J|A|C|S ARTICLES

Published on Web 05/18/2010

Simulation Study of Chiral Two-Dimensional Ultraviolet Spectroscopy of the Protein Backbone

Darius Abramavicius,[†] Jun Jiang,[†] Benjamin M. Bulheller,[‡] Jonathan D. Hirst,*[,‡] and Shaul Mukamel*[,†]

Department of Chemistry, University of California, Irvine, California, and School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

Received March 8, 2010; E-mail: jonathan.hirst@nottingham.ac.uk; smukamel@uci.edu

Abstract: Amide n$-\pi^*$ and $\pi-\pi^*$ excitations around 200 nm are prominent spectroscopic signatures of the protein backbone, which are routinely used in ultraviolet (UV) circular dichroism for structure characterization. Recently developed ultrafast laser sources may be used to extend these studies to two dimensions. We apply a new algorithm for modeling protein electronic transitions to simulate two-dimensional UV photon echo signals in this regime and to identify signatures of protein backbone secondary (and tertiary) structure. Simulated signals for a set of globular and fibrillar proteins and their specific regions reveal characteristic patterns of helical and sheet secondary structures. We investigate how these patterns vary and converge with the size of the structural motif. Specific chiral polarization configurations of the UV pulses are found to be sensitive to aspects of the protein structure. This information significantly augments that available from linear circular dichroism.

Introduction

The diverse functionality of proteins derives from the variety of stable and flexible three-dimensional geometrical structures.[1] Numerous proteins have evolved, each contributing to a specific biological function. Two broad classes of structures can be distinguished: fibrous and globular. Members of the first class of proteins form long fibers; they are usually insoluble in water and perform mostly structural functions. Globular proteins are mesoscopic balls of polypeptide chains. Their functionality ranges from hormones and enzymes to antibodies and cell structural elements. Helical and sheet structural motifs can often be recognized in globular proteins. Fibrillar proteins can be made of long fibers with helical or sheetlike secondary structure. Probing the structure of proteins and their transformations is essential for understanding fundamental cell biology as well as for protein engineering.

The atomistic structure of proteins, if they can be crystallized, can be determined using X-ray crystallography. Nuclear magnetic resonance (NMR) techniques can provide structures in solution. However, some large proteins cannot be studied using such techniques. Examples are amyloid fibrils (because of insolubility) and intermediate steps in protein folding (because of limited time resolution). In this paper, we study the possibility of unravelling and/or refining protein structure using optical methods. Two-dimensional (2D) resonance laser spectroscopy in the infrared and visible regions has become an important technique for studying vibrational and electronic excitations of molecules and their complexes.[2] These properties are unique for specific molecular structures and the underlying electronic dynamics. 2D spectroscopy[3−15] is very sensitive to electronic correlations and ultrafast electronic dynamics in proteins and other systems, such as molecular and semiconductor nanostructures. It has great potential for structure refinement but is still not sufficiently mature to be routinely applied.

Circular dichroism (CD) spectroscopy in the 180−220 nm UV region is a well-established technique for estimating protein

[†] University of California.
[‡] University of Nottingham.

(1) Whitford, D. *Proteins: Structure and function*; Wiley: New York, 2005.
(2) Mukamel, S., Tanimura, Y., Hamm, P., Eds. *Accounts of Chemical Research: Special Issue on Coherent Multidimensional Optical Spectroscopy*; American Chemical Society: Washington, DC, 2009; Vol. 42.
(3) Mukamel, S.; Abramavicius, D.; Yang, L.; Zhuang, W.; Schweigert, I. V.; Voronine, D. *Acc. Chem. Res.* **2009**, *42*, 553–562.
(4) Zanni, M. T.; Ge, N.-H.; Kim, Y. S.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11265–11270.
(5) Sul, S.; Karaiskaj, D.; Jiang, Y.; Ge, N. H. *J. Phys. Chem. B* **2006**, *110*, 19891–19905.
(6) Engel, G. S.; Calhoun, T. R.; Read, E. L.; Ahn, T. K.; Mančal, T.; Cheng, Y. C.; Blankenship, R. E.; Fleming, G. R. *Nature* **2007**, *446*, 782–786.
(7) Read, E. L.; Engel, G. S.; Calhoun, T. R.; Mančal, T.; Ahn, T. K.; Blankenship, R. E.; Fleming, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14203–14208.
(8) Chung, H.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 612–617.
(9) Cho, M. *Chem. Rev.* **2008**, *108*, 1331–1418.
(10) Milota, F.; Sperling, J.; Nemeth, A.; Abramavicius, D.; Mukamel, S.; Kauffmann, H. F. *J. Chem. Phys.* **2009**, *131*, 054510.
(11) Stone, K. W.; Gundogdu, K.; Turner, D. B.; Li, X.; Cundiff, S. T.; Nelson, K. A. *Science* **2009**, *324*, 1169–1173.
(12) Li, Z.; Abramavicius, D.; Mukamel, S. *J. Am. Chem. Soc.* **2008**, *130*, 3509–3515.
(13) Mukamel, S. *Annu. Rev. Phys. Chem.* **2000**, *51*, 691–729.
(14) Zhuang, W.; Hayashi, T.; Mukamel, S. *Angew. Chem.* **2009**, *48*, 3750–3781.
(15) Abramavicius, D.; Palmieri, B.; Voronine, D. V.; Šanda, F.; Mukamel, S. *Chem. Rev.* **2009**, *109*, 2350–2408.

