

# Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution

Carlos R. Baiz, Chunte Sam Peng, Mike E. Reppert, Kevin C. Jones,  
and Andrei Tokmakoff

Department of Chemistry, Massachusetts Institute of Technology, Cambridge MA 02139

## Abstract

We present a method to quantitatively determine the secondary structure composition of globular proteins using coherent two-dimensional infrared (2DIR) spectroscopy of backbone amide I vibrations ( $1550 - 1720 \text{ cm}^{-1}$ ). Sixteen proteins with known crystal structures were used to construct a library of 2DIR spectra, and the fraction of residues in  $\alpha$ -helix,  $\beta$ -sheet, and unassigned conformations was determined by singular value decomposition (SVD) of the measured two-dimensional spectra. The method was benchmarked by removing each individual protein from the set and comparing the composition extracted from 2DIR against the composition determined from the crystal structures. To highlight the increased structural content extracted from 2DIR spectra a similar analysis was also carried out using conventional infrared absorption of the proteins in the library.

## 1. Motivation

Modern structural biology seeks to establish a molecular understanding of biological processes by exploring structure-function relationships in biomolecules. The mechanisms for protein folding, protein-protein interactions, ligand binding and catalysis are fundamentally linked to structure. Membrane proteins, in particular, drive key biological processes such as cell-signaling, and ion transport through the cell membrane. Modern tools such as cryo-electron-microscopy or x-ray crystallography enable detailed measurements of protein structure on multiple length-scales; however, structures are usually measured under non-biological conditions where key dynamic effects that lead to protein stability and function are not captured,[1] and despite the abundance, diversity, and biological importance of membrane proteins, relatively few crystal structures have been solved due to the difficulty in crystallizing the proteins.[2] Aside from NMR spectroscopy, most bio-analytical tools in use today are not sensitive to protein structure, heterogeneity, or conformational dynamics.[3] Consequently, specific classes of proteins such as fibrous proteins, intrinsically disordered proteins, gels, amyloids, and aggregates, have been particularly difficult to characterize. Therefore, even a simple and reliable analysis to quantify the percentage of amino-acids in  $\alpha$ -helix or  $\beta$ -sheet conformations of proteins in solution would constitute a particularly useful bio-structural tool.

Infrared spectroscopy in the amide I region has been used to measure protein structure for over four decades. The vibrational modes in these regions involve primarily backbone C=O and N-H vibrations that are relatively sensitive to secondary structure and are largely free from the influence of side-chain absorptions. To date, however, most infrared measurements have offered either qualitative information or have required significant efforts to vibrationally isolate individual residues through isotope labeling.[4] Attempts to extract structural information from amide-I and II absorption bands in proteins have relied on complex deconvolution and fitting to analyze largely featureless absorption bands.[5-8] While these methods offer semi-quantitative agreement with crystal structures, the results are heavily dependent on the specific set of frequencies and fitting functions used, which greatly limits the usefulness of linear infrared spectroscopy as a structural tool.[9-12] In this paper we present multidimensional spectroscopy as a new technique to measure structure in solution. The method relies upon the specific signatures of secondary structures in the infrared spectrum which allow for a quantitative decomposition of a protein spectrum into combinations of spectra corresponding to individual secondary structures. Our method assumes that each individual secondary structure has a unique spectrum associated with it, and that the total protein spectrum is the sum of each secondary structure spectrum linearly weighted by its content. Although these assumptions are not entirely correct based on our understanding of vibrational coupling and the overlap between secondary structure spectra, the accuracy of our results justify them. Multidimensional spectroscopy provides more unique spectral information than FTIR, which explains why 2DIR is a more accurate predictor of structure.

## 2. Background: Amide-I Multidimensional Infrared Spectroscopy

Over the last ten years, coherent multidimensional spectroscopy has been developed as a technique to study the structure and dynamics of complex systems in solution.[13-18] In particular, two-dimensional infrared (2DIR) spectroscopy has the ability to directly measure the structure, static disorder, conformational flexibility, and solvent exposure of residues within a protein.[19-21] To date, 2DIR spectroscopy has offered new insights into protein structure and heterogeneity,[22] protein dynamics,[23] protein-protein interactions,[24] and ligand binding.[25] However, the amount of structural information extracted from 2DIR has remained largely qualitative, and in most cases the equilibrium structure of the protein under investigation is known *a priori*. In this work, we seek to exploit the structural-sensitivity of 2DIR in order to develop a quantitative method for measuring secondary structure in solution.

Amide-I vibrations involve primarily combinations of backbone carbonyl stretches.[26, 27] The 'local' carbonyl stretches, corresponding to individual amino acids within the protein, form the basis for the delocalized normal modes.[28] The frequencies of the normal modes are given by the frequencies of the local C=O stretches as well as the magnitude of the couplings between these local oscillators. Coupling patterns report on the secondary structure and conformational disorder of the residues: structured regions exhibit highly regular coupling patterns whereas disordered regions exhibit a random coupling pattern.[29] Each oscillator within a particular secondary structure is coupled strongly to other oscillators in the secondary structure, giving rise delocalized normal

modes. The frequency of the delocalized vibrations tends to separate into particular ranges:  $\beta$ -sheets exhibit two peaks near 1630 and 1680  $\text{cm}^{-1}$  whereas  $\alpha$ -helices and unstructured regions appear near 1650  $\text{cm}^{-1}$ . [30, 31] Despite the secondary structure-sensitivity of the peak frequencies, conformational disorder and solvent exposure render the absorption bands broad and largely featureless.

