THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

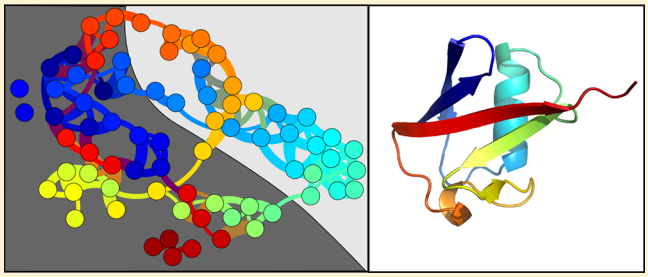
pubs.acs.org/JPCA

# Amide I Two-Dimensional Infrared Spectroscopy: Methods for Visualizing the Vibrational Structure of Large Proteins

Carlos R. Baiz, Mike Reppert, and Andrei Tokmakoff*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

**ABSTRACT:** Network layouts are introduced as a method to visualize couplings between local amide I vibrations in proteins. The method is used to identify groups of strongly coupled oscillators to block-diagonalize the Hamiltonians, considerably reducing the expense associated with computing infrared spectra of large proteins. The quality of linear and nonlinear spectra generated from block-diagonal Hamiltonians is demonstrated by comparison with spectra generated from full Hamiltonian trajectories. A library of six proteins reveals that vibrational couplings within hydrogen-bonded residues in specific secondary structures give rise to the characteristic amide I line shapes whereas other couplings play a minor role. Exciton delocalization analyses indicate that amide I vibrations in proteins remain largely localized to groups of less than ten residues.



## INTRODUCTION

Amide I infrared absorption spectroscopy is a simple and effective method to characterize protein secondary structure in solution.[1−4] Frequencies, line shapes and intensities, contain a wealth of information related to protein conformation. However, broad peaks and spectral overlap give rise to featureless absorption bands, and obscure much of the structural information. Two-dimensional infrared (2D IR) spectroscopy offers improved structural sensitivity by spreading the spectral information onto two frequency axes to measure vibrational couplings between different modes as well as to separate the static and dynamic spectral disorder of the ensemble.[5] Empirical relationships between structure and spectra have been developed in multiple attempts to quantitate secondary structure based on absorption IR spectra, but these attempts met with varying success as the broad bands complicate the structural assignments.[6−8] It was recently demonstrated how the additional spectral information derived from 2D IR spectroscopy can be leveraged to develop more accurate assays of protein structure.[9] However, thus far, structural analysis has remained relatively crude, largely limited to measuring the percentage of residues in $\alpha$-helix, $\beta$-sheet, $\beta$-turn, or unstructured coil conformations. To obtain information beyond bulk secondary structure content, electrostatic frequency maps have gained popularity as a means of connecting infrared spectra with atomistic protein structures, normally derived from molecular dynamics simulations.[10−18] The approach involves constructing an excitonic Hamiltonian based on site energies and couplings computed from electrostatic frequency maps combined with next-neighbor $\varphi/\psi$ coupling maps, and dipole−dipole interaction models to capture long-range interactions. Once a Hamiltonian trajectory is generated, nonlinear spectra can be computed by a static ensemble-averaged Liouville pathway sum of the transitions between ground, one, and two-quanta energy levels,[19,20] or by numerical integration of the Schrödinger equation (NISE).[21,22] Though generating amide I Hamiltonians from molecular dynamics simulations is relatively straightforward, the computational costs associated with converting the Hamiltonian into nonlinear spectra grow rapidly with the size of the system. The number of two-quanta states scales as $N(N+1)/2$, the matrix diagonalization step scales as $O(N^6)$[23] and the Liouville pathway enumeration scales as $O(N^4)$, where $N$ is the number of oscillators. Recent split-operator implementations of the NISE method have reduced the computational scaling to $O(N^4)$[24] and $O(N^3)$[25] by applying the Trotter approximation. Because of poor computational scaling, 2D IR simulations are presently limited to approximately 120−140 oscillators, and although sufficient for small peptides and protein domains, the computational expense prohibits simulation of larger systems such as protein assemblies, oligomers, or disordered aggregates.

