Some research described in this paper was partially funded by the Department of Energy (DE-FG02-99ER14988).

REFERENCES

1. Torii H & Tasumi M; Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. *J Raman Spectrosc* 29(1):81-86 (1998)
2. Kim YS & Hochstrasser RM; Applications of 2D IR Spectroscopy to Peptides, Proteins, and Hydrogen-Bond Dynamics. *J Phys Chem B* 113(24):8231-8251 (2009)
3. Petersen PB & Tokmakoff A; Source for ultrafast continuum infrared and terahertz radiation. *Opt Lett* 35(12):1962-1964 (2010)
4. Strasfeld DB, Ling YL, Gupta R, Raleigh DP, & Zanni MT; Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide. *The journal of physical chemistry. B* 113(47):15679-15691 (2009)
5. Hamm P & Zanni M (2011) *Concepts and Methods of 2d Infrared Spectroscopy* (Cambridge University Press).
6. Cho M (2009) *Two-Dimensional Optical Spectroscopy* (CRC Press).
7. Fayer MD; Dynamics of Liquids, Molecules, and Proteins Measured with Ultrafast 2D IR Vibrational Echo Chemical Exchange Spectroscopy. *Annual Review of Physical Chemistry* 60(1):21 (2009)
8. Hamm P & Woutersen S; Coupling of the Amide I Modes of the Glycine Dipeptide. *Bulletin of the Chemical Society of Japan* 75(5):985-988 (2002)
9. Middleton CT*, et al.*; Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor. *Nat Chem* 4(5):355-360 (2012)
10. Kim YS, Liu L, Axelsen PH, & Hochstrasser RM; 2D IR provides evidence for mobile water molecules in beta-amyloid fibrils. *P Natl Acad Sci USA* 106(42):17751-17756 (2009)
11. Lessing J*, et al.*; Identifying Residual Structure in Intrinsically Disordered Systems: A 2D IR Spectroscopic Study of the GVGXPGVG Peptide. *J Am Chem Soc* 134(11):5032-5035 (2012)
12. Ghosh A, Qiu J, DeGrado WF, & Hochstrasser RM; Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy. *P Natl Acad Sci USA* 108(15):6115-6120 (2011)
13. Manor J*, et al.*; Gating Mechanism of the Influenza A M2 Channel Revealed by 1D and 2D IR Spectroscopies. *Structure* 17(2):247-254 (2009)
14. Wang L, Middleton CT, Zanni MT, & Skinner JL; Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides. *J Phys Chem B* 115(13):3713-3724 (2011)
15. Jansen TL & Knoester J; A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy. *J Chem Phys* 124(4)(2006)
16. Zhuang W, Abramavicius D, Voronine DV, & Mukamel S; Simulation of two-dimensional infrared spectroscopy of amyloid fibrils. *P Natl Acad Sci USA* 104(36):14233-14236 (2007)
17. Munoz V; Conformational dynamics and ensembles in protein folding. *Annu Rev Bioph Biom* 36:395-412 (2007)
18. Kubelka J, Hofrichter J, & Eaton WA; The protein folding 'speed limit'. *Curr Opin Struc Biol* 14(1):76-88 (2004)
19. Ham S, Kim JH, Lee H, & Cho MH; Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA-nD(2)O complexes. *J Chem Phys* 118(8):3491-3498 (2003)
20. Bour P & Keiderling TA; Empirical modeling of the peptide amide I band IR intensity in water solution. *J Chem Phys* 119(21):11253-11262 (2003)
21. Chung HS, Shandiz A, Sosnick TR, & Tokmakoff A; Probing the Folding Transition State of Ubiquitin Mutants by Temperature-Jump-Induced Downhill Unfolding. *Biochemistry-Us* 47(52):13870-13877 (2008)
22. Smith AW*, et al.*; Melting of a beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations. *J Phys Chem B* 114(34):10913-10924 (2010)
23. Schmidt JR, Corcelli SA, & Skinner JL; Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic

structure/molecular dynamics approaches. *J Chem Phys* 121(18):8887-8896 (2004)