Analogous to NMR, a 2DIR spectrum can be interpreted as follows: the vibrations are first excited, or labeled, by two infrared pulses, and after a certain waiting time, a third pulse either induces absorption to higher excited states or stimulates emission down to the ground state. The excitation frequency (horizontal axis) and a detection frequency (vertical axis) are then combined in a two-dimensional plot. Diagonal peaks are due to excitation and detection at the same frequency, and cross peaks arise from excitation of one mode, and detection of a different mode. Cross peaks are observed only if two vibrations share common atoms, or in this context, common residues. For proteins, 2DIR spectroscopy offers enhanced structure-sensitivity by spreading the spectral content onto two frequency axes. For example, in the case of  $\beta$ -sheets, off-diagonal features are observed between the two main peaks at 1630 and 1680  $\text{cm}^{-1}$ , while no cross peaks are observed between  $\alpha$ -helix and  $\beta$ -sheet vibrations since the two modes primarily involve different residues. Further analysis of the main spectral features associated with each secondary structure is provided in the discussion section below.

### 3. Methods

#### a. Protein Selection and Sample Preparation

The set of proteins used in this study (Figure 1) was selected to span a wide variety of structures ranging from mainly  $\alpha$ -helical proteins (Myoglobin) to mainly  $\beta$ -sheet proteins ( $\gamma$ -globulins) while minimizing structural redundancies. The solubility, robustness, and commercial availability of each protein were also considered. All samples were purchased from Sigma-Aldrich (St. Louis, MO) and were used without further purification. The samples were hydrogen/deuterium exchanged in  $\text{D}_2\text{O}$  (Cambridge Isotopes, Andover MA) at 60° C for 1 hour, and lyophilized after freezing in liquid nitrogen. Solid protein samples were stored at -20° C to prevent degradation. The proteins were then re-dissolved in pure  $\text{D}_2\text{O}$  at neutral pH\* to a final concentration of 20 mg/ml. Due to low solubility at neutral pH\*, Fibrinogen, Immunoglobulin G and Insulin samples were dissolved in pH\*=2 DCl solution. The solutions were placed in a temperature-controlled sample cell equipped with two  $\text{CaF}_2$  windows and a 50  $\mu\text{m}$  Teflon spacer. To prevent degradation of the samples, spectra were acquired within approximately one hour after preparing each solution, during this period the samples were kept at room temperature under a dry air atmosphere. All experiments were carried out at 25° C.

#### b. Infrared Spectroscopy

The two-dimensional spectrometer has been described in detail elsewhere.[32, 33] Briefly, 90-fs mid-infrared pulses are generated by an optical parametric amplifier/difference frequency generation setup pumped by a regeneratively amplified Ti:Sapphire laser. Three IR pulses are overlapped at the

sample in a box geometry, and the signal electric field is interferometrically measured in a phase-matched direction by overlapping the emitted signal with a reference pulse. The relative polarization of the excitation and detection pulses is set to  $90^\circ$  so as to enhance off-diagonal features in the spectrum. The excitation frequency is measured by scanning the first two IR pulses and Fourier-transforming the resulting oscillations along the first time delay ( $t_1$ ); the detection frequency is measured using a spectrometer equipped with a  $64 \times 2$ -pixel mercury-cadmium-telluride (MCT) array. The excitation time axis ( $t_1$ ) is scanned from 0 to 3 ps in steps of 4 fs; the frequency resolution along the detection axis is approximately  $4.5 \text{ cm}^{-1}$  per pixel. Dispersed pump-probe spectra, collected in the parallel polarization condition, were used to independently phase the rephasing and non-rephasing spectra. Following Fourier-transformation and phasing, all correlation spectra were interpolated to  $2 \text{ cm}^{-1}$  uniformly-spaced frequency axes for subsequent SVD analysis. An absorption spectrum of each sample was collected immediately before each 2DIR measurement using a Nicolet 380 Fourier-transform infrared spectrometer.  $\text{D}_2\text{O}$  baseline subtraction of FTIR spectra was carried out by fitting a line between two frequency points at  $1550$  and  $1720 \text{ cm}^{-1}$ . Two-dimensional spectra were not corrected for HOD absorption. Since the 2DIR signal depends on the fourth order of the transition dipoles, the small transition dipole of the HOD bend vibration makes its 2DIR peak negligible relative to the amide-I C=O vibrations.

### c. Structure Assignment and Singular Value Decomposition

The x-ray structure of each protein was downloaded from the Brookhaven Protein Databank[34] (PDB) and the secondary structure assignment was calculated with the DSSP program.[35] DSSP uses the atomic coordinates and hydrogen-bonding patterns to assign each residue to the following structural elements:  $\beta$ -bridge, extended  $\beta$ -strand, 3-10 helix,  $\alpha$ -helix, hydrogen bonded turns, and bends. For SVD analysis we combined the  $\beta$ -bridge and extended  $\beta$ -strand into a single structural element  $\beta$ -sheet, and combined the  $\alpha$ -helix and 3-10 helix residues into an effective  $\alpha$ -helix, all other residues were labeled *unassigned*.

