

Dissecting coherent vibrational spectra of small proteins into secondary structural elements by sensitivity analysis

Wei Zhuang, Darius Abramavicius, and Shaul Mukamel*

Department of Chemistry, University of California, Irvine, CA 92697-2025

Edited by Robin M. Hochstrasser, University of Pennsylvania, Philadelphia, PA, and approved March 25, 2005 (received for review November 24, 2004)

The response of proteins to sequences of femtosecond infrared pulses provides a multidimensional view into their equilibrium distribution of structures and snapshot pictures of fast-triggered dynamical events. Analyzing these experiments requires advanced computational tools for assigning regions in the resulting multidimensional correlation plots to specific secondary-structure elements and their couplings. A differential sensitivity analysis technique based on a perturbation of the local (real space) Hamiltonian is developed to achieve that goal. Application to the amide I region of a small globular protein reveals regions associated with the α -helix, β -sheet, and their coupling. Comparison of signals generated in different directions shows that the double-quantum-coherence signal has a higher sensitivity to the couplings compared with the single-quantum-coherence (photon echo) technique.

protein structure determination | two-dimensional infrared

Probing protein structure and folding pathways is one of the central problems of modern structural biology. Elaborate NMR pulse sequences are routinely used in the study of millisecond dynamical events in complex biomolecules with high spatial resolution (1). Optical response functions are particularly sensitive to structural features such as bond strengths, backbone angles, dipole moments, and polarizabilities, as well as the environmental factors that influence the structure such as hydrogen bond strengths and electric fields. Femtosecond optical techniques, thus, offer a good complimentary tool for monitoring fast protein motions and unraveling the underlying couplings between functional groups (2) in real time.

Two-dimensional techniques introduced into NMR in the 1970s revolutionized its resolution and utility (1). Similarly, significant progress has been made over the past decade in developing 2D infrared (2DIR) spectroscopy, based on the application of sequences of femtosecond pulses, into a powerful tool for probing protein structure and providing snapshots of fast-triggered protein folding events (3–6).

Spreading the information in two dimensions greatly enhances the resolution. Furthermore, 2DIR carries additional information regarding the fluctuations of two-exciton-state frequencies that is not available from linear (1D) spectra and provides a direct look at correlations among various segments of the protein. Cross peaks of 2DIR signals of equilibrated structures provide new structural constraints analogous to the nuclear Overhauser effect in 2D NMR. Structural information is extracted from multidimensional NMR data by using extensive simulations (7), through “direct inversion” based on constraint fits guided by simulations or structural optimizations. Their infrared analogues are less resolved, and anharmonic vibrational Hamiltonians are much more complex than spin Hamiltonians, making simulations an even more essential part of the analysis apparatus. Thanks to the ultrafast time scale, the simulation of 2DIR signals with atomic-level details only requires nanosecond trajectories, allowing the microscopic interpretation of spectra and providing a test for the simulated structural trajectories. Most 2DIR simulations carried out so far focused on small

peptides with <10 residues (8–10), which do not have a unique stable structure. 2DIR signatures of ideal infinite size secondary structures were investigated recently (12, 13). Globular proteins are made of several secondary-structure segments, and relating their coherent vibrational response to the ideal motifs constitutes an open computational challenge.

In this article, we demonstrate how sensitivity analysis (14), in conjunction with a local basis Hamiltonian, may be used for assigning various regions in the spectra of a globular protein to specific secondary structures, and show how overlapping regions may be better separated in higher dimensions. This technique may further be applied for selecting the proper pulse sequence for specific structure application. We consider two three-pulse coherent four-wave mixing techniques, whereby three incoming pulses with propagation wave vectors \mathbf{k}_1 , \mathbf{k}_2 , and \mathbf{k}_3 interact with the protein to generate a coherent signal in the directions $\mathbf{k}_I = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ and $\mathbf{k}_{III} = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$. The pulse sequence and time delays (t_1 , t_2 , and t_3) are shown in Fig. 1 *Top Right Upper*.