24. Woutersen S, Mu Y, Stock G, & Hamm P; Hydrogen-bond lifetime measured by time-resolved 2D-IR spectroscopy: N-methylacetamide in methanol. *Chem Phys* 266(2-3):137-147 (2001)
25. Hayashi T, Zhuang W, & Mukamel S; Electrostatic DFT map for the complete vibrational amide band of NMA. *J Phys Chem A* 109(43):9747-9759 (2005)
26. Baiz CR, Peng CS, Reppert ME, Jones KC, & Tokmakoff A; Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* 137(8):1793-1799 (2012)
27. Bloem R, Dijkstra AG, Jansen TLC, & Knoester J; Simulation of vibrational energy transfer in two-dimensional infrared spectroscopy of amide I and amide II modes in solution. *J Chem Phys* 129(5)(2008)
28. Watson TM & Hirst JD; Theoretical studies of the amide I vibrational frequencies of [Leu]-enkephalin. *Mol Phys* 103(11-12):1531-1546 (2005)
29. Jansen TL, Dijkstra AG, Watson TM, Hirst JD, & Knoester J; Modeling the amide I bands of small peptides. *J Chem Phys* 125(4)(2006)
30. Cai KC, Han C, & Wang JP; Molecular mechanics force field-based map for peptide amide-I mode in solution and its application to alanine di- and tripeptides. *Phys Chem Chem Phys* 11(40):9149-9159 (2009)
31. Maekawa H, De Poli M, Moretto A, Toniolo C, & Ge NH; Toward Detecting the Formation of a Single Helical Turn by 2D IR Cross Peaks between the Amide-I and -II Modes. *J Phys Chem B* 113(34):11775-11786 (2009)
32. Barth A; Infrared spectroscopy of proteins. *Bba-Bioenergetics* 1767(9):1073-1101 (2007)
33. Barth A & Zscherp C; What vibrations tell us about proteins. *Q Rev Biophys* 35(4):369-430 (2002)
34. Jackson M & Mantsch HH; The Use and Misuse of Ftir Spectroscopy in the Determination of Protein-Structure. *Crit Rev Biochem Mol* 30(2):95-120 (1995)
35. Krimm S & Bandekar J; Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. *Adv Protein Chem* 38:181-364 (1986)
36. Davydov AS (1971) *Theory of molecular excitons* (Plenum Press).
37. Chung HS & Tokmakoff A; Visualization and characterization of the infrared active amide I vibrations of proteins. *J Phys Chem B* 110(6):2888-2898 (2006)
38. Cheatum CM, Tokmakoff A, & Knoester J; Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy. *J Chem Phys* 120(17):8201-8215 (2004)
39. Torii H & Tasumi M; 3-Dimensional Doorway-State Theory for Analyses of Absorption-Bands of Many-Oscillator Systems. *J Chem Phys* 97(1):86-91 (1992)
40. Torii H & Tasumi M; Application of the 3-Dimensional Doorway-State Theory to Analyses of the Amide-I Infrared Bands of Globular-Proteins. *J Chem Phys* 97(1):92-98 (1992)
41. Abramavicius D, Zhuang W, & Mukamel S; Peptide secondary structure determination by three-pulse coherent vibrational spectroscopies: A simulation study. *J Phys Chem B* 108(46):18034-18045 (2004)
42. Baiz CR, Peng CS, Reppert ME, Jones KC, & Tokmakoff A; Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* (2012)
43. Demirdoven N*, et al.*; Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure. *J Am Chem Soc* 126(25):7981-7990 (2004)
44. Ganim Z*, et al.*; Amide I two-dimensional infrared Spectroscopy of proteins. *Accounts Chem Res* 41(3):432-441 (2008)
45. Ganim Z & Tokmakoff A; Spectral signatures of heterogeneous protein ensembles revealed by MD simulations of 2DIR spectra. *Biophys J* 91(7):2636-2646 (2006)
46. Ham S*, et al.*; Amide I Modes of alpha-Helical Polypeptide in Liquid Water: Conformational Fluctuation,
Phase Correlation, and Linear and Nonlinear Vibrational Spectra. *J Phys Chem B* 108(26):9333 (2004)
47. Nevskaya NA & Chirgadze YN; Infrared-Spectra and Resonance Interactions of