Network graphs reveal the hidden the structure of complex systems by providing an intuitive visual map of the relationships between components.[26−28] Network methods have been applied to model systems in multiple areas of chemistry, biology and social sciences.[29] For example, protein folding is a particularly difficult multidimensional problem that has benefitted from recent Markovian network models to map out the folding landscape and identify metastable states along the protein folding coordinate.[30] Other notable applications include the use of networks to describe the fast folding of peptides under structural constraints, and the fluctuating hydrogen bonding structure of liquid water.[31,32] Some properties are common to many different networks, such as

**Special Issue:** Prof. John C. Wright Festschrift

**Received:** October 29, 2012
**Revised:** November 29, 2012
**Published:** December 11, 2012





dx.doi.org/10.1021/jp310689a | J. Phys. Chem. A 2013, 117, 5955−5961

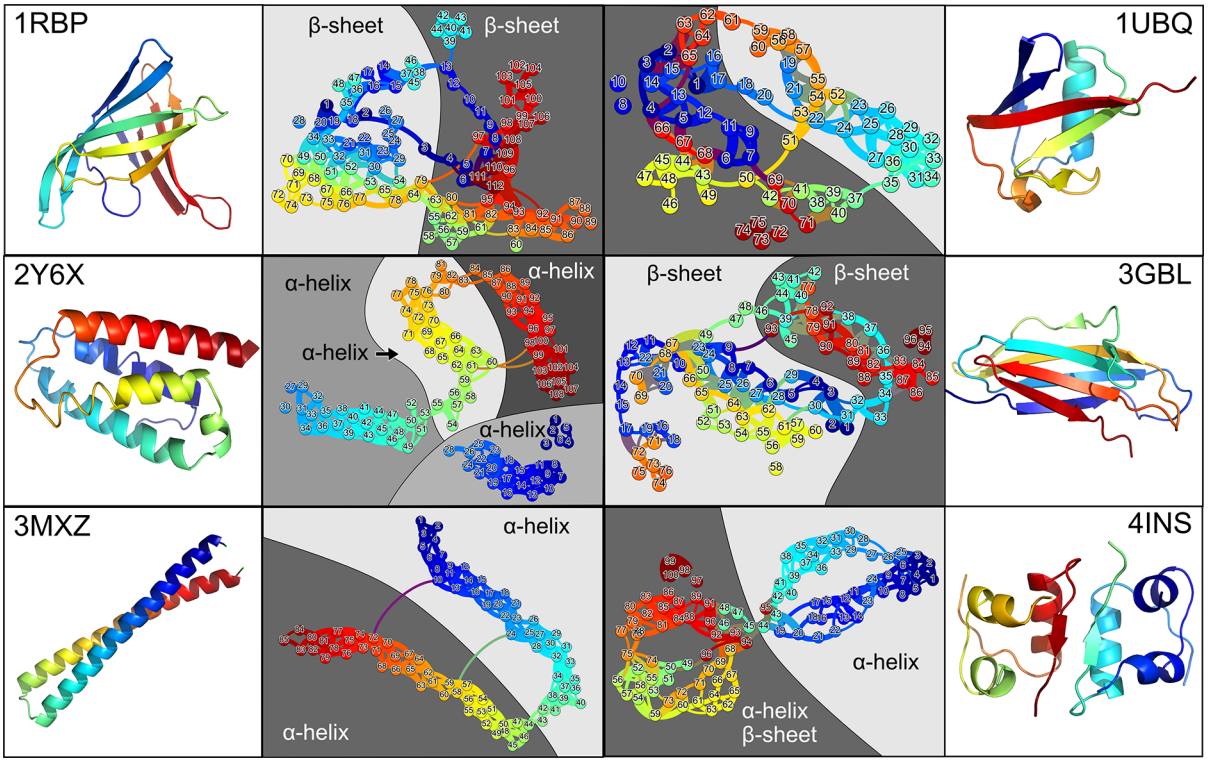



**Figure 1.** Protein structures and network models. Structures and network nodes are colored according to the same coloring scheme. Partitioning of the network is indicated by the different background shades and the respective secondary structures are shown in the figure. Coupling strengths are indicated by the thickness of the edges as well as the distances between nodes. Couplings of less than 2 $cm^{-1}$ are omitted for clarity. Molecules were rendered in PyMOL.[45]

for example, the so-called "small world effect", which implies that the average distance between two nodes is small compared to the size of the network.[33] Another important property, directly applicable to the analyses described here, is *community*, namely that nodes form small tightly clustered communities, which are only loosely connected to the rest of the network.[34]

The principal goal of this paper is to develop a general method, based on analyses of network graphs, to identify groups of weakly coupled oscillators. The Hamiltonians are then block-diagonalized into two or more blocks to reduce computational resources required to simulate spectra. The results also serve to explore the sensitivity of linear IR absorption, nonlinear transient IR absorption and 2D IR spectroscopy to the overall architecture of proteins, namely the spatial arrangement of structural elements that define the three-dimensional structure of a small protein or a protein domain. Although demonstrated in the context of amide I vibrations, the method is general and is applicable to any system of coupled modes.