Both additive and a differential sensitivity analysis techniques were applied to reveal signatures of couplings among secondary motifs and assign different spectral regions to the structures. The former compares the actual spectra with additive spectra obtained by switching off the couplings among different secondary structure motifs. In the latter, we examine derivatives of the spectra with respect to various parameters of the local Hamiltonian. Our simulations show that 2DIR techniques are much more sensitive to the couplings and offer better resolution for the secondary structures and their interactions compared with the linear absorption. We further demonstrate that (t_2 , t_3) correlation plots of \mathbf{k}_{III} are more sensitive to the couplings compared with (t_1 , t_3) correlations in \mathbf{k}_I .

In *The Vibrational Exciton Hamiltonian and Simulation Protocol*, we introduce the vibrational Hamiltonian and the simulation protocol. The additive sensitivity analysis is presented in *Test of Secondary Structure Additivity of Coherent Spectra*, and the differential analysis is developed in *Differential Sensitivity Analysis*. We conclude with a discussion of the results and future perspectives.

The Vibrational Exciton Hamiltonian and Simulation Protocol

Vibrational motions and spectra are commonly described by normal modes. These collective coordinates, which provide a convenient zero-order approximation for the vibrational eigenstates and frequencies, are delocalized and do not directly connect with the structure; it is much easier to trace the structural origin of spectral features by using local vibrational coordinates (15).

Our simulations were carried out for the coherent third-order IR response of a 74-residue protein domain, TB6 (Fig. 1 *Top*

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: 2DIR, 2D infrared.

*To whom correspondence should be addressed. E-mail: smukamel@uci.edu.

© 2005 by The National Academy of Sciences of the USA

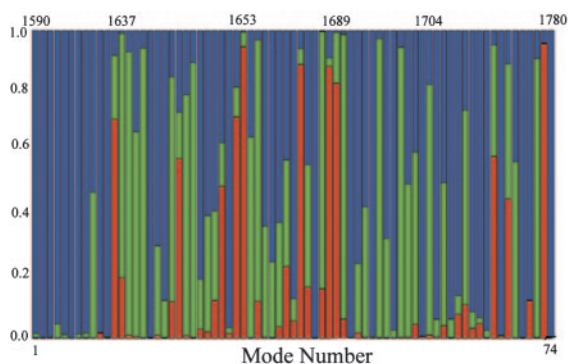
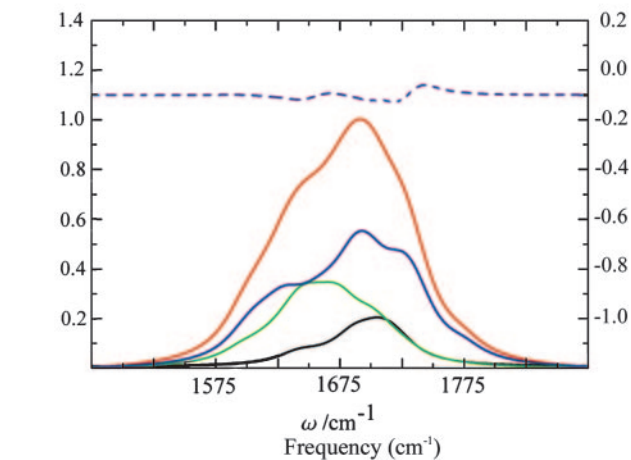
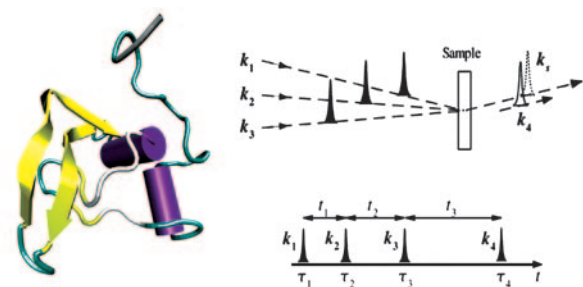


Fig. 1. The molecular structure of TB6 and pulse sequence configuration. (Top Left) The molecular structure of TB6. The Purple sections are the helices, the yellow sections are the β -strands, and the green and white sections are randomly coiling structures. (Top Right Upper) Pulse configuration for heterodyne four-wave mixing techniques. k_1 , k_2 , and k_3 are the input pulses, and k_4 is the signal generated, which is in the same direction as the detection pulse k_4 . (Top Right Lower) The pulse sequence for coherent 2D experiments. t_i are the time intervals between the pulses, and τ_i is the peak time of pulse i . (Middle) Red, linear absorption; black, AS(α); green, AS(β); blue, AS(γ); purple dash, the difference of the additive and the actual spectrum. (Bottom) Motif content of the vibrational eigenstates of TB6. Red, α ; green, β ; blue, random coil. The plot is vs. mode number; the (nonuniform) frequency scale is given at the top.