secondary structure content.[16] It relies on a fundamental property of the interaction between light and matter: isotropic chiral matter tends to rotate the polarization plane of linearly polarized light, and the rotation angle is intimately related to the bulk size and extent of structural chirality.[17] While it is not possible to construct the structure using solely CD data, characterization of structural motifs is possible. α-Helical proteins show a strong positive peak at 190 nm and usually a negatively red-shifted doublet (208 and 222 nm). Sheet-containing proteins are less rigidly organized,[18] and their CD spectra vary from one protein to the other. However, their common feature is a negative amplitude at 180 nm, a positive band at 195 nm, and usually a negative broad peak at approximately 215 nm.[19] The positive features are usually weaker for sheetlike than for α-helical proteins. Elaborate numerical methods based on least-squares fitting, singular-value decomposition, neural network analysis, etc., have been applied to extract the content of various structural motifs from experimental CD spectra.[20−22] Considerable effort has been devoted to simulation of the CD spectra of proteins. Efficient protein parametrization schemes have been developed using a Frenkel exciton (matrix) model. Woody and co-workers have optimized a set of semiempirical parameters, where some parameters were obtained by fitting to experiment.[23] Hirst and co-workers have developed an ab initio parametrization, where all transition parameters and intergroup couplings are calculated from first-principles electronic structure calculations.[18,24] This latter parametrization has a greater predictive capacity, since the parameters are not derived empirically from experiments.

Molecular dynamics (MD) simulations are gradually acquiring the capacity to unravel the structure of peptides and small proteins.[25−28] 2D spectroscopy simulations based on MD trajectories have been successfully employed in conjunction with 2DIR photon echo experiments to probe protein structure and fast folding processes and can be used to evaluate the quality of different MD methods.[14] UV spectroscopy provides a different observation window, which has the advantage over IR, because it can reach shorter pulse durations (5 fs compared to 50 fs) and higher-quality polarization control. The UV region is also important for specific photophysical protein processes, photochemical reactions of peptide bonds, and radiation damage. In a previous two-dimensional ultraviolet spectroscopy (2DUV) study,[29] we have used two exciton parametrization models, the first-principles one[18] and the semiempirical one,[23] and we used a diagonal Gaussian disorder of transition energies to model the spectral line width. Both parametrization models yielded very similar absorption and CD spectra of myoglobin. 2DUV, however, showed different signals and can, thus, be used to refine the model. In this paper, we employ a newly developed, quantum mechanics/molecular mechanics (QM/MM) approach[30] to improve the modeling of the amide electronic transition energies. Compared to the protein model of earlier work,[18] in which the transition energy is fixed and independent of the amino acid configuration, we allow the transition energy to fluctuate with the MD trajectory, using an electrostatic model. In contrast to our previous study,[29] the new model defines the transition frequency fluctuations from the microscopic electrostatic interaction among the amide groups, side chains, and water. Our new model can be transferred to different solvents and environments (e.g., membranes) and can accurately describe spectral line broadenings in various UV experiments. Simulations of absorption and CD based on this model[30] showed that the water fluctuations are fast and induce homogeneous spectral broadenings. Side chain fluctuations, however, are much stronger and slower and cause a large inhomogeneous broadening. This dominates the CD signal, and some of the fine details of the spectra are washed out. The 2DUV photon echo technique has the capacity to eliminate certain types of inhomogeneities and improve the resolution. Recent advances in laser technology have led to the development of coherent ultrashort pulses in the UV.[31] Ultrafast attosecond broadband UV and XUV pulses under development could make 2DUV feasible.[32−35] Here we simulate 2D UV signals of protein backbones of several proteins and their segments and demonstrate that the cross-peaks in 2DUV carry unique signatures of protein structure.

Methods

Several simulation techniques for nonlinear optical signals of molecular systems and their complexes relevant for the resonance electronic spectroscopy have been developed.[14,15,36] We use a computational protocol based on the exciton model, i.e., the matrix of single-exciton transition energies and intergroup (and intragroup) interactions calculated using electrostatic interactions between transition charge densities. The level scheme for amide $n-\pi^*$ and $\pi-\pi^*$ transitions, their transition dipoles, and interamide interactions are implemented as in the package DichroCalc.[18,37]

Fluctuations in the molecular environment induce spectral broadenings. To account for these effects, we use a QM/MM approach. We consider a dilute solution so that interactions between distinct protein molecules are negligible. MD simulations of a single protein molecule in water were performed with a time step of 1 fs using NAMD version 2.7,[25] with the CHARMM27[26] force field, and the TIP3P water model,[38] and cubic periodic boundary conditions. Long-range electrostatic interactions were computed using particle-mesh Ewald (PME),[27,28] and a real space cutoff of 12 Å was used for nonbonded interactions. The NPT ensemble was used. All MD simulations started with a 5000-step minimization and heating for 600 ps from 0 K to room temperature (310 K). After a

(16) Berova, N., Nakanishi, K., Woody, R. W., Eds. *Circular Dichroism. Principles and Applications*, 2nd ed.; John Wiley & Sons, Inc.: New York, 2000.
(17) Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 2004.
(18) Bulheller, B. M.; Rodger, A.; Hirst, J. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2020–2035.
(19) Greenfield, N. J. *Anal. Biochem.* **1996**, *235*, 1–10.
(20) Greenfield, N. J. *Methods Enzymol.* **2004**, *383*, 282–317.
(21) Whitmore, L.; Wallace, B. A. *Nucleic Acids Res.* **2004**, *32*, W668–W673.
(22) Whitmore, L.; Wallace, B. A. *Biopolymers* **2008**, *89*, 392–400.
(23) Woody, R. W. *Monatsh. Chem.* **2005**, *136*, 347–366.
(24) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 9636–9644.
(25) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
(26) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(27) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(29) Li, Z.; Abramavicius, D.; Zhuang, W.; Mukamel, S. *Chem. Phys.* **2007**, *341*, 29–36.
(30) Jiang, J., Abramavicius, D., Bulheller, B. M., Hirst, J. D., Mukamel, S. *J. Phys. Chem. B* **2010**, in press.
(31) Corkum, P., De Silvestri, S., Nelson, K., Riedle, E., Schoenlein, R., Eds. *Ultrafast Phenomena XVI*; Springer: Berlin, 2009.
(32) Bressler, C.; Chergui, M. *Chem. Rev.* **2004**, *104*, 1781–1812.
(33) Kapteyn, H.; Cohen, O.; Christov, I.; Murnane, M. *Science* **2007**, *317*, 775–778.
(34) http://ssrl.slac.stanford.edu/lcls.
(35) Krausz, F.; Ivanov, M. *Rev. Mod. Phys.* **2009**, *81*, 163–234.
(36) Cho, M. *Two-Dimensional Optical Spectroscopy*; CRC Press: Boca Raton, FL, 2009.
(37) Bulheller, B. M.; Hirst, J. D. *Bioinformatics* **2009**, *25*, 539–540.
(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

