THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

Signatures of the Protein Folding Pathway in Two-Dimensional Ultraviolet Spectroscopy

Jun Jiang,*[,†,‡] Zaizhi Lai,[§] Jin Wang,*[,§,∥] and Shaul Mukamel*[,‡]

[†]Department of Chemical Physics, University of Science and Technology of China, No. 96, JinZhai Road Baohe District, Hefei 230026, China

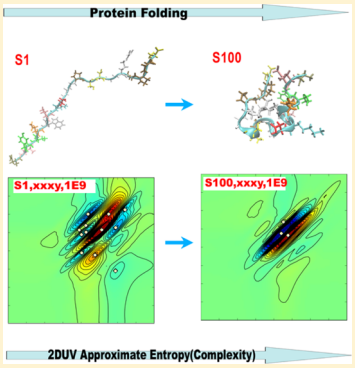
[‡]Chemistry Department, University of California Irvine, 433A Rowland Hall, Irvine, California 92697, United States

[§]Department of Chemistry and Physics, University of New York at Stony Brook, Stony Brook, New York 11794, United States

[∥]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, No. 5625, Ren Min Street, Changchun, Jilin 130021, China

S *Supporting Information*

ABSTRACT: The function of protein relies on their folding to assume the proper structure. Probing the structural variations during the folding process is crucial for understanding the underlying mechanism. We present a combined quantum mechanics/molecular dynamics simulation study that demonstrates how coherent resonant nonlinear ultraviolet spectra can be used to follow the fast folding dynamics of a mini-protein, Trp-cage. Two dimensional ultraviolet signals of the backbone transitions carry rich information of both local (secondary) and global (tertiary) structures. The complexity of signals decreases as the conformational entropy decreases in the course of the folding process. We show that the approximate entropy of the signals provides a quantitative marker of protein folding status, accessible by both theoretical calculations and experiments.



SECTION: Biophysical Chemistry and Biomolecules

Protein folding is an important biological process whereby polypeptides form a unique functional 3D structure. Experimental probing of protein folding pathways along the free-energy landscape[1] is critical for the understanding of its mechanism, and the manipulation of protein structure and function.[2] Thanks to significant advances in computation power, the atomic level description of protein folding has become feasible through molecular dynamics (MD) simulations.[3,4] The experimental monitoring of the folding process at the atomic level is difficult due to the lack of suitable fast tools. Nuclear magnetic resonance (NMR) spectra[5] give high-quality data but require extensive data processing. Circular dichroism (CD) spectroscopy[6] probes protein secondary structures with low resolution and cannot reveal tertiary or global structures, making it difficult to view different folding stages. More importantly, the absence of a good indicator of global protein geometry makes it hard to correlate the simulated structure dynamics with experimentally probed folding status.

Coherent ultrafast 2D electronic spectroscopy has emerged recently as a new structure refinement tool.[7−9] The photon echo signal records the response of molecules to sequences of ultrafast (femtosecond) coherent laser pulses, whose 2D correlation plots reflect the variation of the nonlinear response functions with the controlled time intervals. They reveal electronic couplings in proteins and interactions among residues, which characterize time evolving structures. For instance, we have demonstrated in theory the utilization of 2D infrared (IR)[10] and 2D Raman[11] to capture the local change of certain structural elements during protein folding. Recent advances in laser sources[12−14] make it possible to extend this technique into the ultraviolet (UV) regime. UV photoresponse of peptide involves electronic transitions of each peptide bond along the protein backbone. In comparing to IR and Raman spectroscopic tools, which mostly reflect local details, two-dimensional ultraviolet (2DUV) spectroscopy of proteins, carrying abundant information about inter- and intramolecular interactions, is particularly useful in identifying both the local details and global geometries of protein structures.[15−17]

In this work, we perform quantum mechanics/molecular dynamics (QM/MM) simulations of 2DUV spectroscopy of the folding process of the Trp-cage peptide. Trp-cage is one of the most widely studied peptides[18] and is a convenient model system for theoretical investigations of its fast nanosecond folding dynamics. The 2DUV technique offers a good indicator of protein global geometry and provides an powerful tool to probe the protein folding process.