Singular value decomposition (SVD) is a matrix-diagonalization technique used to describe the data in terms of a set of principal component spectra. Within this method a matrix of spectra  $M$ , is decomposed into three matrices such that,

$$M = USV^T, \quad (1)$$

where the superscript  $T$  indicates a transpose. The SVD process is analogous to eigenvalue decomposition: Column-wise the matrix  $U$  contains a set of orthogonal basis spectra (SVD components), row-wise  $V$  contains the set of coefficients associated with the projection of each input spectrum ( $M$ ) onto the individual basis spectra and  $S$  is a diagonal matrix containing a set of weighting coefficients that represent the relative contribution of each basis spectrum in  $U$  to the input data set  $M$ . In our analysis FTIR spectra are input as one-dimensional arrays (vectors) where each entry corresponds to the absorbance at each frequency. Likewise, each two-dimensional spectrum is column-wise reshaped into a one dimensional array.

Some of the U columns, V columns, and S diagonal values are zero or near zero. We discard these components and only keep the nonzero parts. As a result of this rejection, S becomes a square, invertible matrix.

If we assume that the protein spectra,  $M$ , are the linear combination of the secondary structure spectra weighted by their content, then we can decompose the spectra into:

$$\Sigma P^T = M \quad (2)$$

where  $\Sigma = \{m^{(\beta)}, m^{(\alpha)}, m^{(\text{una})}\}$  and  $P = \{p^{(\beta)}, p^{(\alpha)}, p^{(\text{una})}\}$ . The columns of  $\Sigma$  are the pure spectra of a  $\beta$ -sheet,  $\alpha$ -helix, and unassigned structure, respectively. The columns of  $P$  contain the fraction of residues per protein in the  $\beta$ -sheet,  $\alpha$ -helix, and unassigned conformation calculated using the DSSP program.

To transform the SVD results from equation (1) into the spectra and fractions in equation (2), we insert  $XX^T=I$  into equation (1):

$$M = USXX^{-1}V^T. \quad (3)$$

and solve for  $X^{-1}$  such that:

$$X^{-1}V^T = P^T. \quad (4)$$

$X^{-1}$  can then be inverted and multiplied to give  $\Sigma$ :

$$USX = \Sigma. \quad (5)$$

To assign the secondary structure  $P^{(u)} = \{p^{(\beta)}(u), p^{(\alpha)}(u), p^{(\text{una})}(u)\}$  of an “unknown” spectrum  $m^{(u)}$ , we solve the equation:

$$\Sigma \left( P^{(u)} \right)^T = m^{(u)}. \quad (6)$$

Substitution of previous equations allows us to solve for  $P^{(u)}$  with:

$$P^{(u)} = P^T V S^{-1} U^T m^{(u)}. \quad (7)$$

The SVD procedure is cross-validated for each sample by removing a specific spectrum from the initial set, building an SVD basis with the remaining spectra, and using the removed spectrum as the unknown. Finally the predicted percentage of secondary structure is compared with the percentage computed from the PDB structure.

## 4. Results

### a. Interpretation of Amide-I Spectra

Experimental FTIR and 2DIR spectra are shown in Figures 2 and 3 respectively. Qualitatively, absorption spectra associated with primarily  $\alpha$ -helical proteins are characterized by a single band centered near  $1650\text{ cm}^{-1}$  with an approximate diagonal width of  $50\text{ cm}^{-1}$ , while proteins composed of primarily  $\beta$ -sheet show two peaks centered near  $1620$  and  $1680\text{ cm}^{-1}$  resulting from vibrations whose main transition dipoles lie perpendicular and parallel to the  $\beta$ -strands respectively. The amplitude ratio of  $\nu_{\text{parallel}}$  to  $\nu_{\text{perp}}$  reports on the size of the  $\beta$ -sheet.[36] In this paper we limit the analysis to the number of residues in  $\alpha$ -helix and  $\beta$ -sheet conformations. In principle, 2DIR spectroscopy is sensitive to the size and shape of secondary structures as well, but a larger and more diverse protein library would be required for such analysis. Mixed  $\alpha/\beta$  proteins show a combination of the features associated with helix, sheet and unstructured regions where the relative contributions of the individual features depends on the fraction of residues that compose each secondary structure. Peaks assigned to unstructured regions are centered near  $1640\text{ cm}^{-1}$  and thus tend to overlap with the  $\alpha$ -helix peaks. This overlap increases the difficulty of unambiguously assigning  $\alpha$ -helix and unstructured conformations in proteins. In general, diagonal linewidths report on the static heterogeneity (disorder) of the amide-I frequencies whereas the anti-diagonal linewidths report on the sub-picosecond dynamic fluctuations that are a result of protein-solvent interactions; and the ratio of diagonal to anti-diagonal linewidth serves as a measure of structural rigidity and solvent exposure.[23] In this context it should be pointed out that since vibrations are highly delocalized; residue-level resolution can be achieved through isotope labeling.[4]