2 ns equilibration, we simulated a 16 ns trajectory and recorded structures every 400 fs. The protein structures were stable during the MD simulations. An ensemble of 2000 MD snapshots was recorded every 400 fs.

The electronic transitions are described by the Hamiltonian

$$\hat{H}_s = \sum_{ma} \varepsilon_{ma}^{(s)} \hat{B}_{ma}^{\dagger} \hat{B}_{ma} + \sum_{manb}^{m \neq n} J_{ma,nb}^{(s)} \hat{B}_m^{\dagger} \hat{B}_n \qquad (1)$$

where m denotes the peptide unit, a represents the transition of the unit (either n$-\pi^*$ or $\pi-\pi^*$), and (s) denotes the snapshot. $\varepsilon_{ma}^{(s)}$ represents the transition energy of state a of the mth unit, and $J_{ma,nb}^{(s)}$ denotes the resonant coupling. The exciton creation ($\hat{B}_{ma}^{\dagger}$) and annihilation ($\hat{B}_{ma}$) operators have commutation relations $[\hat{B}_{ma}, \hat{B}_{nb}^{\dagger}] = \delta_{mn}(1 - 2\hat{B}_{mb}^{\dagger}\hat{B}_{ma})$.

The transition energies of the mth peptide unit, ε_{ma}, were calculated by considering the electric field at this unit induced by water, the remaining peptide groups, and the amino acid side chains, as described in ref 30. The interamide interactions and transition amplitudes were calculated with the DichroCalc package.[18] Single-exciton eigenstates were calculated by numerical diagonalization of eq 1. This model reproduces realistic spectral broadening effects based on a MD trajectory using an electrostatic model.[30]

The 2DUV signal is obtained using four coherent short time-ordered laser pulses, labeled by their wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$, and $\mathbf{k}_4$. The laser pulses have the same carrier frequency and are resonant with peptide backbone electronic transitions. The signal is detected at direction $\mathbf{k}_4 = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ as a function of three delay times: t_1, between pulses $\mathbf{k}_1$ and $\mathbf{k}_2$; t_2, between pulses $\mathbf{k}_2$ and $\mathbf{k}_3$; and t_3, between pulses $\mathbf{k}_3$ and $\mathbf{k}_4$. The 2D spectrum is obtained by two-dimensional Fourier transform $t_1 \rightarrow \Omega_1$ and $t_3 \rightarrow \Omega_3$. We set the second delay t_2 to zero. The 2DUV signal was calculated in the quasi-particle representation,[15] whereby each exciton is considered as a quasi-particle moving in real space. If two excitons approach the same amide group, they scatter due to Pauli exclusion. The scattering matrix was calculated with eq 325 of ref 15. We then calculate the homogeneous two-dimensional $\mathbf{k}_1$ photon echo signal for this snapshot (s) using eqs 119−122 and 127 of ref 15. A single-exciton dephasing rate (homogeneous broadening) ($\gamma = 250$ cm^{-1}) was added to represent the spectral line width induced by fast fluctuations of water molecules. The Hamiltonian and the homogeneous signal are calculated by this procedure at each snapshot of the full MD trajectory. The full inhomogeneous signal is obtained by averaging over 2000 MD snapshots for absorption and CD spectra, and 500 MD snapshots for 2D signals. All calculated 2DUV spectra were normalized to 1 and then were plotted from $-R$ to R ($0 < R < 1$) to enhance weak features.

The UV region relevant for the spectroscopy of the peptide backbone is from 180 to 220 nm. A 200 nm photon has a period T of 0.67 fs. We assume a Gaussian pulse envelope as a function of frequency. The length of the pulse with five optical periods is 3.33 fs [full width at half-maximum (fwhm)]. Its spectral bandwidth is 8840 cm^{-1} (the variance of the Gaussian shape is 3754 cm^{-1}). That approximately corresponds in wavelength to the 35 nm bandwidth (fwhm). We used these pulse shape parameters with a carrier central frequency ω_0 of 52000 cm^{-1} (192 nm) so that the full absorption band is covered (the full CD band is broader). Calculations were performed for both xxxx and xxxy polarization configurations: the xxxx configuration is the simplest, and the xxxy configuration is, like CD, induced by molecular chirality. We, thus, denote CD and 2D xxxy signals as chirality induced (CI). They are ∼100 times weaker than regular ones but have better structural resolution.