## METHODS

**Molecular Dynamics Simulations.** Molecular structures from the protein databank (PDB)[35] were solvated in a box of SPC/E waters, energy minimized and equilibrated at 300 K and 1 atm using the all-atom OPLS force field. Protein atoms were position-restrained using a harmonic potential. Production simulations were carried out for 500 ps, with 2 fs step and snapshots saved every 1 ps. The GROMACS 3.3.1 package was used for all simulations.[36]

**Excitonic Model and Spectral Simulations.** Amide I modes were modeled as a set of bilinearly coupled oscillators using a Frenkel exciton Hamiltonian of the type[37]

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

where $\varepsilon_n$ represent the site frequencies, $J_{m,n}$ represent the coupling constants, and $\Delta$ the two-quanta anharmonicity (set to 16 $cm^{-1}$). Site energies and local transition dipole moments were computed from the projection of the electric field and gradients at the C, N, O, H positions of the amide moiety using the electrostatic map of Jansen and Knoester, including the recently developed map for proline residues.[14,17] Nearest neighbor couplings ($J_{i,i+1}$) were computed using a $\varphi/\psi$ map derived from ab initio calculations, and through-space coupling constants were calculated using the transition charge coupling (TCC) model.[15] Two-quantum transition dipoles were generated by harmonic scaling of the one-quantum dipoles.

Absorption spectra were computed in the static ensemble as

$$I(\omega) = \left\langle \sum_{k=1}^{N} |\vec{\mu}_k|^2 (\omega - \omega_k) \right\rangle \tag{2}$$


dx.doi.org/10.1021/jp310689a | J. Phys. Chem. A 2013, 117, 5955−5961

where $\omega_k$ and $\mu_k$ represent the normal-mode frequencies and transition dipoles of the one-quantum Hamiltonian respectively. Spectra were convolved with a 5 $cm^{-1}$ full-width-at-half-maximum (fwhm) Lorentzian function.

Rephasing (R) and nonrephasing (NR) 2D IR spectra were computed as

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

$$S^{NR}_{ijkl}(\omega_1,\omega_3) \propto \sum_{a,b=1}^{N} \delta(\omega_1-\omega_{a0})\{(\langle \mu_i^{0a}\mu_j^{0b}\mu_k^{b0}\mu_l^{a0}\rangle + \langle \mu_i^{0a}\mu_j^{a0}\mu_k^{0b}\mu_l^{b0}\rangle)\delta(\omega_3-\omega_{a0}) - \sum_{c=1}^{N(N+1)/2} \langle \mu_i^{0a}\mu_j^{0b}\mu_k^{ac}\mu_l^{cb}\rangle\delta(\omega_3-\omega_{cb})\} \quad (3)$$

where $\omega_1$ and $\omega_3$ represent the excitation and detection frequencies, respectively, and $\mu_i^A\mu_j^B\mu_k^C\mu_l^D$ are the lab-frame transition dipole moment components. Summation indices $a$ and $b$ run over one-quantum states, whereas index $c$ represents two-quantum states. Orientational contributions to the response functions were computed for an isotropic ensemble.[38] The perpendicular polarization geometry ($ZZYY$) was used to enhance cross peaks. Rephasing and nonrephasing contributions were convolved with 5 $cm^{-1}$ fwhm two-dimensional Lorentzian functions. Dispersed pump–probe (DPP) spectra were computed by projecting absorptive 2D IR spectra onto the detection axis ($\omega_3$).

**Network Model.** Network graphs, shown in Figure 1, provide an intuitive means of visualizing Hamiltonian matrices.[29] Oscillators are represented as nodes, and the coupling constants represented by the thickness—or weight—of the connecting edges. To generate the network, we computed an ensemble-averaged coupling matrix using the root-mean-squared (RMS) value of each coupling constant ($J_{i,j}$) from the ensemble of Hamiltonians. Coupling constants fluctuate by approximately ~20% relative to their RMS strength over the entire 10 ns trajectory. Thus, in principle, a very short trajectory, or even a single snapshot, would be sufficient to generate a relatively accurate network graph. It is worth pointing out that there are alternative approaches to calculating the coupling strengths. For example, a method that also takes into account differences in site energies, consists of computing the pairwise inverse participation ratio (IPR) as a measure of coupling.[39] Within the models described here, we found that the pairwise IPR is well correlated with the strength of the coupling constants.

Graph layouts were optimized using the ForceAtlas algorithm as implemented in Gephi 0.8.1.[40] Similar to molecular dynamics, the ForceAtlas method represents the graph as a system of masses (nodes) connected by springs (edges). The force constants are proportional to the weight of the edges, which are in turn proportional to the RMS coupling constants. An arbitrary repulsion term is generally added to prevent node overlap. The graph layout is then optimized by minimizing the "energy" of the system given the balance between attraction and repulsion forces. Layouts are not unique, but the nodes are similarly grouped starting from random initial configurations.

Once the layout is generated, the nodes are partitioned into two or more groups such that the number and weight of edges between different groups is minimized. In this work, we partitioned the Hamiltonian by visually inspecting the graph layout and grouping nodes into two or more subsets. The sub-Hamiltonians were then treated independently for the spectral simulations, and individual spectra were simply summed together to generate the final spectrum. For ease of visualization, network plots exclude couplings smaller than 2 $cm^{-1}$, though all coupling constants within each block were included in the spectral simulations.