The amide-I vibrations shift approximately  $5\text{-}10\text{ cm}^{-1}$  upon deuteration; while this is not a large shift relative to the peak width, it could likely contribute to the errors in the assigned structure, it is therefore important to ensure that proteins are fully deuterated. The Amide-II modes, centered near  $1550\text{ cm}^{-1}$  and involving primarily C=N stretch and N-H bend, tend to shift by approximately  $100\text{ cm}^{-1}$  to  $1450\text{ cm}^{-1}$  upon deuteration (Amide-II'), therefore it is relatively straightforward to check for partial deuteration by comparing the intensity of the  $1550\text{ cm}^{-1}$  band to the  $1450\text{ cm}^{-1}$  band.[31] All the samples used here showed negligible amide-II peak intensities compared to amide-II' following H/D exchange. It is also important to note that certain side-chain absorptions overlap the amide-I peaks. Different protonation states will shift these absorption bands and may thus interfere with the structural assignment of spectra.

## b. Singular Value Decomposition

Once spectra are decomposed and structure projections are computed (Eqs. 1 and 4), it is informative to reconstruct the 2DIR spectra corresponding to purely  $\alpha$ -helix,  $\beta$ -sheet components (Eq. 3). Note that since the unassigned structure percentages are not linearly independent from the helix and sheet, the number of components in the analysis is effectively reduced to two. Figure 4 below shows the reconstructed SVD spectra; the  $\alpha$ -helix spectrum shows a single peak centered at  $1650\text{ cm}^{-1}$  whereas the  $\beta$ -sheet spectrum exhibits two peaks near  $1630$  and  $1670\text{ cm}^{-1}$  with the corresponding cross-peaks, giving the spectrum the characteristic “Z-shape” associated with primarily  $\beta$ -sheet proteins. SVD spectra are in agreement with our spectral assignment described in

the previous section as well as with experimental and simulated spectra of idealized structures published previously.[37]

### c. Secondary Structure Prediction

The SVD method is validated by removing each protein from the initial set, creating an SVD basis with the remaining structures, and using the new basis to analyze the “unknown” structure (see Methods). Figure 5 shows the fraction of each secondary structure predicted from SVD analysis along with the fractions obtained from the x-ray structure. Overall there is an excellent correlation between the predicted structures in solution with those derived from the crystal structure. The largest error is observed in Myoglobin, perhaps due to the fact that this protein lies at the edge of the conformational space sampled by the protein library. Alternatively, the large error may suggest that this sample was somehow anomalous. The root-mean-squared errors in comparing the SVD prediction with the x-ray structure are 12.5, 9.2 and 9.1% for  $\alpha$ -helix,  $\beta$ -sheet, and unstructured components respectively; however removing Myoglobin from the set reduces the values to 6.7, 7.7, and 8.2% respectively. Naturally, the  $\beta$ -sheet component is the most distinct and thus the errors in predicting the percentage of residues in this conformation tend to be smaller. Distinguishing between  $\alpha$ -helices and unstructured regions has remained challenging for amide-I infrared spectroscopy as both peaks appear in the same region of the spectrum. For comparison, we performed the same SVD decomposition on the measured FTIR spectra. The RMS errors obtained for an FTIR-derived SVD basis are: 19.5, 8.3 and 21.5% respectively, consistent with published benchmarks.[38-40] This simple comparison highlights the structural-sensitivity gain associated with projecting the spectral information onto two frequency axes and measuring the diagonal as well as the anti-diagonal linewidths. Incomplete background subtraction is one of the largest sources of error in FTIR spectroscopy. Differences in the H<sub>2</sub>O background, however do not affect the 2DIR data since the non-linear interactions largely suppress the broad background signal, and the spectra are corrected for any distortions due to absorption of the IR pulses by H<sub>2</sub>O.

## Discussion

Circular dichroism (CD) spectroscopy is a standard method to measure protein conformation in solution. Although the light-matter interactions of amide-I IR spectroscopy and ultraviolet CD spectroscopy are quite different, both methods measure the structure of the backbone of proteins by exciting transitions that are delocalized over multiple amide-I units and where the local structure of the units affects the absorption frequencies and intensities.[41] The errors associated with secondary structure determination by 2DIR are comparable to those obtained from circular dichroism. The accepted RMSE values for CD are: 9%  $\alpha$ -helix, 12% anti-parallel  $\beta$ -sheet, 8% parallel  $\beta$ -sheet, 7%  $\beta$ -turn, and 9% unassigned.[42] These results show that 2DIR spectroscopy is a viable alternative to CD spectroscopy with the added advantage of ultrafast time resolution and the ability to isotope-label individual residues for increased structural resolution.

The observed errors between structure determination techniques in solution compared to crystal structure reflect, in part, on the structural heterogeneity and the inherent differences in structure that

are induced upon crystallization: it is well-known that water plays a central role in maintaining the balance between entropic and enthalpic contributions that determine the structure and conformational flexibility needed for protein function.[43]

Another source of error in our measurements is the protein library itself. To minimize errors, the SVD basis should reflect the region of conformational space of the unknown protein to be analyzed. For example, constructing a basis set of purely  $\beta$ -sheet proteins and using such basis to analyze an  $\alpha$ -helical protein would likely result in a large error. A particularly large basis set is unlikely to capture the structural details of an individual proteins and therefore likely to produce an “averaged-out” measure of structure. To circumvent this limitation, it may be possible to perform a multi-level SVD analysis where the protein is initially screened using a diverse basis set, and based on the initial results the SVD basis is then further tailored to include only proteins that are similar in structure to the unknown protein. Such analysis would, however, be susceptible to errors in biasing the basis set towards a particular structure that may not accurately reflect the structure of the unknown protein.