It can be shown using response function theory that the CD signal corresponds to the linear response function xy tensor component, where the excitation is at x polarization and the detection is polarized perpendicularly at y.[39] The 2D spectrum at xxxy polarizations corresponds to the case in which the first excitation has y polarization, and the second excitation has x polarization, just like in CD. The remaining two xx polarized interactions (absorption-like) perform the detection. The diagonal section of the 2DUV spectrum corresponds to the case in which the excitation and detection have the same energies, so the excited states are detected directly without interaction with other transitions (no coherent or incoherent energy transfer is involved). In this case, the diagonal of the 2DUV xxxy spectrum reflects the detection of CD-like features, which appear similar to the CD spectrum. Finite pulse bandwidths and nonlinear rescaling of peak amplitudes may distort this relation. The same consideration applies to 2DUV xxxx and linear absorption, which corresponds to the xx type linear response function. Therefore, the diagonal of the 2D xxxx spectrum should reflect the absorption spectrum. Both xxxx and xxxy pulse configurations can be realized in a collinear pulse configuration. For notation and details, see ref 15.

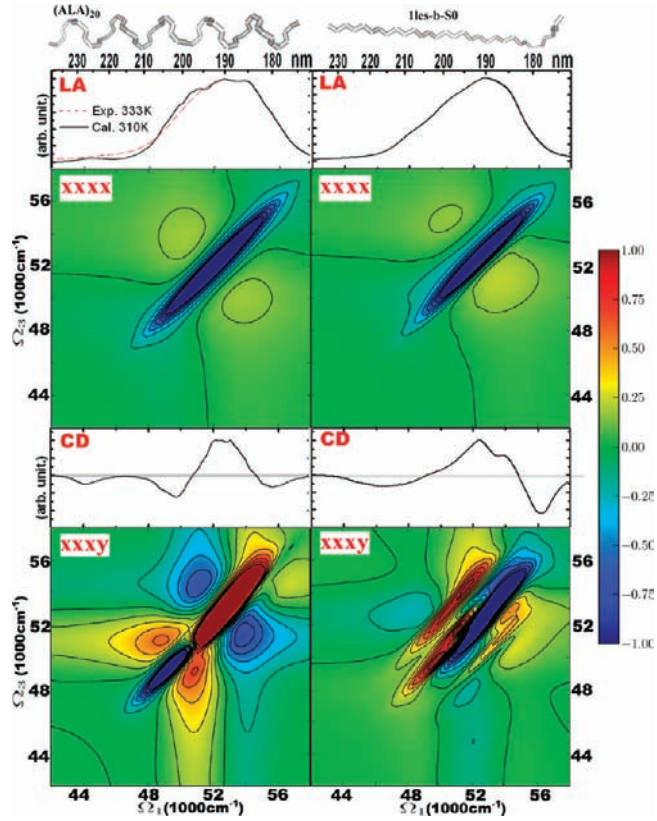

Figure 1. Dependence of the linear and 2D signals on the backbone configuration: helical (left column) vs extended strand (right column). From top to bottom: structures and absorption, 2DUV xxxx, CD, and 2DUV xxxy spectra. The extended strand is taken from lectil lectin (Thr1−Phe11 of chain C of Protein Data Bank entry 1les), and the α-helical conformation is a 20-residue polyalanine chain. The 2DUV spectral normalization (R) factors: −0.35 for helical xxxx, −0.16 for helical xxxy, −0.38 for strand xxxx, and −0.31 for strand xxxy.

Results

Connecting 2D Signals to Secondary Structure Elements. We first present the 2DUV spectra of typical protein structural motifs. In Figure 1, we compare the spectra of extended and helical conformations of the peptide backbone, shown in the top panels: the helical fragment is a small polyalanine α-helix, and the strand is excised from lentil lectin (Protein Data Bank

(39) Wagersreiter, T.; Mukamel, S. *J. Chem. Phys.* **1996**, *105*, 7995.