Amide-One and Amide-2 Vibrations of Alpha-Helix. *Biopolymers* 15(4):637-648 (1976)
48. Woutersen S & Hamm P; Nonlinear two-dimensional vibrational spectroscopy of peptides. *J. Phys.: Condens. Matter* 14(39):1035 - 1062-1035 - 1062 (2002)
49. Wang JP & Hochstrasser RM; Characteristics of the two-dimensional infrared spectroscopy of helices from approximate simulations and analytic models. *Chem Phys* 297(1-3):195-219 (2004)
50. Kubelka J, Silva RAGD, & Keiderling TA; Discrimination between peptide 3(10)- and alpha-helices. Theoretical analysis of the impact of alpha-methyl substitution on experimental spectra. *J Am Chem Soc* 124(19):5325-5332 (2002)
51. Silva RAGD*, et al.*; Discriminating 3(10)-from alpha helices: Vibrational and electronic CD and IR absorption study of related Aib-containing oligopeptides. *Biopolymers* 65(4):229-243 (2002)
52. Maekawa H, Toniolo C, Moretto A, Broxterman QB, & Ge NH; Different spectral signatures of octapeptide 3(10) and alpha-helices revealed by two-dimensional infrared spectroscopy. *J Phys Chem B* 110(12):5834-5837 (2006)
53. Maekawa H, Toniolo C, Broxterman QB, & Ge NH; Two-dimensional infrared spectral signatures of 3(10)- and alpha-helical peptides. *J Phys Chem B* 111(12):3222-3235 (2007)
54. Wang JP; Conformational dependence of anharmonic vibrations in peptides: Amide-I modes in model dipeptide. *J Phys Chem B* 112(15):4790-4800 (2008)
55. Hahn S, Kim S-S, Lee C, & Cho M; Characteristic two-dimensional IR spectroscopic features of antiparallel and parallel beta-sheet polypeptides: simulation studies. *The Journal of Chemical Physics* 123(8):084905-084905 (2005)
56. Karjalainen EL, Ravi HK, & Barth A; Simulation of the Amide I Absorption of Stacked beta-Sheets. *J Phys Chem B* 115(4):749-757 (2011)
57. Kubelka J & Keiderling TA; Differentiation of beta-sheet-forming structures: Ab initio-based simulations of IR absorption and vibrational CD for model peptide and protein beta-sheets. *J Am Chem Soc* 123(48):12048-12058 (2001)
58. Lee C & Cho MH; Local amide I mode frequencies and coupling constants in multiple-stranded antiparallel beta-sheet polypeptides. *J Phys Chem B* 108(52):20397-20407 (2004)
59. Moore WH & Krimm S; Transition Dipole Coupling in Amide I Modes of Beta Polypeptides. *P Natl Acad Sci USA* 72(12):4933-4935 (1975)
60. Maekawa H & Ge NH; Comparative Study of Electrostatic Models for the Amide-I and -II Modes: Linear and Two-Dimensional Infrared Spectra. *J Phys Chem B* 114(3):1434-1446 (2010)
61. Choi JH, Kim JS, & Cho MH; Amide I vibrational circular dichroism of polypeptides: Generalized fragmentation approximation method. *J Chem Phys* 122(17)(2005)
62. DeFlores LP, Ganim Z, Nicodemus RA, & Tokmakoff A; Amide I '-II ' 2D IR Spectroscopy Provides Enhanced Protein Secondary Structural Sensitivity. *J Am Chem Soc* 131(9):3385-3391 (2009)
63. Middleton CT, Buchanan LE, Dunkelberger EB, & Zanni MT; Utilizing Lifetimes to Suppress Random Coil Features in 2D IR Spectra of Peptides. *J Phys Chem Lett* 2(18):2357-2361 (2011)
64. Torii H & Tasumi M; Model-Calculations on the Amide-I Infrared Bands of Globular-Proteins. *J Chem Phys* 96(5):3379-3387 (1992)
65. Wang L*, et al.*; 2DIR Spectroscopy of Human Amylin Fibrils Reflects Stable beta-Sheet Structure. *J Am Chem Soc* 133(40):16062-16071 (2011)
66. Middleton CT, Woys AM, Mukherjee SS, & Zanni MT; Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods* 52(1):12-22 (2010)