Computational models and symmetry considerations for idealized  $\beta$ -sheets suggest that 2DIR spectroscopy is sensitive to the size and type of  $\beta$ -structure.[28, 36] To date a systematic measurement of pure  $\beta$ -sheet proteins has not been carried out. More generally, quantifying the spectral signatures of various structural motifs, with particular attention to heterogeneity and site disorder, will further enhance the capabilities of 2DIR spectroscopy as a structural technique. Similarly, the  $\sim 12\text{ cm}^{-1}$  shift that occurs upon deuteration, also suggests that additional information can be derived from the 2DIR of partially deuterated proteins, analogous to amide-I/II 2DIR hydrogen/deuterium exchange experiments,[31] which report directly on the solvent exposure of secondary structures within a protein. Finally, an underlying assumption of the analysis presented here is that the spectra of individual secondary structures are additive. However, the delocalized nature of amide-I vibrations coupled with the inherent nonlinearity of 2DIR spectroscopy, renders the peaks sensitive to tertiary contacts. Preliminary simulations suggest that the couplings between secondary structures are relatively weak ( $\sim 2\text{ cm}^{-1}$ ), and thus peak distortions due to coupling between secondary structures is only a small source of error within the SVD analysis. Future studies will benefit from the tertiary structure information present in the spectra but will require a significant effort in developing a large basis set as well as computational modeling to understand the spectral signatures of particular tertiary contacts.

## 5. Summary and Outlook

We have presented a general method to measure protein structure in solution based on coherent two-dimensional infrared spectroscopy. Spectra are analyzed by singular-value decomposition, a well-known analysis method which remains free from the inherent subjectivity and bias of phenomenological fitting models. The model was tested on a set of sixteen commercially-available globular proteins with well-defined structures. The results presented here show that 2DIR spectroscopy is capable of quantitatively determining the structure of *stock* proteins without the need for labeling. One key advantage, however, of 2DIR spectroscopy lies in its picosecond time resolution and ability to spread the spectral information over two axes by correlating pump and



probe frequencies. Combined with a laser-induced temperature-jump[44-46] the method can be used to determine protein folding/unfolding pathways in solution and expand the current understanding of protein folding and function.

**Acknowledgment:** Funding for this project was provided by the National Science Foundation (CHE-0616575 and CHE-0911107) as well as a grant from Agilent Technologies.

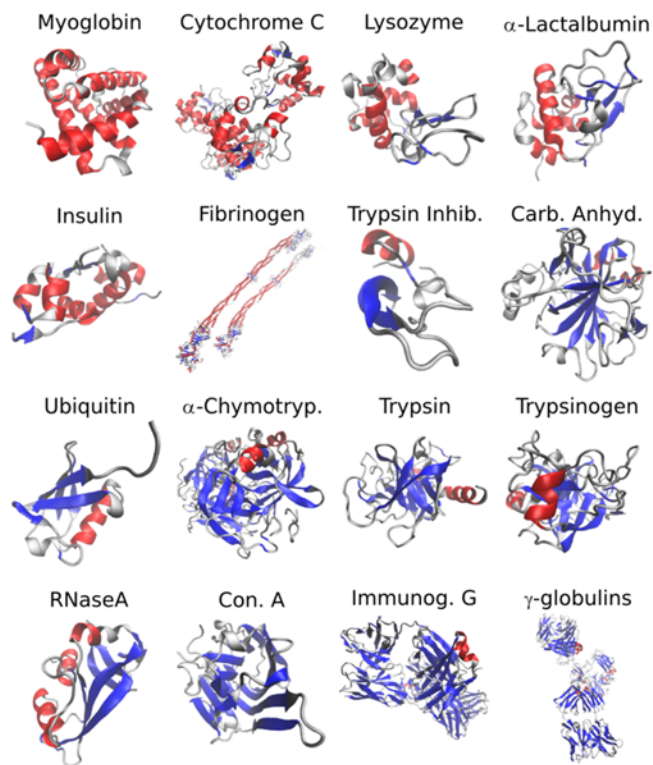


Figure 1: PDB Structures of the proteins used for SVD analysis.  $\alpha$ -helix and  $\beta$ -sheet structures are colored in red and blue respectively. The proteins are arranged in order of increasing  $\beta$ -sheet contents as calculated by the DSSP program.