Left) (16). TB6 is one of the 11 genetic mutations in TB domains from human fibrillin 1, which cause the Marfan syndrome. There are 10 residues in its α -helix section (α), 22 β -sheets (β), and 42 randomly coiled residues (γ). This protein was chosen because it is large enough to have well defined secondary structures, yet sufficiently small to make the simulations feasible (13).

We start with the vibrational exciton Hamiltonian (11, 12).

$$\hat{H}_s = \sum_m \varepsilon_m \hat{B}_m^\dagger \hat{B}_m - \sum_m \frac{\Delta_m}{2} \hat{B}_m^\dagger \hat{B}_m^\dagger \hat{B}_m \hat{B}_m + \sum_{m \neq n} J_{mn} \hat{B}_m^\dagger \hat{B}_n$$

Table 1. The averages and variances of local mode frequencies

Frequency	$\bar{\omega}$, cm^{-1}	δ , cm^{-1}
$\varepsilon_m^\alpha(t)$	1,684.7	22.71
$\varepsilon_m^\beta(t)$	1,667.8	29.83
$\varepsilon_m^\gamma(t)$	1,681.2	42.09

$\varepsilon_m^\nu(t)$ is the frequency of the m th mode of ν type at time t .

Here, \hat{B}_m^\dagger (\hat{B}_m) is the creation (annihilation) operator for the m th amide I mode, localized within the amide units ($\text{O}=\text{C}-\text{N}-\text{H}$ segments), with frequency ε_m and diagonal anharmonicity Δ_m . These operators satisfy the Bose commutation relations $[\hat{B}_m, \hat{B}_n^\dagger] = \delta_{mn}$. J_{mn} is the harmonic intermode coupling.

All parameters of the Hamiltonian fluctuate because of conformational changes of the backbone, and solvent and side-chain dynamics. The vibrational system is thus perturbed by a time-dependent electric field causing line broadening and frequency shifts. The present study focuses on the primary peak pattern. We thus calculated the averaged values of all frequency shifts and did not model the line broadening microscopically. A Lorentzian lineshape with a 20-cm^{-1} homogeneous linewidth was assumed for all transitions.

The fundamental frequency of each of the 74 local modes is given by $\varepsilon_m(t) = \varepsilon - \delta\varepsilon_m(t)$, where $\varepsilon = 1,707\text{ cm}^{-1}$ is the frequency of isolated NMA molecule in the gas phase (17, 18). The water molecules as well as all protein atoms other than the amide units are treated as a bath, which generates the fluctuating electrostatic field responsible for the frequency shift $\delta\varepsilon_m(t)$. The instantaneous frequency shift was calculated for each mode by using the linear correlation formula with the instantaneous electric potential (19). The point charges of the peptide unit atoms ($\text{O}=\text{C}-\text{N}-\text{H}$) were obtained from ref. 19, and those of the other protein atoms were taken from the CHARMM27 force field (20). The TIP3 model was used for the water charges and the simulation. An ensemble of $\delta\varepsilon_m(t)$ was generated by using the simulated structural trajectory obtained from a molecular dynamics simulation starting with the NMR resolved structure from the Protein Data Bank (ID code 1APJ), which includes all of the atoms of one TB6 and 13,608 water molecules. The simulation was carried out by using the CHARMM package with the CHARMM27 force field (20). A 10-\AA cutoff was used for nonbonded interactions. The long-range electrostatic interactions were calculated using the Ewald Sum method (21). The TB6 domain was imbedded in a 76-\AA cube of TIP3 water. The cutoff distance for the Lennard-Jones forces was set to 12 \AA . All water molecules overlapping with the protein were removed. An adopted basis Newton-Raphson energy minimization (5×10^4 steps) (20) was performed to release the internal tension. In this and all steps that follow, the protein backbone atom coordinates were held fixed at the NMR resolved structure. After energy minimization, the system was heated to 300 K in 20 ps, and a 5-ns equilibration in NPT (constant number, pressure, and temperature) ensemble (22) with 1-fs time steps was performed to get the correct density of the system. The extended system method was used to hold the temperature and pressure constant (22), and the final box length was 74.67 \AA . This was followed by a 10-ns equilibration of the system in the NVE (constant number, volume, and energy) ensemble with a 2-fs time step [the SHAKE algorithm (23) was used in the NVE ensemble equilibration to constrain the bonds connected with hydrogen atoms]. A 1-ns NVE dynamics was run with 2-fs time steps. One hundred snapshots of the frequency shifts $\delta\varepsilon_m(t)$ were recorded, and the average value of $\varepsilon_m(t)$ for each amide I mode was then used as ε_m in the Hamiltonian Eq. 1.