The Trp-cage contains 20 residues with the sequence "NLYIQWLKDGG PSSGRPPPS". The extended conformation

Received: February 1, 2014
Accepted: March 19, 2014
Published: March 19, 2014

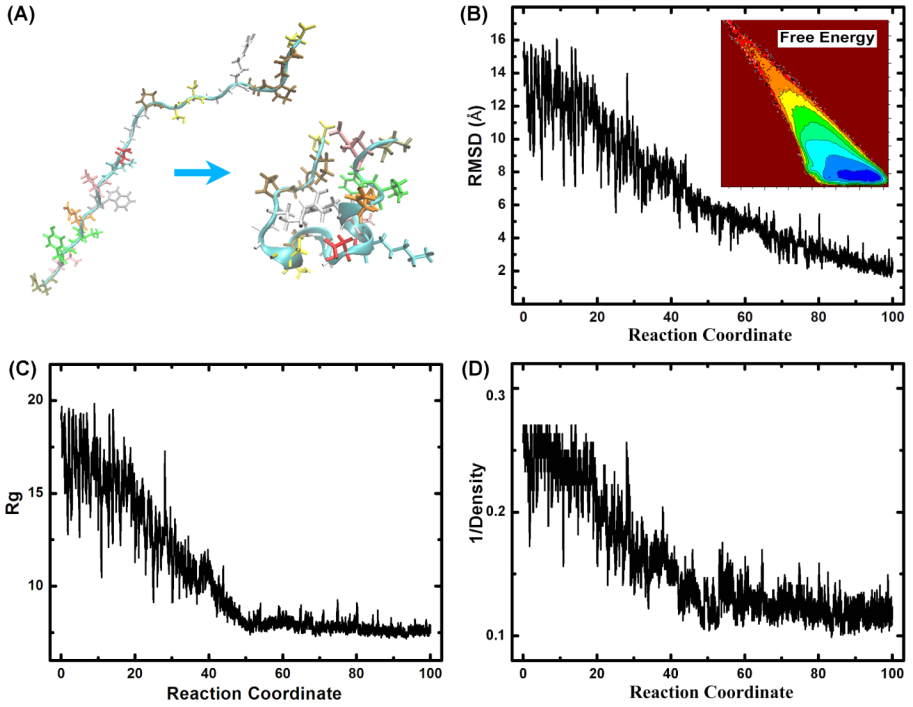


Figure 1. (A) From the unfolded strand to folded cage structure of a Trp-cage protein (PDB code: 1L2Y). The backbone trace is shown as a ribbon, and the side chains are depicted with wires. The RMSD (B), R_g (C), and inverse of packing density (D) along the free-energy landscape of Trp-cage folding process (from S1 to S100). The inset in panel B shows the free-energy landscape.

was built up as the initial structure, based on which 50 trajectories with different initial conditions were simulated at 315 K for 200 ns (details in the Supporting Information). These trajectories cover 10 μs simulations of protein folding, enough for building the free-energy landscape. The free energy was calculated as $F = -\log(P)$, where P was the population obtained from all the 10 μs MD simulated data, as shown in the inset graph of Figure 1A. 100 state points along the folding pathway on the energy landscape denoted S1, S2, ..., S100, were selected by the clustering method. To model the geometric fluctuations, we chose 100 subconformations (MD snapshots) around each state point. The variations of RMSD (root mean square deviation) and R_g (Radius of gyration) along the folding path are shown in Figure 1B,C, respectively. Because Trp-cage folding is also a packing process from a strand to a compact cage, the evolution of tertiary structure can be characterized by the protein packing density.[19] We computed the packing density as the average of the number of residue's Cα atoms within a 9 Å radius of the Cα atom of a given residue. The evolution of the inverse of packing density displayed in Figure 1C is consistent with the R_g. The packing density is closely related to the protein conformation entropy, and its evolution suggests that Trp-cage conformation entropy decreases along the packing (folding) process. Obviously, both the R_g and packing density (conformation entropy) are good markers of the protein folding status. Unfortunately, these are not readily accessible experimentally.