**Protein Set.** The set of proteins was selected from the PDB such that the following criteria were satisfied: (1) The two main secondary structures ($\alpha$-helices and $\beta$-sheets) are represented in the set. (2) Proteins are smaller than 120 residues, yet contain at least two structures: helix–helix, helix–sheet, and sheet–sheet, with the exception of the $\beta$-barrel protein 1RBP. (3) Structures are derived from high-resolution (<2 Å) X-ray diffraction data. (4) Structures do not contain any metal centers or cofactors. Table 1 shows the set of proteins selected for analysis, and corresponding structures are shown in Figure 1. In addition, the protein set was selected such that different secondary structure motifs were represented: $\beta$-barrel (1RBP), tightly packed $\alpha$-helices (2Y6X) two loosely packed $\alpha$-helices (3MXZ), $\alpha$-helix and mixed parallel/antiparallel $\beta$-sheets (1UBQ), two antiparallel $\beta$-sheets (3GBL), $\alpha$-helix and antiparallel $\beta$-sheet (4INS).

**Table 1. Set of Proteins Used for Network Analysis and Spectral Simulations**[a]

| protein | PDB | $N$ | structure |
|---|---|---|---|
| retinol binding protein | 1RBP (24–136) | 113 | $\beta$-barrel |
| ubiquitin | 1UBQ | 76 | mixed $\alpha/\beta$ |
| Psb27 assembly factor | 2Y6X | 109 | $\alpha$-helix |
| $\beta$2-microglobulin | 3GBL | 97 | $\beta$-sheet |
| tubulin folding cofactor A | 3MXZ (1–87) | 87 | $\alpha$-helix |
| insulin dimer | 4INS | 102 | mixed $\alpha/\beta$ |

[a]The table includes the protein databank (PDB) code and number of residues ($N$).

**Vibrational Delocalization.** To gain an understanding of how the network models affect the character of the excitonic states in the proteins, we examine the degree of exciton delocalization. The inverse participation ratio, or delocalization length, expressed as[39]

$$P_\alpha = [\sum_{i=1}^{N} p_{i,\alpha}{}^2]^{-1} \quad (4)$$

describes the number of oscillators that contribute to the eigenstate $\alpha$, where $p_{i,\alpha}$ represents the contribution (population) of $i$th oscillator to the eigenstate. The value of $P$ ranges from one for a localized state to $N$ for a fully delocalized state. We also define a *network* inverse participation ratio as

$$\tilde{P}_\alpha = [\sum_{S} (\sum_{i\in\{S\}}^{N_s} p_{i,\alpha})^2]^{-1} \quad (5)$$


dx.doi.org/10.1021/jp310689a | J. Phys. Chem. A 2013, 117, 5955–5961

where the inner summation involves oscillators belonging to a specific section ($S$) of the network and thus can be interpreted as an effective *population* of that section, and the outer summation runs over the sections of the network. Similarly, the value of $\tilde{P}_\alpha$ ranges from one to the number of sections in the network.

## DISCUSSION

Network graphs for the six proteins analyzed here are shown in Figure 1. It is worth pointing out the following observations: First, the layouts of the network graphs mirror the three-dimensional structures of the proteins. For example, $\alpha$-helices show strong coupling between $N$ and $N + 3$ pairs, and coupling patterns in $\beta$-sheets generally reflect the hydrogen bonding contacts between strands. Second, the network graphs also show that nearest-neighbor ($N$, $N + 1$) couplings are relatively weak. Third, through-space couplings between different secondary structures are weak, as evidenced by the lack of edges connecting nodes within different secondary structures. Finally, the network layouts provide an intuitive, albeit not unique, way of block-diagonalizing the Hamiltonian such that the smallest number of edges is disrupted. Not surprisingly, the Hamiltonian breakups tend to align with the secondary structure of the protein, suggesting that in most cases simple inspection of the protein structure may be sufficient to determine a suitable partitioning scheme. However, the networks can be helpful in cases where it is not clear how to break up the Hamiltonian a priori, such as in the case of the $\beta$-barrel protein 1RBP.