67. Shim SH*, et al.*; Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *P Natl Acad Sci USA* 106(16):6614-6619 (2009)
68. Gnanakaran S & Hochstrasser RM; Conformational preferences and vibrational frequency distributions of short peptides in

relation to multidimensional infrared spectroscopy. *J Am Chem Soc* 123(51):12886-12898 (2001)
69. Hamm P, Lim MH, & Hochstrasser RM; Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J Phys Chem B* 102(31):6123-6138 (1998)
70. Decatur SM; Elucidation of residue-level structure and dynamics of polypeptides via isotopp-edited infrared spectroscopy. *Accounts Chem Res* 39(3):169-175 (2006)
71. Barber-Armstrong W, Donaldson T, Wijesooriya H, Silva RAGD, & Decatur SM; Empirical relationships between isotope-edited IR spectra and helix geometry in model peptides. *J Am Chem Soc* 126(8):2339-2345 (2004)
72. Fang C & Hochstrasser RM; Two-dimensional infrared spectra of the C-13=O-18 isotopomers of alanine residues in an alpha-helix. *J Phys Chem B* 109(39):18652-18663 (2005)
73. Bandekar J; Amide Modes and Protein Conformation. *Biochim Biophys Acta* 1120(2):123-143 (1992)
74. Barth A; The infrared absorption of amino acid side chains. *Prog Biophys Mol Bio* 74(3-5):141-173 (2000)
75. Nagarajan S*, et al.*; Differential Ordering of the Protein Backbone and Side Chains during Protein Folding Revealed by Site-Specific Recombinant Infrared Probes. *J Am Chem Soc* 133(50):20335-20340 (2011)
76. Fournier F*, et al.*; Optical fingerprinting of peptides using two-dimensional infrared spectroscopy: proof of principle. *Analytical biochemistry* 374(2):358-365 (2008)
77. Fournier F*, et al.*; Protein identification and quantification by two-dimensional infrared spectroscopy: implications for an all-optical proteomic platform. *P Natl Acad Sci USA* 105(40):15352-15357 (2008)
78. Guo R*, et al.*; Potential for the detection of molecular complexes and determination of interaction geometry by 2DIR: Application to protein sciences. *Faraday Discussions* 150:161-161 (2011)
79. Lindquist BA, Furse KE, & Corcelli SA; Nitrile groups as vibrational probes of biomolecular structure and dynamics: an overview. *Phys Chem Chem Phys* 11(37):8119-8132 (2009)
80. Oh KI*, et al.*; Nitrile and thiocyanate IR probes: Molecular dynamics simulation studies. *J Chem Phys* 128(15)(2008)
81. Choi JH, Oh KI, & Cho MH; Azido-derivatized compounds as IR probes of local electrostatic environment: Theoretical studies. *J Chem Phys* 129(17)(2008)
82. Thielges MC*, et al.*; 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* 115(38):11294-11304 (2011)
83. Hendrickson WA, Horton JR, & Lemaster DM; Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (Mad) - a Vehicle for Direct Determination of 3-Dimensional Structure. *Embo J* 9(5):1665-1672 (1990)
84. Wang L, Xie J, & Schultz PG; Expanding the genetic code. *Annu Rev Bioph Biom* 35:225-249 (2006)
85. Hill JR*, et al.*; Vibrational Dynamics of Carbon-Monoxide at the Active-Site of Myoglobin - Picosecond Infrared Free-Electron Laser Pump-Probe Experiments. *J Phys Chem-Us* 98(43):11213-11219 (1994)
86. Fayer MD; Fast protein dynamics probed with infrared vibrational echo experiments. *Annual Review of Physical Chemistry* 52:315-356 (2001)
87. King JT, Arthur EJ, Brooks CL, & Kubarych KJ; Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *The Journal of Physical Chemistry B* 116(19):5604-5611 (2012)
88. Jansen TLC & Knoester J; Waiting Time Dynamics in Two-Dimensional Infrared Spectroscopy. *Accounts Chem Res* 42(9):1405-1411 (2009)