Quantum mechanics (QM) calculations were employed to compute ultraviolet signals based on MD snapshots (details in the Supporting Information). Calculations of the excited states by CASSCF/SCRF (the complete-active space self-consistent-field) method has provided parameters for the transition energies of isolated peptide bond unit (see Figure S1 in the Supporting Information), the resonant couplings, and electric and magnetic dipole moments.[20] The electronic structures and electrostatic potential distributions of the peptide bonds and amino side chains in proteins were computed at the density functional theory (DFT) level B3LYP/6-311++G**. The Frenkel exciton Hamiltonian is the most practical model to describe the photoresponse of molecular aggregates[21]

$$\hat{H} = \sum_{me} \varepsilon_{me} \hat{B}_{me}^{\dagger} \hat{B}_{me} + \sum_{me,nf}^{m \neq n} J_{me,nf} \hat{B}_{me}^{\dagger} \hat{B}_{nf} \tag{1}$$

where me is the e electronic transition on the mth chromophore. $\hat{B}_{me}^{\dagger}$ is the creation Pauli operator that promotes the chromophore m into the excited state e, and $\hat{B}_{me}$ is the corresponding annihilation operator. The excited-state energy ε_{me} can be calculated as the summation of the excited energy of the isolated chromophore ($\varepsilon_{me,0}$) and the inter- and intramolecular electrostatic corrections (i.e., Coulomb interactions between electric transition dipoles and surrounding electrostatic fields)

$$\varepsilon_{me} = \varepsilon_{me,0} + \sum_{k}^{k \neq m} \frac{1}{4\pi\epsilon} \int\int d\mathbf{r}_m \, d\mathbf{r}_k \frac{[\rho_m^{ee}(\mathbf{r}_m) - \rho_m^{gg}(\mathbf{r}_m)]\rho_k^{gg}(\mathbf{r}_k)}{|\mathbf{r}_m - \mathbf{r}_k|} \tag{2}$$

Here k runs over all atomic or molecular sites surrounding the excited chromophore, ε is the dielectric constant, $\mathbf{r}_m$ and $\mathbf{r}_k$ are the positions coordinates, and ρ_m^{gg} and ρ_k^{gg} (ρ_m^{ee}) represent the ground (excited) state charge density.

Meanwhile, the resonant coupling between singly excited states me and nf is given by $\langle 0|\hat{B}_{me}\hat{H}\hat{B}_{nf}^{\dagger}|0\rangle = J_{me,nf}$ where $|0\rangle$ is the ground state.