The average values and variances of ε_m for the various modes of each secondary structure type are given in Table 1, and the

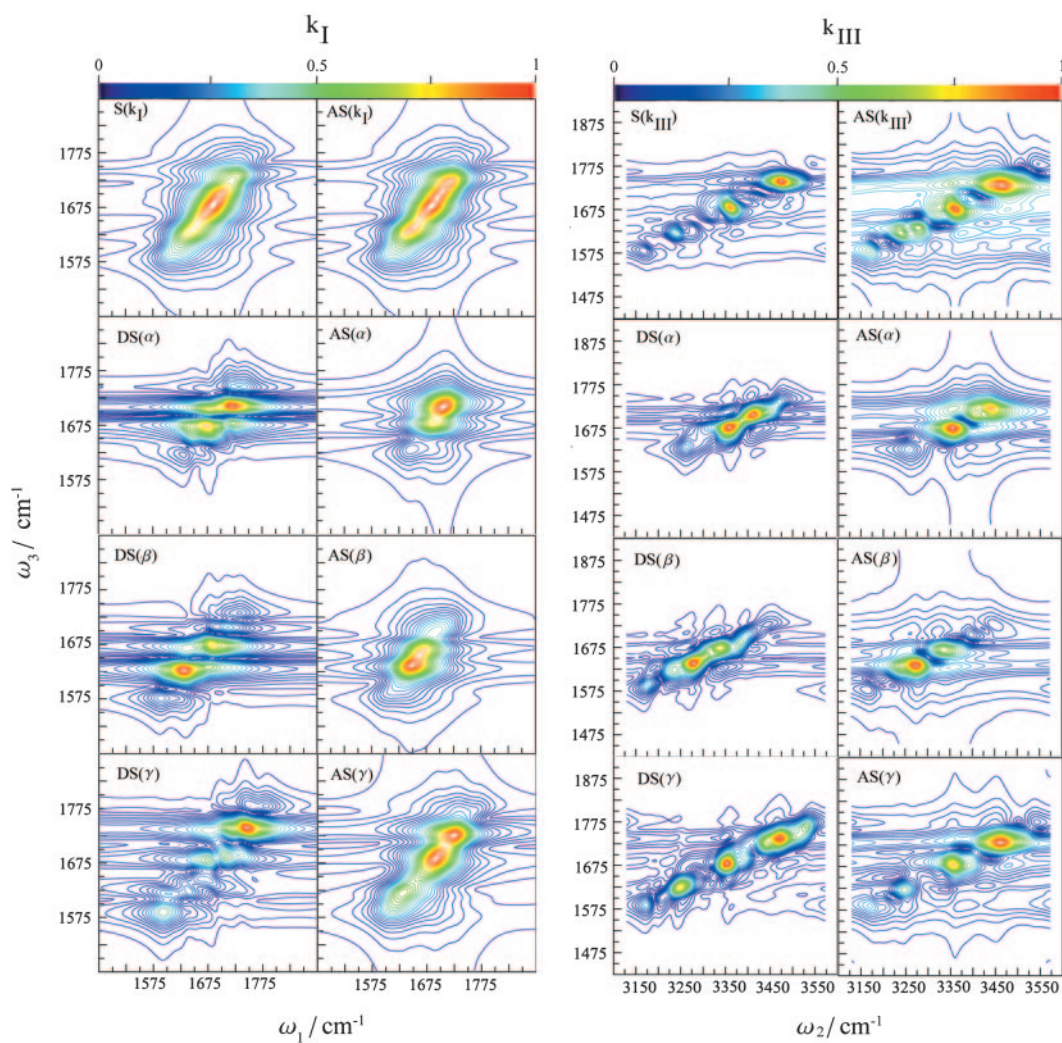


Fig. 2. Two-dimensional coherent vibrational spectra of TB6. (*Left*) Sensitivity analysis for k_I . In the left column is shown the total simulated spectrum $S(k_I)$ and its derivatives with respect to the various secondary structures, $DS(\alpha)$, $DS(\beta)$, and $DS(\gamma)$. In the right column is shown the additive spectrum obtained by neglecting inter secondary structure couplings $AS(k_I)$, and the contributions of each motifs to the spectrum [$AS(\alpha)$, $AS(\beta)$, and $AS(\gamma)$]. (*Right*) Sensitivity analysis for k_{III} . In the left column is shown the total simulated spectrum $S(k_{III})$ and its derivatives with respect to the various secondary structures, $DS(\alpha)$, $DS(\beta)$, and $DS(\gamma)$. In the right column is shown the additive spectrum obtained by neglecting inter secondary structure couplings, $AS(k_{III})$, and the contributions of each motifs to the spectrum [$AS(\alpha)$, $AS(\beta)$, and $AS(\gamma)$].

structure of type ν . Obviously, $DS(\nu)$ of each motif depends on the couplings with other motifs. This is an inevitable consequence of the nonadditivity of the spectra when various segments interact strongly. $DS(\nu)$ may, thus, not be viewed as the “pure” spectrum of motif ν but rather as the spectral feature which is strongly affected by that motif.

The lower six panels of k_I in Fig. 2 show that for k_I , the DS and AS peak positions are very close. This is because the couplings between different amide I modes are not very strong. However, the DS spectra have additional peaks. Also, because the AS represent the spectra of the isolated structures and the DS include their couplings with other parts of the protein, we observe peak shifts. Consequently, the two overlapping peaks in AS for α are well separated in DS. Similar trends are observed in the k_{III} spectra shown in Fig. 2. We have operationally defined the frequency regime bounded by the 35% contour of the highest peak of $DS(\alpha)$ in Fig. 2 as the α regime. This is shown by the black contour in the middle panels of Fig. 4. The β regime defined similarly is given by the red contour. These two regions are clearly better separated in k_{III} than in k_I .