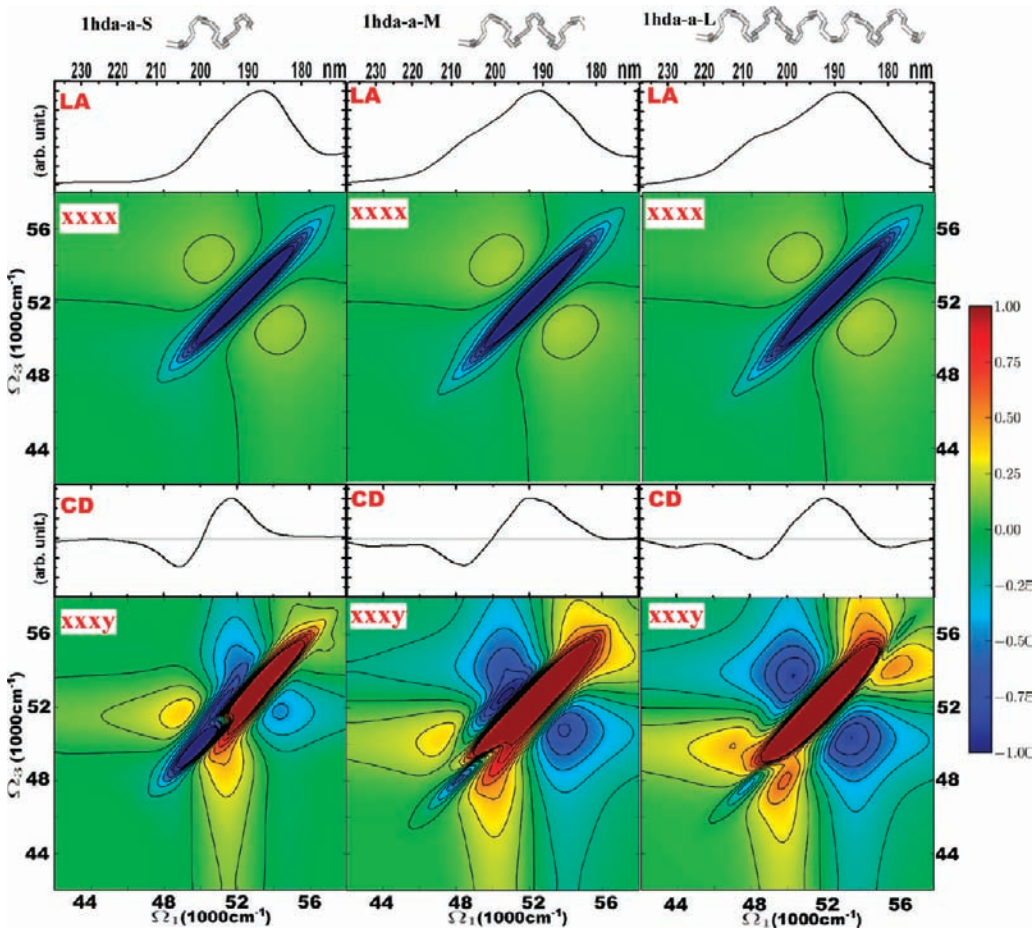

Figure 2. Linear and 2DUV signals of helical motifs of different lengths. The helices are taken from hemoglobin (PDB entry 1hda): 1hda-a-S, Pro124−Asp129 of chain D; 1hda-a-M, Pro124−Val134 of chain D; 1hda-a-L, Pro124−His143 of chain D. From top to bottom: structures and absorption, 2DUV xxxx, CD, and 2DUV xxxy spectra. The 2DUV spectral normalization (*R*) factors: −0.37 for S xxxx, −0.23 for S xxxy, −0.35 for M xxxx, −0.12 for M xxxy, −0.35 for L xxxx, and −0.11 for L xxxy.

entry 1les). For the sake of completeness, we also show linear absorption (LA) and CD spectra. The two LA signals are similar and show very broad asymmetric peaks at 53000 cm^{-1}. These features are mapped approximately onto the diagonal line ($\Omega_3 = -\Omega_1$) of the nonchiral 2DUV xxxx spectrum. For the helix and the sheet, it shows strong diagonal features and relatively weak cross-peaks. The patterns consist of negative (green-to-blue) and positive (green-to-red) features. The diagonal negative peak of the helix is slightly broader than that of the strand. The cross-peaks in the spectrum of the helix are symmetrically distributed around the negative diagonal peak, while the lower cross-peak (on the Ω_3 axis) of the extended strand is stronger.

We now turn to CI signals. The CD spectrum of the two conformations is very different below 50000 cm^{-1}: for the strand there is a single broad negative peak centered at 45500 cm^{-1}, while the helix shows a negative doublet (44000 and 48000 cm^{-1}).[20,23] Spectra along the diagonal approximately reflect the CD in the 48000−56000 cm^{-1} region. The diagonal section of the strand spectrum has a negative peak at 54000 cm^{-1} and a positive peak at 49500 cm^{-1}. Strong cross-peaks asymmetrically appear on both sides of the negative and positive peaks. The 2DUV pattern of the helical structure is markedly different. The negative 54000 cm^{-1} diagonal peak of the strand shifts to lower values of Ω_3, and a strong positive peak develops at 52000 cm^{-1}. The positive peak of the strand at 49500 cm^{-1} corresponds to the negative peak at 48000 cm^{-1} in the helix. This spectral redistribution around the diagonal results in a relatively symmetric cross-peak pattern. The cross-peak symmetry in both the xxxx and xxxy spectra is, thus, attributed to the helical structure and its interactions between neighboring turns of the helix.

We next explore how the symmetric cross-peak pattern develops with the length of the helix. In Figure 2, we show the 2DUV signals for three helical segments with different lengths excised from hemoglobin (PDB entry 1hda). We label the three helical segments as short (S), medium (M), and long (L). Their structures are shown at the top of Figure 2: S contains two helical turns, M has three helical turns, and L comprises five helical turns. The absorption of these helices demonstrates the development of the 49000 cm^{-1} shoulder with helix length. The variation of 2DUV xxxx with the helix length is relatively weak: the length of the negative region on the diagonal increases from the short to the medium helix. There are stronger differences in the CI signals. Differences in CD are mostly apparent in the growth of the negative 44000 cm^{-1} peak with helix length. The negative diagonal peak at 48000−50000 cm^{-1} significantly varies in the 2DUV xxxy spectra for different helices. That peak is strong for the small helix and gradually decreases with helix length. It could, thus, be used to determine the effective length of helices in globular proteins.

In Figure 3, we depict the variation of the 2DUV signal with the size of the sheet motif. Three motifs were taken from lectin: small (S), medium (M), and large (L). S has two strands, M