89. Mukamel S (1999) *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, USA).
90. Hunt NT; 2D-IR spectroscopy: ultrafast insights into biomolecule structure and function. *Chemical Society reviews* 38(7):1837-1848 (2009)
91. Baiz CR & Kubarych KJ; Ultrabroadband detection of a mid-IR continuum by

chirped-pulse upconversion. *Opt Lett* 36(2):187-189 (2011)
92. Zanni MT, Ge NH, Kim YS, & Hochstrasser RM; Two-dimensional IR spectroscopy can be designed to eliminate the diagonal peaks and expose only the crosspeaks needed for structure determination. *Proceedings of the National Academy of Sciences* 98(20):11265-11265 (2001)
93. Khalil M, Demirdoven N, & Tokmakoff A; Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution. *J Phys Chem A* 107(27):5258-5279 (2003)
94. Roberts ST, Loparo JJ, & Tokmakoff A; Characterization of spectral diffusion from two-dimensional line shapes. *J Chem Phys* 125(8)(2006)
95. Gruebele M; The fast protein folding problem. *Annual Review of Physical Chemistry* 50:485-516 (1999)
96. Bredenbeck J & Hamm P; Transient 2D-IR spectroscopy: Towards a molecular movie. *Chimia* 61(1-2):45-46 (2007)
97. Chung HS, Khalil M, Smith AW, & Tokmakoff A; Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev Sci Instrum* 78(6)(2007)
98. Callender R & Dyer RB; Probing protein dynamics using temperature jump relaxation spectroscopy. *Curr Opin Struc Biol* 12(5):628-633 (2002)
99. Dumont C, Emilsson T, & Gruebele M; Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat Methods* 6(7):515-U570 (2009)
100. Gutman M & Nachliel E; The Dynamic Aspects of Proton-Transfer Processes. *Biochim Biophys Acta* 1015(3):391-414 (1990)
101. Fabian H & Naumann D; Methods to study protein folding by stopped-flow FT-IR. *Methods* 34(1):28-40 (2004)
102. Baiz C, Nee M, McCanne R, & Kubarych K; Ultrafast nonequilibrium Fourier-transform two-dimensional infrared spectroscopy. *Opt. Lett.* 33(21):2533-2535 (2008)
103. Xiong W*, et al.*; Transient 2D IR Spectroscopy of Charge Injection in Dye-Sensitized Nanocrystalline Thin Films. *J Am Chem Soc* 131(50):18040-+ (2009)
104. Jones KC, Ganim Z, Peng CS, & Tokmakoff A; Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J Opt Soc Am B* 29(1):118-129 (2012)
105. Jones KC, Ganim Z, & Tokmakoff A; Heterodyne-Detected Dispersed Vibrational Echo Spectroscopy. *J Phys Chem A* 113(51):14060-14066 (2009)
106. Marecek J*, et al.*; A simple and economical method for the production of C-13, O-18-labeled Fmoc-amino acids with high levels of enrichment: Applications to isotope-edited IR studies of proteins. *Org Lett* 9(24):4935-4937 (2007)
107. Lin YS, Shorb JM, Mukherjee P, Zanni MT, & Skinner JL; Empirical Amide I Vibrational Frequency Map: Application to 2D-IR Line Shapes for Isotope-Edited Membrane Peptide Bundles. *J Phys Chem B* 113(3):592-602 (2009)
108. Roy S*, et al.*; Solvent and conformation dependence of amide I vibrations in peptides and proteins containing proline. *J Chem Phys* 135(23)(2011)