This differential sensitivity analysis can also be used to identify the regions corresponding to inter-secondary-structure couplings: By perturbing the off-diagonal elements in the local Hamiltonian, it should be possible to probe directly the effect of couplings on the spectra. This should identify the spectral domains that are sensitive to real-space interactions, and to molecular structure. Fig. 4 *Bottom* shows that the DS of the k_{III} technique with respect to $J_{\alpha\beta}$ is clearly more localized than k_I . Because of the large variance of the couplings, we applied a different perturbation protocol, whereby all $J_{\alpha\beta}$ were multiplied by a factor 1.001. A comparison of Fig. 4 *Middle* with Fig. 4 *Bottom* shows that for both techniques, the strong peaks of DS with respect to $J_{\alpha\beta}$ appear around the region where the α and β features overlap. The peak intensity is determined by the ratio of the coupling and the energy difference of the two local modes. When the coupling is much weaker than the energy difference (the weak coupling limit), the changes of the eigenstate energies are also small and the difference spectrum is weak. The difference spectrum signal will increase as this ratio increases. In the strong coupling limit, the eigenstates will have similar contribu-

tions. The bottleneck will be the repeated Hamiltonian diagonalizations and carrying out the multiple summations over eigenstates.

Full-scale inhomogeneous averaging over 100 configurations in TB6 will require 1,000 machine hours on an AMD Opteron 244. The present sensitivity analysis will require repeating this calculation 40 times. The nonlinear exciton equations (11, 12) provide the most promising method for large-scale inhomogeneous averaging. They use a single-exciton basis and view the nonlinear response in terms of scattering between excitons (quasi-particles) rather than transitions among eigenstates. The expensive computation of multiple exciton states is totally avoided. The bottleneck in this calculation is the calculation of a tetradic frequency-dependent exciton scattering matrix. Our recent calculations show that useful approximations that exploit

the short-range nature of exciton couplings may reduce computational cost and will scale very favorably with protein size.

Sensitivity analysis can further be applied to connect chirally sensitive techniques such as vibrational IR circular dichroism and Raman optical activity (30) to structure, and to analyze hydrogen bonding networks. Once the various regions and cross peaks have been assigned, it should be possible to use pulse shaping learning algorithms (11, 31, 32) to selectively magnify or suppress the contributions of desired motifs, to study them separately. This could be particularly useful for probing protein folding pathways.