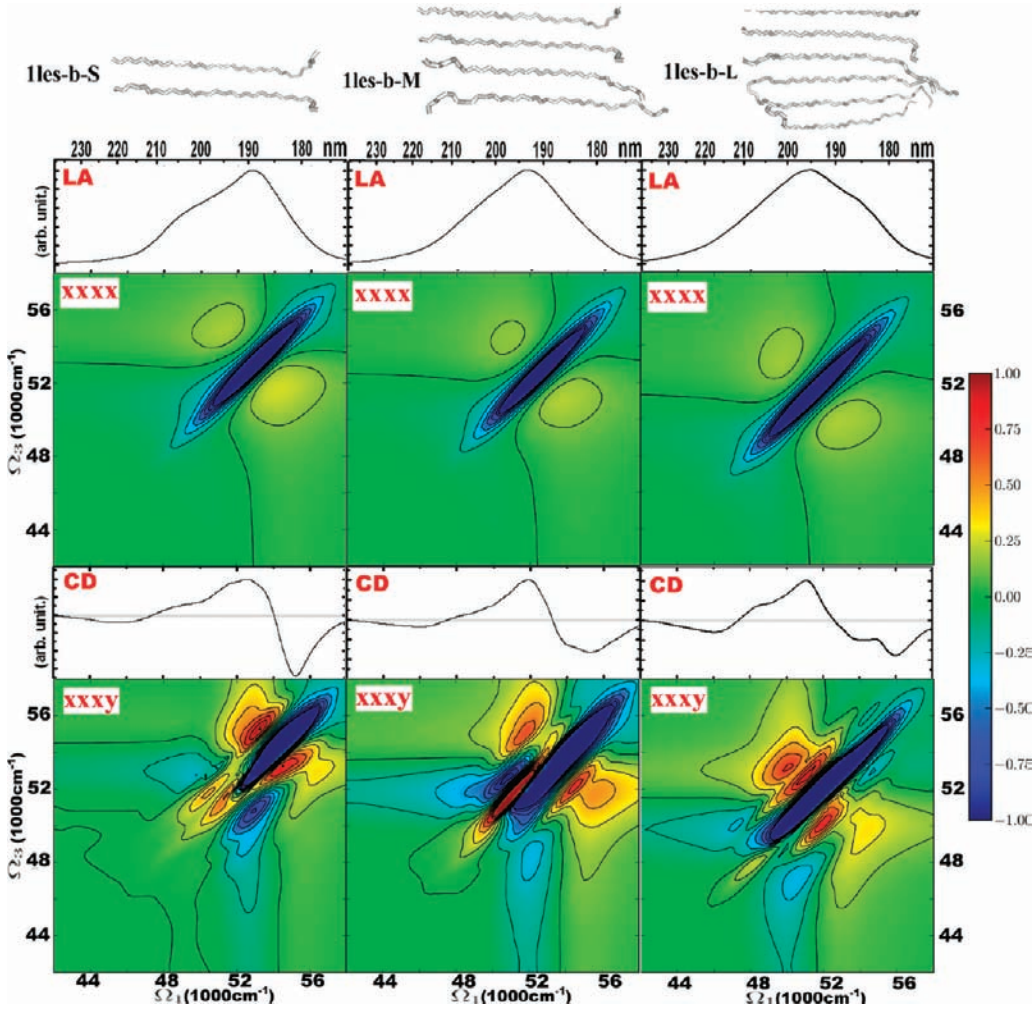

Figure 3. Linear and 2DUV signals of the β-sheets of various sizes. The sheet elements were obtained from lectin (PDB entry 1les) as follows. 1les-b-S is made of two strands: Thr1−Phe11 of chain C and Val37−Leu46 of chain D of lectin (PDB entry 1les). 1les-b-M is made of four strands: Thr1−Phe11 of chain C, Val37−Leu46 of chain D, Val60−Val70 of chain C, and Glu158−Ala169 of chain C. 1les-b-L is made of six strands: Thr1−Phe11 of chain C, Val37−Leu46 of chain D, Val60−Val70 of chain C, Glu158−Ala169 of chain C, Val1−Pro11 of chain D, and Val37−Leu46 of chain D. From top to bottom: structures and absorption, 2DUV xxxx, CD, and 2DUV xxxy spectra. The 2DUV spectral normalization (R) factors: −0.35 for S xxxx, −0.34 for S xxxy, −0.37 for M xxxx, −0.32 for M xxxy, −0.36 for L xxxx, and −0.12 for L xxxy.

four, and L six, in each case making an antiparallel β-sheet. The absorption signals of all three motifs are very similar. The 2D xxxx spectra of the sheets are also very similar and feature a single diagonal peak at 51200 cm^{-1}. The chiral CD spectra differ mostly in the shape of dispersive region between 48000 and 56000 cm^{-1}. These changes can be rationalized as follows. There is a gradual increase in a positive shoulder at 50000 cm^{-1} with the size of the sheet. These differences are amplified in the chiral 2D xxxy spectra in the 48000−52000 cm^{-1} region. The main negative diagonal peak becomes more elongated with sheet size. We also notice a variation in the fine structure of the cross-peaks.

2D UV Signals of Globular and Fibrillar Proteins. In Figure 4, we present the 2DUV spectra of four proteins with characteristic secondary and tertiary structures. Hemoglobin (PDB entry 1hda) is a globular protein made of interconnected helices. Less than 20% is nonhelical. Rabbit skeletal α-tropomyosin (PDB entry 2d3e) is a long helical protein that forms fibers. Its core consists of two intertwined α-helices. Lentil lectin (PDB entry 1les) is mostly a twisted β-sheet globular protein. Monellin (PDB entry 1mol) is a small globular protein made of a sheet and an α-helix, which allows us to study both structural motifs and their interactions.

The absorption spectra of these proteins are similar; they do not vary when many amides are involved. The same can be concluded for the 2DUV xxxx signal of these proteins, which is almost identical, showing one strong negative diagonal peak and a pair of weaker positive cross-peaks, positioned symmetrically on both sides of the diagonal. The CD features agree with experiment: the CD spectra of helical proteins have all features observed experimentally. Likewise, the CD spectra of the sheetlike proteins are similar to each other and to the experimental sheet spectra. The 2DUV CI xxxy signals of the proteins show rich spectral features, which strongly depend on structure. Helical proteins show a very strong positive diagonal peak at around 52000 cm^{-1}. The pattern is symmetric along the diagonal for the regular long helices (tropomyosin). The asymmetry in the hemoglobin spectrum can, thus, be attributed to the less regular distribution of the helices and is thus a signature of specific tertiary structure. Sheet-containing proteins exhibit a very strong diagonal negative peak at 54000 cm^{-1}.