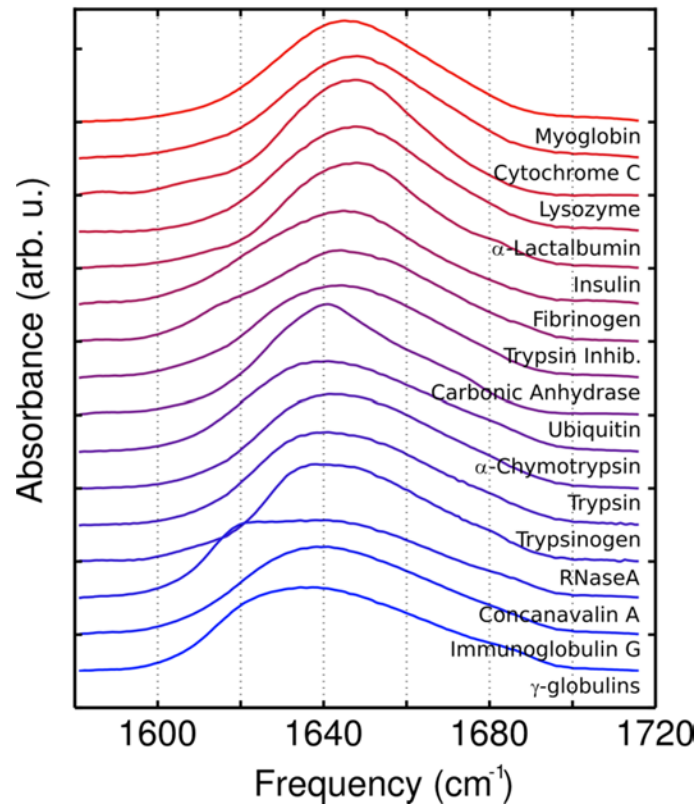


Figure 2: Amide-I linear absorption spectra. Peaks are normalized to the area in the 1580 to 1720  $\text{cm}^{-1}$  and vertically offset for clarity.

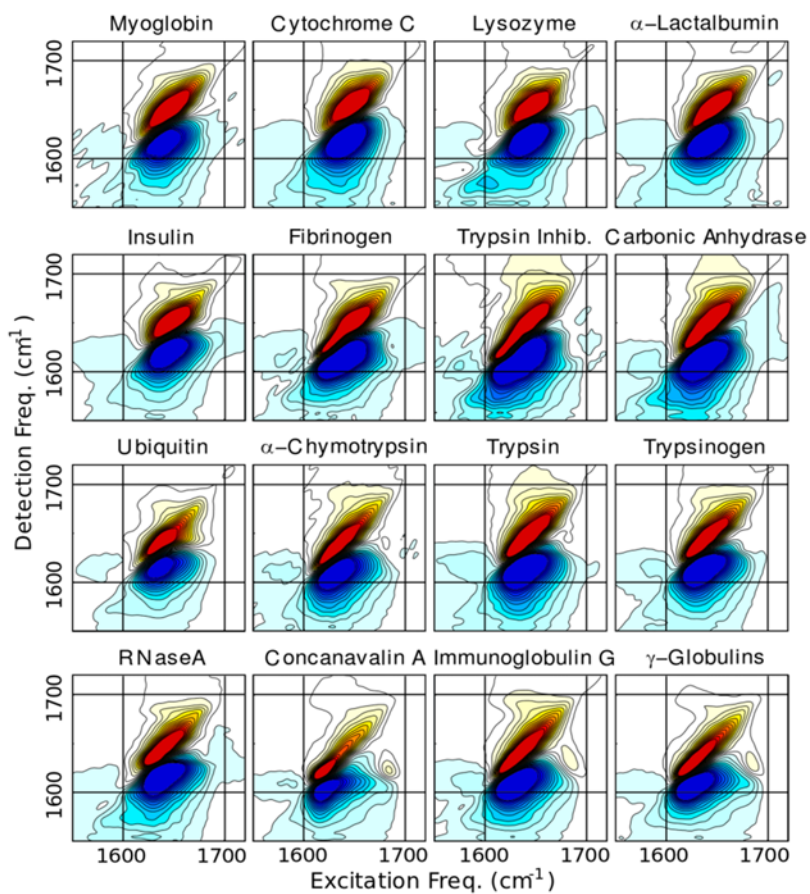


Figure 3: Correlation 2DIR spectra of the proteins acquired in the perpendicular polarization geometry. Contours are plotted from  $\pm 50\%$  of the maximum amplitude in 4% intervals and spectra are arranged in order of increasing  $\beta$ -sheet content (see Figure 1).

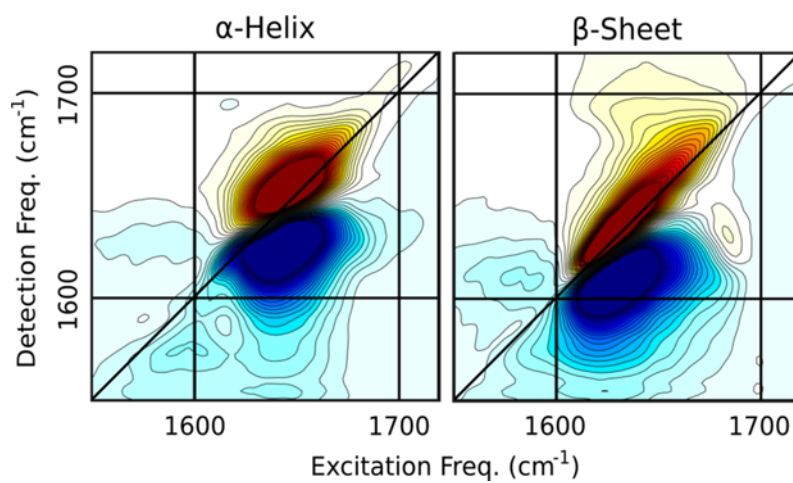


Figure 4: Spectra reconstructed from the SVD decomposition. Contours are plotted from  $\pm 50\%$  of the maximum amplitude in 4 % intervals and spectra are arranged in order of increasing  $\beta$ -sheet content (see Figure 1).