Simulated amide I absorption spectra are shown Figure 2. Primarily $\alpha$-helical proteins (2Y6X, 3MXZ) exhibit two peaks centered around 1640−1660 $cm^{-1}$,[41,42] whereas $\beta$-sheets exhibit two bands centered near 1640 and 1670 $cm^{-1}$, which correspond to vibrational states with transition dipole moments perpendicular and parallel to the $\beta$-strands respectively.[20,43]

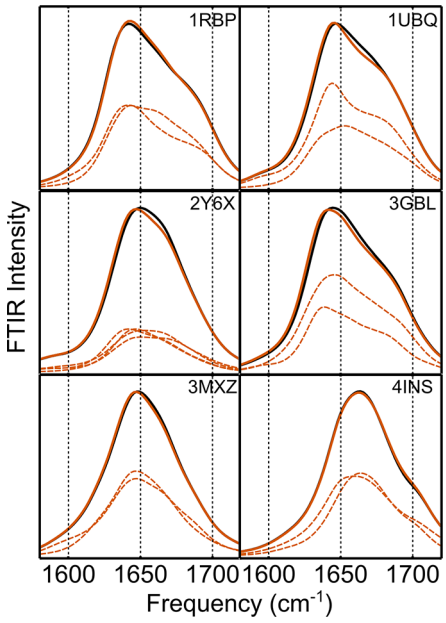



**Figure 2.** Simulated IR absorption spectra using the full Hamiltonian (black curves) and the block-diagonalized Hamiltonians (orange curves). Spectral contributions from individual sections are shown below each spectrum (dashed).

Absorption, dispersed pump-probe, and 2D IR spectra (Figures 2, 3, and 4, respectively) simulated using the network analysis approach are virtually identical to spectra simulated with the full Hamiltonians, suggesting that there is little structural information contained in the weak coupling constants between different secondary structures. In other words, the peak positions, line widths, and intensities of the peaks are determined predominantly by the strong coupling constants between near neighbors and hydrogen-bonded residue pairs. The results can be partially explained by the fact that site energy fluctuations are large relative to the coupling constants neglected in the block-diagonalization. The site disorder tends to localize the excitons to groups of ~6 oscillators (see below), suggesting that it is improbable for an exciton to delocalize over multiple sections of the network. These results indicate that amide I spectroscopy is in general not sensitive to the spatial arrangement of secondary structures.

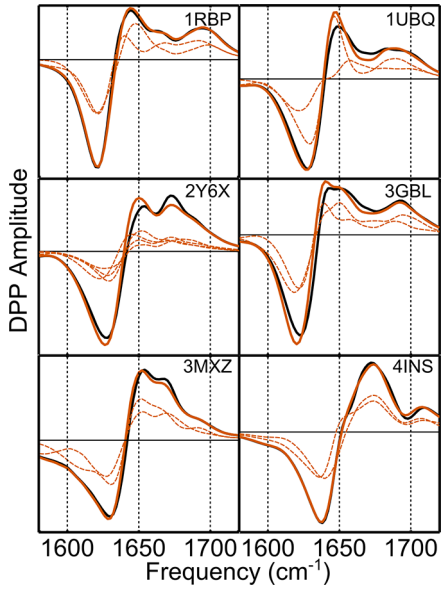



**Figure 3.** Dispersed pump−probe spectra for full Hamiltonian curves (black) and block-diagonalized (orange) along with the contributions from each section of the network (dashed).

The best agreement between the full-Hamiltonian and network spectra is observed for $\alpha$-helical proteins, such as 3MXZ. These results likely reflect the fact that steric interactions between side chains prevent helices from coupling strongly. Despite being a purely $\alpha$-helical protein, the differences for 2Y6X are more pronounced than for 3MXZ; this is in part a result of block-diagonalization into four subspaces. Recently, Barth and co-workers[44] pointed out that helix−helix coupling leads to additional spectral shifts in helical multimers, usually found in membrane proteins. Therefore, the methods presented here require further validation in the case of membrane proteins.

## VIBRATIONAL FREQUENCY DISORDER AND DELOCALIZATION

The inverse participation ratio (eq 4) provides a measure of the delocalization *length* of the excitons. Large site frequency disorder, relative to the coupling constants, tends to localize the excitons. Figure 5 shows the average inverse participation ratio as a function of frequency. States in the low or high frequency