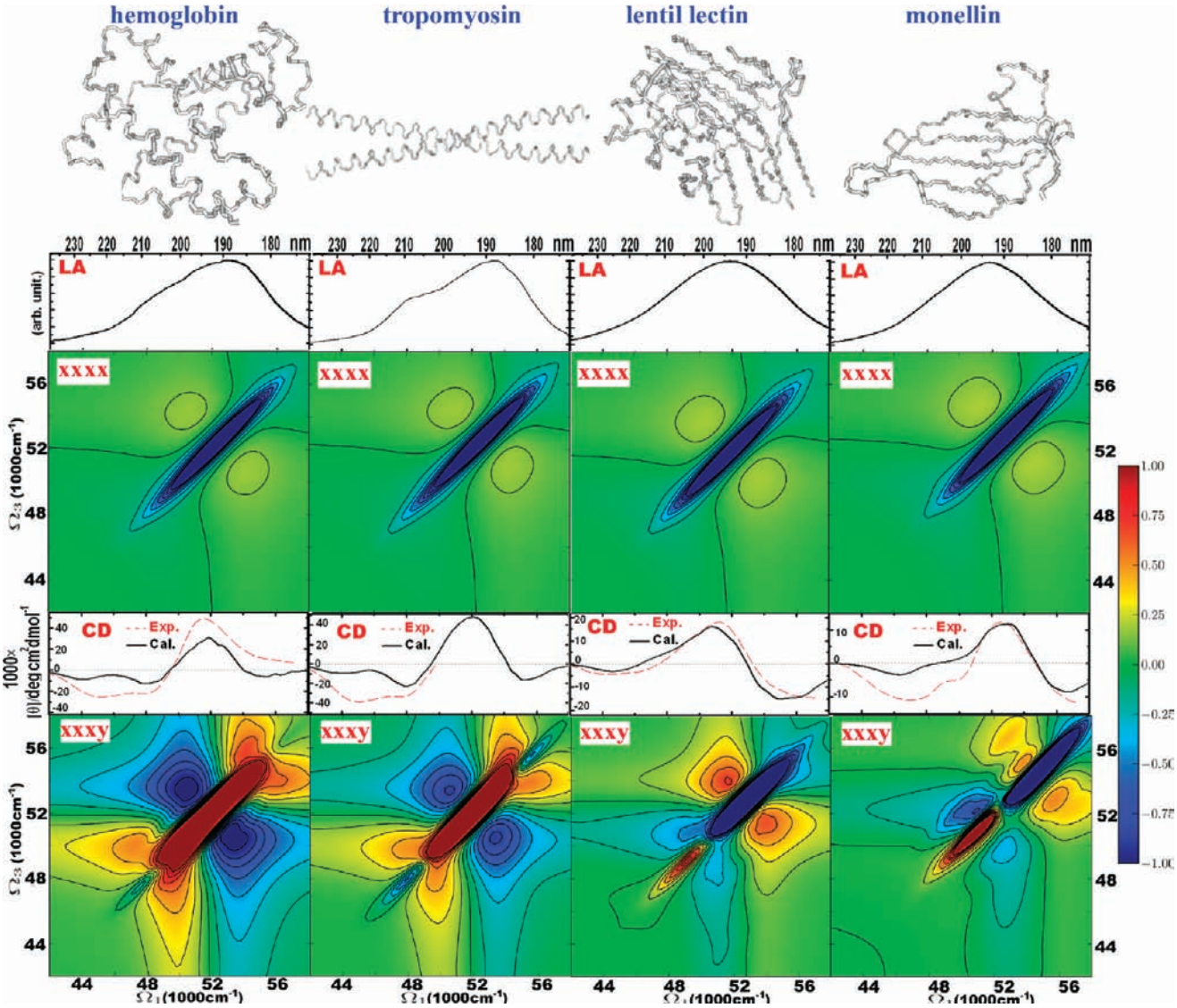

Figure 4. Simulated linear and 2D signals of helical proteins, hemoglobin (PDB entry 1hda), and tropomyosin (PDB entry 2d3e), and of sheet-containing proteins, lentil lectin (PDB entry 1les) and monellin (PDB entry 1mol). From top to bottom: structures and absorption, 2DUV xxxx, CD, and 2DUV xxxy spectra. The simulated CD is given in absolute units. The 2DUV spectral normalization (R) factors: −0.38 for hemoglobin xxxx, −0.07 for hemoglobin xxxy, −0.36 for tropomyosin xxxx, −0.11 for tropomyosin xxxy, −0.33 for for lentil lectin xxxx, −0.14 for lentil lectin xxxy, −0.32 for monellin xxxx, and −0.21 for monellin xxxy.

Since lectin has a large content of sheet and monellin is a mixture of helix and sheet, this difference can be related to the strength of the positive diagonal peak at 50000 cm^{-1} correlating with the helical content of monellin. Overall, the sheets show weaker cross-peaks than the helices.

Discussion

We have investigated the variation of 2DUV signals with respect to the size and geometry of secondary structure motifs. The absorption spectrum and the nonchiral xxxx 2DUV signal do not depend strongly on the secondary structure of a protein. The signal is dominated by a single negative diagonal peak, and cross-peaks are weak, indicating that the amide $\pi-\pi^*$ transition is dominated by inhomogeneities and interamide interactions are weak. This is markedly different from the amide vibrational bands, where 2DIR spectroscopy shows strong cross-peak patterns.

Our previous simulation of 2DUV spectra of helical proteins used a cruder parametrization of amide transitions.[29] The simulation parameters presented here are more accurate and show much better agreement with CD experiments, especially for helical proteins. Our previous simulations showed stronger cross-peaks in the 2D signals. This arose from smaller dephasing rates; as a result, the peaks across the diagonal ($|\Omega_3| = \text{const} - |\Omega_1|$) were narrower, their overlap was smaller, and this provided less cancellation. In the model presented here, the 250 cm^{-1} homogeneous dephasing rate was estimated from absorption line shapes by including only water-induced fluctuations. Inhomogeneity was obtained by adding averaging over fluctuations of the whole protein. In this way, we simulate all broadening mechanisms microscopically, and the simulated spectra must be more accurate. The weak cross-peaks in the simulated 2D xxxx signal are an indication that the interamide electrostatic interaction is weak compared to the amide transition energy fluctuations.