We thank Dr. Thomas la Cour Jansen for many helpful discussions. This work was supported by National Institutes of Health Grant 1R01 GM59230-03 and National Science Foundation Grant CHE-0132571.

1. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley Interscience, New York).
2. Kulbelka, J., Hofrichter, J. & Eaton, W. A. (2004) *Curr. Opin. Struct. Biol.* **14**, 76–88.
3. Mukamel, S. (2000) *Annu. Rev. Phys. Chem.* **51**, 691–729.
4. Bredenbeck, J., Helbing, J., Behrendt, R., Renner, C., Moroder, L., Wachtveitl, J. & Hamm, P. (2003) *J. Phys. Chem. B* **107**, 8654–8660.
5. Asplund, M. C., Zanni, M. T. & Hochstrasser, R. M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8219–8224.
6. Golonzka, O., Khalil, M., Demirdoven, N. & Tokmakoff, A. (2001) *Phys. Rev. Lett.* **86**, 2154–2157.
7. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. (2005) *Nature* **433**, 128–132.
8. Kwac, K. & Cho, M. H. (2003) *J. Chem. Phys.* **119**, 2247–2255.
9. Woutersen, S., Pfister, R., Hamm, P., Mu, Y. G., Kosov, D. S. & Stock, G. (2002) *J. Chem. Phys.* **117**, 6833–6840.
10. Jansen, T. I. C., Zhuang, W. & Mukamel, S. (2004) *J. Chem. Phys.* **121**, 10577–10598.
11. Mukamel, S. & Abramavicius, D. (2004) *Chem. Rev.* **104**, 2073–2098.
12. Abramavicius, D., Zhuang, W. & Mukamel, S. (2004) *J. Phys. Chem. B* **108**, 18034–18045.
13. Demirdoven, N., Cheatum, C. M., Chung, H. S., Khalil, M., Knoester, J. & Tokmakoff, A. (2004) *J. Am. Chem. Soc.* **126**, 7981–7990.
14. Tomovic, R. & Vukobratovic, M. (1972) *General Sensitivity Theory* (Elsevier, New York).
15. Dreyer, J., Moran, A. M. & Mukamel, S. (2003) *J. Phys. Chem. B* **107**, 5967–5985.
16. Yuan, X. M., Downing, A. K., Knott, V. & Handford, P. A. (1997) *EMBO J.* **16**, 6659–6666.
17. Torii, H. & Tasumi, M. (1998) *J. Raman Spectrosc.* **29**, 81–86.
18. Torii, H. & Tasumi, M. (1992) *J. Chem. Phys.* **96**, 3379–3387.
19. Cha, S. Y., Ham, S. H. & Cho, M. H. (2002) *J. Chem. Phys.* **117**, 740–750.
20. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217.
21. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995) *J. Chem. Phys.* **103**, 8577–8593.
22. Hoover, W. G. (1985) *Phys. Rev. A* **31**, 1695–1697.
23. Vangunsteren, W. F. & Berendsen, H. J. C. (1977) *Mol. Phys.* **34**, 1311–1327.
24. Hochstrasser, R. M. (2001) *Chem. Phys.* **266**, 273–284.
25. Andrews, D. L. & Thirunamachandran, T. (1977) *J. Chem. Phys.* **67**, 5026–5033.
26. Leegwater, J. A. & Mukamel, S. (1994) *J. Chem. Phys.* **101**, 7388–7398.
27. Scheurer, C. & Mukamel, S. (2001) *J. Chem. Phys.* **115**, 4989–5004.
28. Mukamel, S. & Tortschanoff, A. (2002) *Chem. Phys. Lett.* **357**, 327–335.
29. Fulmer, E. C., Mukherjee, P., Krummel, A. T. & Zanni, M. T. (2004) *J. Chem. Phys.* **120**, 8067–8078.
30. Barron, L. D. (2004) *Molecular Light Scattering and Optical Activity* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.
31. Abramavicius, D. & Mukamel, S. (2004) *J. Chem. Phys.* **120**, 8373–8378.
32. Levis, R. J. & Rabitz, H. A. (2002) *J. Phys. Chem. A* **106**, 6427–6444.