dx.doi.org/10.1021/jp310689a | J. Phys. Chem. A 2013, 117, 5955−5961

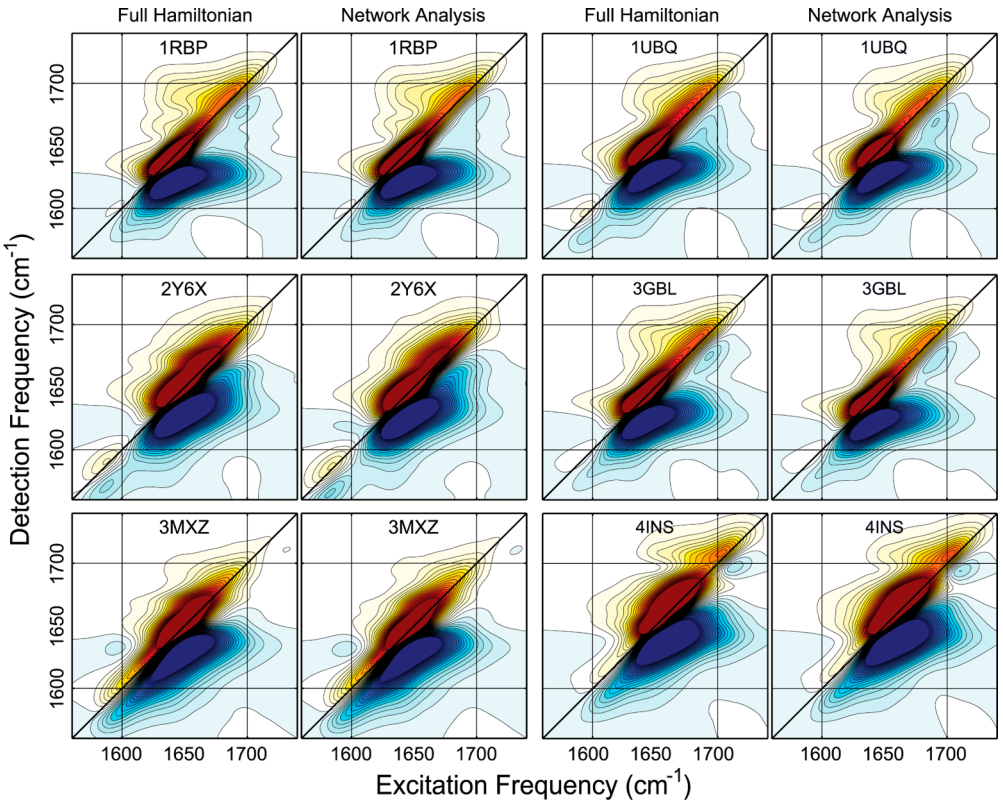



**Figure 4.** Two-dimensional infrared (2D IR) spectra of ubiquitin (PDB: 1UBQ) calculated using the full Hamiltonian trajectories and the block-diagonalized Hamiltonians extracted from the network analysis.

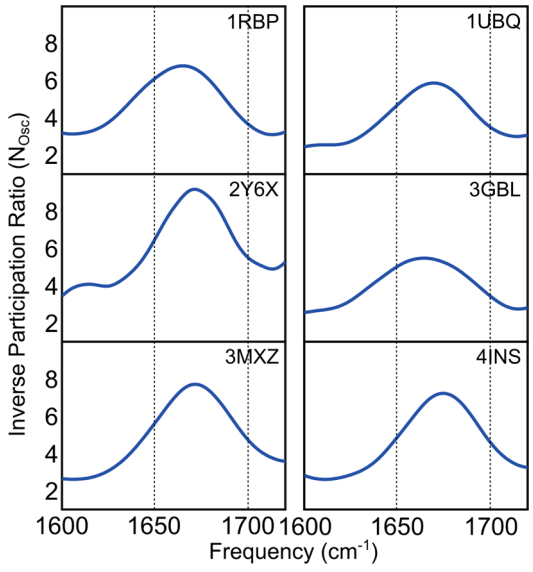



**Figure 5.** Inverse participation ratio (average delocalization length, eq 4) as a function of frequency for all six proteins in the set.

regions (<1620, >1700 $cm^{-1}$) are localized to approximately 2−4 oscillators whereas states in the center of the band are more delocalized, spanning a space composed of 6−8 oscillators. This observed delocalization length is approximately the same number of residues as a small secondary structure motif, suggesting that excitations are localized to individual hydrogen-bonded structures (i.e., $\alpha$-helix or $\beta$-sheet).

Dynamic site fluctuations, not captured by our model, in the high-disorder limit can lead to narrower effective site energy distributions, inducing additional delocalization, whereas in low-disorder systems, dynamics can localize states. Because the static analysis reveals that states are relatively localized, here we only explore the effects of dynamic delocalization. The site frequency-frequency time correlation function decays to about half of its initial value within the first 150 fs.[19] To simulate the effect of dynamic delocalization within our static model, we applied a five-frame moving-window average to the site energy trajectories (1 ps/frame sampling rate). The window function reduces the site energy distribution by approximately a factor of 2, whereas the inverse participation ratio only increases by approximately 1−1.5 oscillators. These results suggest that fast fluctuations are not an important contribution to exciton delocalization in these systems.