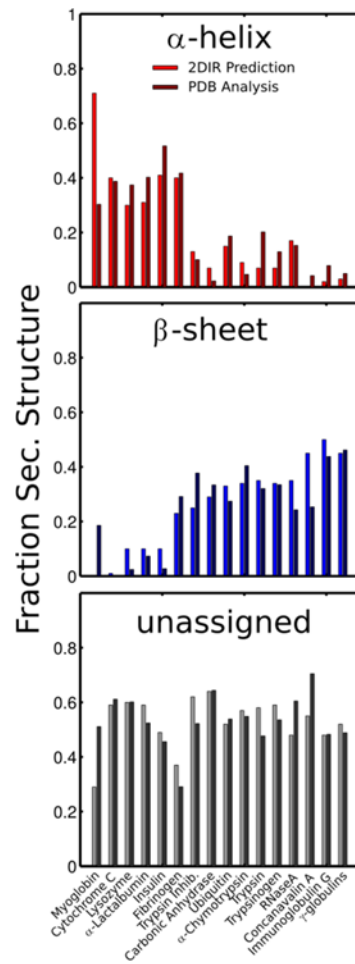


Figure 5: Comparison between amount of secondary structure predicted by SVD analysis of the 2DIR spectra and the amount extracted from the x-ray structures using the DSSP program (see text for details).

## References

1. Frank, J., *Single-particle imaging of macromolecules by cryo-electron microscopy*. Annual Review of Biophysics and Biomolecular Structure, 2002. **31**: p. 303-319.
2. White, S.H., *Biophysical dissection of membrane proteins*. Nature, 2009. **459**(7245): p. 344-346.
3. Braun, W., *Distance Geometry and Related Methods for Protein-Structure Determination from Nmr Data*. Quarterly Reviews of Biophysics, 1987. **19**(3-4): p. 115-157.
4. Middleton, C.T., et al., *Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy*. Methods, 2010. **52**(1): p. 12-22.
5. Byler, D.M. and H. Susi, *Examination of the Secondary Structure of Proteins by Deconvolved Ftir Spectra*. Biopolymers, 1986. **25**(3): p. 469-487.
6. Ruegg, M., V. Metzger, and H. Susi, *Computer Analyses of Characteristic Infrared Bands of Globular Proteins*. Biopolymers, 1975. **14**(7): p. 1465-1471.
7. Fasman, G.D., *Protein Conformational Prediction*. Trends in Biochemical Sciences, 1989. **14**(7): p. 295-299.
8. Goormaghtigh, E., J.M. Ruyschaert, and V. Raussens, *Evaluation of the information content in infrared spectra for protein secondary structure determination*. Biophysical Journal, 2006. **90**(8): p. 2946-2957.
9. Jackson, M. and H.H. Mantsch, *The Use and Misuse of Ftir Spectroscopy in the Determination of Protein-Structure*. Critical Reviews in Biochemistry and Molecular Biology, 1995. **30**(2): p. 95-120.
10. Jackson, M. and H.H. Mantsch, *Artifacts Associated with the Determination of Protein Secondary Structure by Atr-Ir Spectroscopy*. Applied Spectroscopy, 1992. **46**(4): p. 699-701.
11. Hubner, W., H.H. Mantsch, and H.L. Casal, *Beware of Frequency-Shifts*. Applied Spectroscopy, 1990. **44**(4): p. 732-734.
12. Surewicz, W.K., H.H. Mantsch, and D. Chapman, *Determination of Protein Secondary Structure by Fourier-Transform Infrared-Spectroscopy - a Critical-Assessment*. Biochemistry, 1993. **32**(2): p. 389-394.
13. Hamm, P., M.H. Lim, and R.M. Hochstrasser, *Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy*. Journal of Physical Chemistry B, 1998. **102**(31): p. 6123-6138.
14. Khalil, M., N. Demirdoven, and A. Tokmakoff, *Coherent 2D IR spectroscopy: Molecular structure and dynamics in solution*. Journal of Physical Chemistry A, 2003. **107**(27): p. 5258-5279.
15. Fayer, M.D., *Dynamics of Liquids, Molecules, and Proteins Measured with Ultrafast 2D IR Vibrational Echo Chemical Exchange Spectroscopy*. Annual Review of Physical Chemistry, 2009. **60**(1): p. 21.
16. Ganim, Z., et al., *Amide I two-dimensional infrared Spectroscopy of proteins*. Accounts of Chemical Research, 2008. **41**(3): p. 432-441.
17. Baiz, C.R., et al., *Two-Dimensional Infrared Spectroscopy of Metal Carbonyls*. Accounts of Chemical Research, 2008. **In Press**.
18. Roberts, S.T., K. Ramasesha, and A. Tokmakoff, *Structural Rearrangements in Water Viewed Through Two-Dimensional Infrared Spectroscopy*. Accounts of Chemical Research, 2009. **42**(9): p. 1239-1249.
19. Fayer, M.D., *Fast protein dynamics probed with infrared vibrational echo experiments*. Annual Review of Physical Chemistry, 2001. **52**: p. 315-356.