We observed a strong dependence of the chiral xxxy 2DUV signal on the secondary structure. This signal shows characteristic patterns for sheets and helices. By comparing signals of

different helical proteins, we conclude that the helical protein signal is dominated by a strong positive diagonal peak. The signal for different proteins is very similar; however, the differences indicate sensitivity to tertiary structure.

Can the observed features be related to long-range structural order of proteins, which is preserved through hydrogen bonding? To address this issue, we should determine whether 2DUV spectra depend on nearest-neighbor interactions, which are a function of rotation angles between neighboring peptide groups, or on long-range electrostatic interactions that may be inherent in specific helical or sheet structures. Since the interamide interactions are very weak (see 2DUV xxxx signals), we expect that long-range interactions do not a play significant role in shaping 2DUV spectra of the proteins. Instead, the diagonal features depend on the dihedral angles between neighboring amides through the resulting shifts in the transition energies. That is why the negative diagonal feature of a single extended strand is similar to that of the whole β-sheet. We also found that some nonhelical proteins show a strong positive peak in the 2DUV xxxy signal (not shown). This can be related to specific turns in the protein structure: the helical strand and extended strand (or sheet structure) have different angles between neighboring amides. The difference between diagonal features in the 2DUV xxxy signals, thus, originates from different local configurations. The cross-peaks between these diagonal peaks, while weak, are induced by the long-range interactions.

Another important question addressed here is the variation of the 2DUV spectra with the size of specific structural motifs. Our simulations showed that only the cross-peak regions of the chirality-induced 2DUV xxxy signal is sensitive to size. This supports the conclusion that the cross-peaks may be related to the long-range interactions between amide groups. However, the spectral variation with size (for helices and for sheets) is not monotonic in this set of protein structures, and it is difficult to establish the relation between the size of a structural motif and the 2DUV spectrum.

Further studies are necessary to explore these issues in more detail. Our 2DUV simulations demonstrate the high sensitivity of the chirality-induced 2D signals to the secondary structure of proteins. Other types of 2D signals have been used in electronic spectroscopy of photosynthetic molecules[40,41] and semiconductor quantum wells[42] to reveal electronic correlations of excitons. These signals may be informative in UV studies of proteins as well.

Coherent 2D experiments in the UV protein backbone region will become available in the near future. It is important to develop efficient and accurate simulation models. Our simulations include few important assumptions. (i) We assumed that only single excitation can occupy an amide group, and we also neglected excited state absorption of amide groups into their electronic states lying around $\sim$100000 cm^{-1} (from their ground states). Such electronic states (Rydberg-like) are substantially destabilized in condensed phases relative to the gas phase and appear to have a limited influence. (ii) We neglected charge-tranfer (CT) states of the protein backbone.[43] These may be important in shaping the higher-energy (180 nm) area of UV spectra.[44] (iii) Electronic interactions with near-resonant side chain excitations were neglected. (iv) We used a simple dephasing model as induced by fast water molecule fluctuations. It is, however, known that the water molecule fluctuation correlation function may extend to picoseconds;[45] thus, the homogeneous model (Markovian approximation) may fail. Other fluctuations were assumed to be much slower than the experimental time scale. A full fluctuation analysis may be performed in the future. (v) Dissipative energy transport between amides was neglected. This may not be a significant factor for our simulation scheme, but it may affect spectral line shapes through lifetime broadening. (vi) The 2DUV signal was simulated for a laser carrier frequency of 52000 cm^{-1} with a 8000 cm^{-1} bandwidth enhancing the $\pi-\pi^*$ transitions close to this region. By tuning the carrier frequency, we can enhance different electronic transitions, and the resulting signal may exhibit different patterns. Using laser pulses with different central frequencies (two- or three-color setup) may enhance the cross-peak regions. The full bandwidth of the amide CD in the UV region is 170$-$250 nm, which corresponds to $\sim$18000 cm^{-1}. Different sections of this bandwidth may be studied using state-of-the-art laser techniques.

Acknowledgment. We gratefully acknowledge the support of the National Institutes of Health (Grants GM059230 and GM091364) and the National Science Foundation (Grant CHE-0745892). B.M.B. was the grateful recipient of an Early-Stage Researcher Short Visit award from the Collaborative Computational Project for Biomolecular Simulation. J.D.H. thanks the Leverhulme Trust for a research fellowship.

Note Added after ASAP Publication. An uncorrected version of this paper published ASAP May 18, 2010, the corrected version reposted June 2, 2010.

JA101968G

(40) Read, E. L.; Schlau-Cohen, G. S.; Engel, G. S.; Wen, J.; Blankenship, R. E.; Fleming, G. R. *Biophys. J.* **2008**, *95*, 847–856.
(41) Abramavicius, D.; Voronine, D. V.; Mukamel, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8525.
(42) Stone, K. W.; Gundogdu, K.; Turner, D. B.; Li, X.; Cundiff, S. T.; Nelson, K. A. *Science* **2009**, *324*, 1169–1173.
(43) Oakley, M. T.; Hirst, J. D. *J. Am. Chem. Soc.* **2006**, *128*, 12414–12415.
(44) Bulheller, B. M.; Miles, A. J.; Wallace, B. A.; Hirst, J. D. *J. Phys. Chem. B* **2008**, *112*, 1866–1874.
(45) Eaves, J. D.; Tokmakoff, A.; Geissler, P. L. *J. Phys. Chem. A* **2005**, *109*, 9424–9436.