109. Van Amerongen H, Valkunas L, & Van Grondelle R (2000) *Photosynthetic Excitons* (World Scientific).
110. Moffitt W; Optical Rotatory Dispersion of Helical Polymers. *J Chem Phys* 25(3):467-478 (1956)
111. Miyazawa T; Perturbation Treatment of the Characteristic Vibrations of Polypeptide Chains in Various Configurations. *J Chem Phys* 32(6):1647-1652 (1960)
112. Krimm S & Abe Y; Intermolecular Interaction Effects in Amide I Vibrations of Beta Polypeptides. *P Natl Acad Sci USA* 69(10):2788-2792 (1972)
113. Sung J & Silbey RJ; Four wave mixing spectroscopy for a multilevel system. *The Journal of Chemical Physics* 115(20):9266-9287 (2001)
114. Faeder SMG & Jonas DM; Two-dimensional electronic correlation and relaxation spectra: Theory and model calculations. *J Phys Chem A* 103(49):10489-10505 (1999)
115. Khalil M, Demirdoven N, & Tokmakoff A; Obtaining absorptive line shapes in two-

dimensional infrared vibrational correlation spectra. *Phys Rev Lett* 90(4)(2003)
116. Golonzka O & Tokmakoff A; Polarization-selective third-order spectroscopy of coupled vibronic states. *J Chem Phys* 115(1):297-309 (2001)
117. Jonas DM; Two-dimensional femtosecond spectroscopy. *Annual Review of Physical Chemistry* 54:425-463 (2003)
118. Torii H; Effects of intermolecular vibrational coupling and liquid dynamics on the polarized Raman and two-dimensional infrared spectral profiles of liquid N,N-dimethylformamide analyzed with a time-domain computational method. *J Phys Chem A* 110(14):4822-4832 (2006)
119. Jansen TL & Knoester J; Nonadiabatic effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide. *J Phys Chem B* 110(45):22910-22916 (2006)
120. Gaigeot MP; Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys Chem Chem Phys* 12(14):3336-3359 (2010)
121. Ingrosso F, Monard G, Farag MH, Bastida A, & Ruiz-Lopez MF; Importance of Polarization and Charge Transfer Effects to Model the Infrared Spectra of Peptides in Solution. *J Chem Theory Comput* 7(6):1840-1849 (2011)
122. Jeon J & Cho M; Direct quantum mechanical/molecular mechanical simulations of two-dimensional vibrational responses: N-methylacetamide in water. *New J Phys* 12(2010)
123. Gorbunov RD, Kosov DS, & Stock G; Ab initio-based exciton model of amide I vibrations in peptides: Definition, conformational dependence, and transferability. *J Chem Phys* 122(22)(2005)
124. Byler DM & Susi H; Examination of the Secondary Structure of Proteins by Deconvolved Ftir Spectra. *Biopolymers* 25(3):469-487 (1986)
125. DeFlores LP, Ganim Z, Ackley SF, Chung HS, & Tokmakoff A; The anharmonic vibrational potential and relaxation pathways of the amide I and II modes of N-methylacetamide. *J Phys Chem B* 110(38):18973-18980 (2006)
126. Pande V; Folding@home: Sustained petaflops for production calculations today, exaflops soon? *Abstr Pap Am Chem S* 242(2011)
127. Pande V; Simulating protein folding in vitro and in vivo. *Biochem Cell Biol* 88(2):409-409 (2010)