20. Kim, Y.S., et al., *2D IR provides evidence for mobile water molecules in beta-amyloid fibrils*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(42): p. 17751-17756.
21. Shim, S.H., et al., *Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(16): p. 6614-6619.
22. Ghosh, A., et al., *Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(15): p. 6115-6120.
23. Smith, A.W., et al., *Melting of a beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations*. Journal of Physical Chemistry B, 2010. **114**(34): p. 10913-10924.
24. Ganim, Z., K.C. Jones, and A. Tokmakoff, *Insulin dimer dissociation and unfolding revealed by amide I two-dimensional infrared spectroscopy*. Physical Chemistry Chemical Physics, 2010. **12**(14): p. 3579-3588.
25. Bredenbeck, J., et al., *Protein ligand migration mapped by nonequilibrium 2D-IR exchange spectroscopy*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(36): p. 14243-14248.
26. Kim, Y., et al., *Cyanide binding and active site structure in heme-copper oxidases: Normal coordinate analysis of iron-cyanide vibrations of  $a(3)(2+)$ CN(-) complexes of cytochromes  $ba(3)$  and  $aa(3)$* . Biospectroscopy, 1998. **4**(1): p. 1-15.
27. Barth, A., *Infrared spectroscopy of proteins*. Biochimica Et Biophysica Acta-Bioenergetics, 2007. **1767**(9): p. 1073-1101.
28. Chung, H.S. and A. Tokmakoff, *Visualization and characterization of the infrared active amide I vibrations of proteins*. Journal of Physical Chemistry B, 2006. **110**(6): p. 2888-2898.
29. Moore, W.H. and S. Krimm, *Transition Dipole Coupling in Amide I Modes of Beta Polypeptides*. Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(12): p. 4933-4935.
30. Ganim, Z., et al., *Amide I two-dimensional infrared Spectroscopy of proteins*. Accounts of Chemical Research, 2008. **41**(3): p. 432-441.
31. DeFlores, L.P., et al., *Amide I 'II' 2D IR Spectroscopy Provides Enhanced Protein Secondary Structural Sensitivity*. Journal of the American Chemical Society, 2009. **131**(9): p. 3385-3391.
32. Jones, K.C., Z. Ganim, and A. Tokmakoff, *Heterodyne-Detected Dispersed Vibrational Echo Spectroscopy*. Journal of Physical Chemistry A, 2009. **113**(51): p. 14060-14066.
33. Chung, H.S., et al., *Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics*. Review of Scientific Instruments, 2007. **78**(6).
34. Valafar, H., F. Valafar, and J.H. Prestegard, *One and two dimensional statistical data-mining of protein databank for crystallization patterns of proteins*. Metmbs'01: Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, 2001: p. 67-72.
35. Joosten, R.P., et al., *A series of PDB related databases for everyday needs*. Nucleic Acids Research, 2011. **39**: p. D411-D419.
36. Smith, A.W., et al., *Two-dimensional infrared spectroscopy of beta-sheets and hairpins*. Biophysical Journal, 2004. **86**(1): p. 619a-619a.
37. Cheatum, C.M., A. Tokmakoff, and J. Knoester, *Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy*. Journal of Chemical Physics, 2004. **120**(17): p. 8201-8215.

38. Pribic, R., *Principal Component Analysis of Fourier-Transform Infrared and/or Circular-Dichroism Spectra of Proteins Applied in a Calibration of Protein Secondary Structure*. Analytical Biochemistry, 1994. **223**(1): p. 26-34.
39. Pribic, R., et al., *Protein Secondary Structure from Fourier-Transform Infrared and or Circular-Dichroism Spectra*. Analytical Biochemistry, 1993. **214**(2): p. 366-378.
40. Rahmelow, K. and W. Hubner, *Secondary structure determination of proteins in aqueous solution by infrared spectroscopy: A comparison of multivariate data analysis methods*. Analytical Biochemistry, 1996. **241**(1): p. 5-13.
41. Jiang, J., et al., *Simulation of Two-Dimensional Ultraviolet Spectroscopy of Amyloid Fibrils*. Journal of Physical Chemistry B, 2010. **114**(37): p. 12150-12156.
42. Vanstokkum, I.H.M., et al., *Estimation of Protein Secondary Structure and Error Analysis from Circular-Dichroism Spectra*. Analytical Biochemistry, 1990. **191**(1): p. 110-118.
43. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition*. Annual Review of Biophysics and Biomolecular Structure, 2006. **35**: p. 389-415.
44. Chung, H.S., et al., *Transient 2D IR spectroscopy of ubiquitin unfolding dynamics*. Biophysical Journal, 2007: p. 207a-207a.
45. Chung, H.S., M. Khalil, and A. Tokmakoff, *Using nonlinear IR spectroscopy to probe early events in the thermal unfolding of proteins*. Abstracts of Papers of the American Chemical Society, 2004. **227**: p. U332-U332.
46. Chung, H.S. and A. Tokmakoff, *Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy*. Proteins-Structure Function and Bioinformatics, 2008. **72**(1): p. 474-487.