The present models indicate that large site disorder leads to localized states, providing additional insight into why block-diagonalizing the amide I Hamiltonian has little effect on the computed spectra. The plots also show that, even for the most delocalized states, the delocalization length is significantly smaller than the number of oscillators in each network subspace, indicating that excitons are largely localized to single secondary structures. Consequently, the length and frequencies of the excitons remain relatively unperturbed upon block-diagonalization of the Hamiltonian. Root-mean-squared network inverse participation ratios (Table 2) range from one to the number of sections in the network: two sections for all proteins in the set except for 2Y6X where the Hamiltonian is divided into four sections. The values show that, on average, the excitons are mostly localized to one of the subspaces only ∼25% of the exciton is delocalized over two sections. This number is larger for 2Y6X where the size of the network sections is smaller (∼27 oscillators) compared to the other


dx.doi.org/10.1021/jp310689a | J. Phys. Chem. A 2013, 117, 5955−5961

Table 2. Root-Mean-Squared (RMS) Inverse Participation Ratio ($P_\alpha$) and Network Inverse Participation Ratio ($\tilde{P}_\alpha$)

| | $P_\alpha$ (RMS) | $\tilde{P}_\alpha$ (RMS) |
|---|---|---|
| 1RBP | 6.56 | 1.27 |
| 1UBQ | 5.47 | 1.26 |
| 2Y6X | 8.53 | 1.65 |
| 3GBL | 5.37 | 1.27 |
| 3MXZ | 6.83 | 1.29 |
| 4INS | 6.51 | 1.24 |

proteins in the set. The results suggest that smaller values of the network inverse participation ratio will produce spectra which are in better agreement with the full Hamiltonian spectra. Because the network inverse participation ratio can be generated at little computational expense, the analysis may be used in conjunction with optimization algorithms to determine the optimal partitioning of the Hamiltonian.

## CONCLUSIONS AND SUMMARY

We have presented a new approach to block-diagonalizing Hamiltonian trajectories through a network analysis to identify groups of strongly coupled residues. The excellent agreement observed between the full Hamiltonian and network analysis spectra suggest that amide I 2D IR spectra are relatively unaffected by couplings between secondary structures of intra- or intermolecular nature such as in the insulin dimer (4INS). The results indicate that, to within a good approximation, spectra of globular proteins can be represented as linear combinations of secondary-structure component spectra. Finally, the network analysis approach presented here provides an intuitive and practical way to reduce the computational expense of amide I spectral simulations in proteins. Although the selection of the Hamiltonian blocks was done manually, the procedure can be automated using a combination of network community detection algorithms[34] combined with nonlinear optimization algorithms.

In closing, we emphasize that the analyses presented here remain solely a result of theoretical modeling. Importantly, electrostatic maps, which have only been tested and benchmarked on systems consisting of a few amino acids, may not accurately describe amide I vibrations in proteins. Though the models qualitatively reproduce experimental spectra, the model predictions on delocalization length and couplings between secondary structures are difficult to test experimentally. We believe that multidimensional spectroscopy, combined with isotope labeling will provide additional spectral information on amide I vibrations that will enable further refinement of the vibrational models.

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: tokmakof@mit.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Thomas la Cour Jansen for the insightful comments on the manuscript. This work was supported by the National Science Foundation (CHE-0911107 and CHE-1212557) and a grant from Agilent Technologies. M.R. thanks the National Science Foundation for a Graduate Research Fellowship.