128. Lindorff-Larsen K, Piana S, Dror RO, & Shaw DE; How Fast-Folding Proteins Fold. *Science* 334(6055):517-520 (2011)
129. Chung HS, Ganim Z, Jones KC, & Tokmakoff A; Transient 2D IR spectroscopy of ubiquitin unfolding dynamics. *P Natl Acad Sci USA* 104(36):14237-14242 (2007)
130. Cochran AG, Skelton NJ, & Starovasnik MA; Tryptophan zippers: Stable, monomeric beta-hairpins. *P Natl Acad Sci USA* 98(10):5578-5583 (2001)
131. Mukherjee P, Kass I, Arkin IT, & Zanni MT; Structural disorder of the CD3 xi transmembrane domain studied with 2D IR spectroscopy and molecular dynamics simulations. *J Phys Chem B* 110(48):24740-24749 (2006)
132. Remorino A, Korendovych IV, Wu YB, DeGrado WF, & Hochstrasser RM; Residue-Specific Vibrational Echoes Yield 3D Structures of a Transmembrane Helix Dimer. *Science* 332(6034):1206-1209 (2011)
133. Hamm P, Helbing J, & Bredenbeck J; Two-dimensional infrared spectroscopy of photoswitchable peptides. *Annual Review of Physical Chemistry* 59:291-317 (2008)
134. Kaziannis S*, et al.*; Femtosecond to microsecond photochemistry of a [FeFe]hydrogenase enzyme model compound. *The journal of physical chemistry. B* 114(46):15370-15379 (2010)
135. Chung HS & Tokmakoff A; Temperature-de pendent downhill unfolding of ubiquitin. II. Modeling the free energy surface. *Proteins* 72(1):488-497 (2008)
136. Chung HS & Tokmakoff A; Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins* 72(1):474-487 (2008)
137. Bredenbeck J, Helbing J, Nienhaus K, Nienhaus GU, & Hamm P; Protein ligand migration mapped by nonequilibrium 2D-IR

exchange spectroscopy. *P Natl Acad Sci USA* 104(36):14243-14248 (2007)

138. Finkelstein IJ, Ishikawa H, Kim S, Massari AM, & Fayer MD; Substrate binding and protein conformational dynamics measured by 2D-IR vibrational echo spectroscopy. *P Natl Acad Sci USA* 104(8):2637-2642 (2007)
139. Bandaria JN*, et al.*; Characterizing the dynamics of functionally relevant complexes of formate dehydrogenase. *P Natl Acad Sci USA* 107(42):17974-17979 (2010)
140. Chung JK, Thielges MC, & Fayer MD; Dynamics of the folded and unfolded villin headpiece (HP35) measured with ultrafast 2D IR vibrational echo spectroscopy. *P Natl Acad Sci USA* 108(9):3578-3583 (2011)
141. Urbanek DC, Vorobyev DY, Serrano AL, Gai F, & Hochstrasser RM; The Two-Dimensional Vibrational Echo of a Nitrile Probe of the Villin HP35 Protein. *J Phys Chem Lett* 1(23):3311-3315 (2010)
142. Fang C*, et al.*; Two-dimensional infrared spectra reveal relaxation of the nonnucleoside inhibitor TMC278 complexed with HIV-1 reverse transcriptase. *P Natl Acad Sci USA* 105(5):1472-1477 (2008)
143. Ganim Z, Jones KC, & Tokmakoff A; Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy. *Phys Chem Chem Phys* 12(14):3579-3588 (2010)
144. Ganim Z, Tokmakoff A, & Vaziri A; Vibrational excitons in ionophores: experimental probes for quantum coherence-assisted ion transport and selectivity in ion channels. *New J Phys* 13:113030 (2011)