## REFERENCES

(1) Krimm, S.; Bandekar, J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. *Adv. Protein Chem.* **1986**, *38*, 181−364.
(2) Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta-Bioenerg.* **2007**, *1767* (9), 1073−1101.
(3) Barth, A.; Zscherp, C. What vibrations tell us about proteins. *Q. Rev. Biophys.* **2002**, *35* (4), 369−430.
(4) Jackson, M.; Mantsch, H. H. The Use and Misuse of Ftir Spectroscopy in the Determination of Protein-Structure. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30* (2), 95−120.
(5) Hamm, P.; Zanni, M. *Concepts and Methods of 2d Infrared Spectroscopy*; Cambridge University Press: Cambridge, U.K., 2011.
(6) Goormaghtigh, E.; Ruysschaert, J. M.; Raussens, V. Evaluation of the information content in infrared spectra for protein secondary structure determination. *Biophys. J.* **2006**, *90* (8), 2946−2957.
(7) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by Deconvolved Ftir Spectra. *Biopolymers* **1986**, *25* (3), 469−487.
(8) Ruegg, M.; Metzger, V.; Susi, H. Computer Analyses of Characteristic Infrared Bands of Globular Proteins. *Biopolymers* **1975**, *14* (7), 1465−1471.
(9) Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* **2012**, *137* (8), 1793−1799.
(10) Bour, P.; Keiderling, T. A. Empirical modeling of the peptide amide I band IR intensity in water solution. *J. Chem. Phys.* **2003**, *119* (21), 11253−11262.
(11) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.* **2004**, *121* (18), 8887−8896.
(12) Ham, S.; Kim, J. H.; Lee, H.; Cho, M. H. Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA-nD(2)O complexes. *J. Chem. Phys.* **2003**, *118* (8), 3491−3498.
(13) Hamm, P.; Woutersen, S. Coupling of the Amide I Modes of the Glycine Dipeptide. *Bull. Chem. Soc. Jpn.* **2002**, *75* (5), 985−988.
(14) Jansen, T. L.; 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.* **2006**, *124* (4).
(15) Jansen, T. L.; Knoester, J. Nonadiabatic effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide. *J Phys Chem B* **2006**, *110* (45), 22910−22916.
(16) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J. Melting of a beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations. *J Phys Chem B* **2010**, *114* (34), 10913−10924.
(17) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. Solvent and conformation dependence of amide I vibrations in peptides and proteins containing proline. *J. Chem. Phys.* **2011**, *135* (23), 234507.
(18) Lessing, J.; Roy, S.; Reppert, M.; Baer, M.; Marx, D.; Jansen, T. L.; Knoester, J.; Tokmakoff, A. Identifying Residual Structure in Intrinsically Disordered Systems: A 2D IR Spectroscopic Study of the GVGXPGVG Peptide. *J. Am. Chem. Soc.* **2012**, *134* (11), 5032−5035.
(19) Ganim, Z.; Tokmakoff, A. Spectral signatures of heterogeneous protein ensembles revealed by MD simulations of 2DIR spectra. *Biophys. J.* **2006**, *91* (7), 2636−2646.
(20) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. Signatures of beta-sheet secondary structures in linear and two-dimensional infrared spectroscopy. *J. Chem. Phys.* **2004**, *120* (17), 8201−8215.
(21) Jansen, T. L. C.; Knoester, J. Waiting Time Dynamics in Two-Dimensional Infrared Spectroscopy. *Acc. Chem. Res.* **2009**, *42* (9), 1405−1411.
(22) 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* **2006**, *110* (14), 4822−4832.
(23) Anderson, E. *LAPACK users' guide*, 3rd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1999; p xxi, 407 pp.
(24) Paarmann, A.; Hayashi, T.; Mukamel, S.; Miller, R. Nonlinear response of vibrational excitons: simulating the two-dimensional infrared spectrum of liquid water. *J. Chem. Phys.* **2009**, *130* (20), 204110.
(25) Liang, C.; Jansen, T. An Efficient N3-Scaling Propagation Scheme for Simulating Two-Dimensional Infrared and Visible Spectra. *J. Chem. Theory Comput.* **2012**, *8* (5), 1706.
(26) Strogatz, S. H. Exploring complex networks. *Nature* **2001**, *410* (6825), 268−276.
(27) Barabasi, A. L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286* (5439), 509−512.
(28) Albert, R.; Barabasi, A. L. Statistical mechanics of complex networks. *Rev Mod Phys* **2002**, *74* (1), 47−97.
(29) Newman, M. E. J. *Networks: an introduction*; Oxford University Press: Oxford , New York, 2010; p xi, 772 pp.
(30) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52* (1), 99−105.
(31) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E. H. G.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. $\alpha$-Helix folding in the presence of structural constraints. *Proc. Natl. Acad. Sci.* **2008**, *105* (28), 9588−9593.
(32) Rao, F.; Garrett-Roe, S.; Hamm, P. Structural Inhomogeneity of Water by Complex Network Analysis. *J. Phys. Chem. B* **2010**, *114* (47), 15598−15604.
(33) Amaral, L. A. N.; Scala, A.; Barthelemy, M.; Stanley, H. E. Classes of small-world networks. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (21), 11149−11152.
(34) Girvan, M.; Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (12), 7821−7826.
(35) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301−D303.
(36) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(37) Hamm, P.; Lim, M. H.; Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **1998**, *102* (31), 6123−6138.
(38) Golonzka, O.; Tokmakoff, A. Polarization-selective third-order spectroscopy of coupled vibronic states. *J. Chem. Phys.* **2001**, *115* (1), 297−309.
(39) Bell, R. J.; Dean, P. Atomic vibrations in vitreous silica. *Discuss. Faraday Soc.* **1970**, *50*, 55−61.
(40) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *3rd International AAAI Conference on Weblogs and Social Media*; AAAI: Palo Alto, CA, 2009.
(41) Torii, H.; Tasumi, M. 3-Dimensional Doorway-State Theory for Analyses of Absorption-Bands of Many-Oscillator Systems. *J. Chem. Phys.* **1992**, *97* (1), 86−91.
(42) Nevskaya, N. A.; Chirgadze, Y. N. Infrared-Spectra and Resonance Interactions of Amide-I and Amide-II Vibrations of Alpha-Helix. *Biopolymers* **1976**, *15* (4), 637−648.
(43) Demirdoven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure. *J. Am. Chem. Soc.* **2004**, *126* (25), 7981−7990.
(44) Karjalainen, E.-L.; Barth, A. Vibrational coupling between helices influences the amide I infrared absorption of proteins: application to bacteriorhodopsin and rhodopsin. *J. Phys. Chem. B* **2012**, *116* (15), 4448−4456.
(45) Schrödinger, L. *The PyMOL Molecular Graphics System*, Version 1.4.1; 2012.