dx.doi.org/10.1021/jz5002264 | J. Phys. Chem. Lett. 2014, 5, 1341−1346

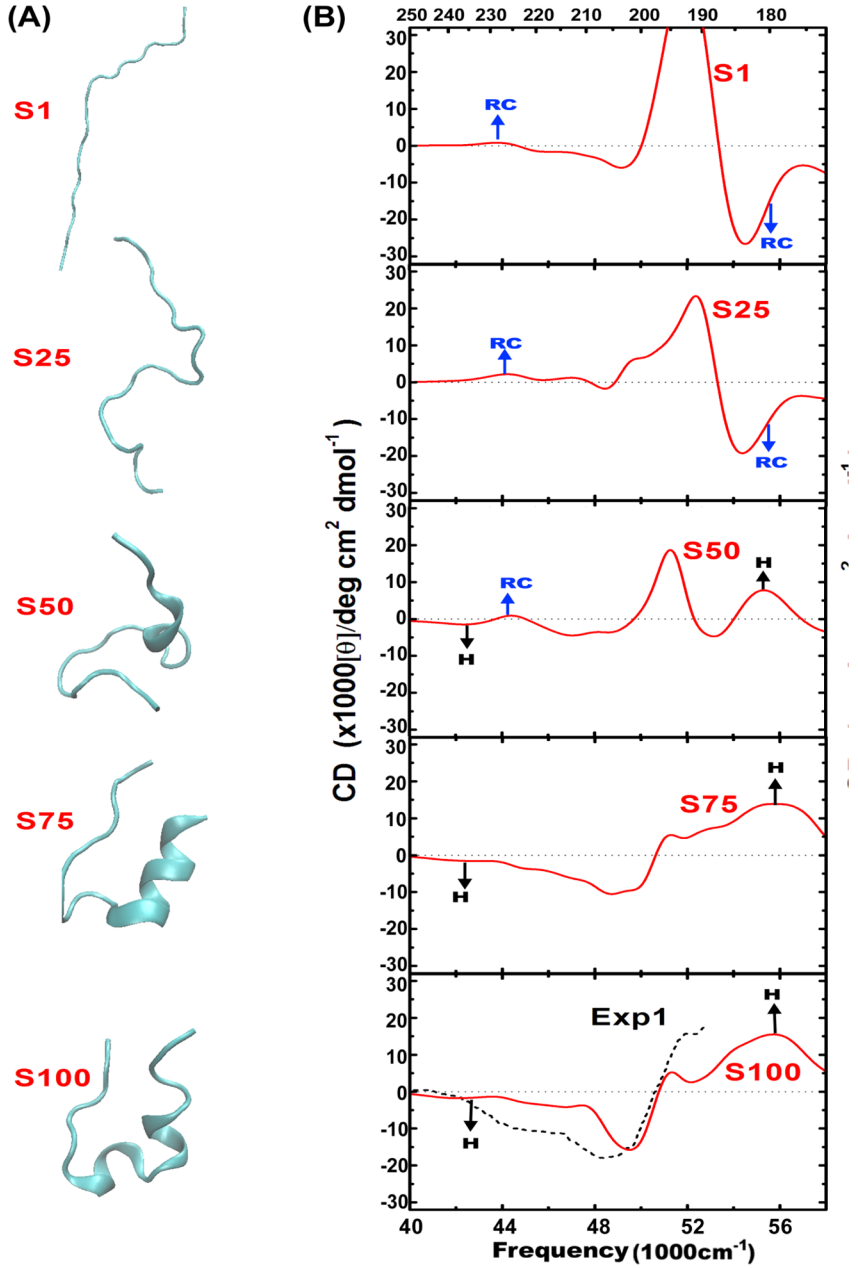


Figure 2. Structure (A) and CD spectra (B) of five states (from top to bottom:S1, S25, S50, S75, and S100) along the Trp-cage folding process. Spectra are averaged over 100 MD snapshots for each state. We labeled CD signals of the random coil and Helix as RC and H, respectively. Black dotted CD curve is from experiment Exp1[23] for the folded Trp-cage protein.

$$J_{me,nf} = \frac{1}{4\pi\varepsilon}\int\int \mathrm{d}\mathbf{r}_m\,\mathrm{d}\mathbf{r}_n \frac{\rho_{me}^{eg}(\mathbf{r}_m)\rho_{nf}^{gf}(\mathbf{r}_n)}{|\mathbf{r}_m - \mathbf{r}_n|} \quad (3)$$

where $\rho_{me}^{eg}(\mathbf{r}_m)$ and $\rho_{nf}^{gf}(\mathbf{r}_n)$ are the transition charge densities.

Here ε_{me} and $J_{me,nf}$ are calculated using our exciton Hamiltonian with electrostatic fluctuations (EHEF) algorithm[21] at a high ab initio level. EHEF also enables us to combine the QM and MM outputs to construct the effective exciton Hamiltonian of any MD snapshots (details in the Supporting Information). The Hamiltonian is applied to the response function framework based on which UV spectra are computed using the SPECTRON code.[22] 2DUV calculations were performed for the nonchiral (xxxx, xyyx, and xyxy) and chirality-induced (xxxy) pulse polarization configurations (details in the Supporting Information). Chiral 2D signals record interferences among transitions at different parts of the whole protein and thus provide richer spectral features compared with their nonchiral counterparts. The signals are displayed on a nonlinear scale that interpolates between logarithmic for small values and linear for large values, thus revealing both the strong and weak features (details in the Supporting Information).

The structures of five states, S1 (initial unfolded peptide), S25, S50, S75, and S100 (final folded peptide), are depicted in Figure 2A. Computed CD signals of S1, S25, S50, S75, and S100 (averaged over 100 MD snapshots for each state) are displayed in Figure 2B. Simulations of S100 with the final folded structure are in a good agreement with the experimental CD[23] of the folded Trp-cage peptide. CD is the standard 1D spectroscopic technique widely used for identifying protein secondary structures. CD signals reflect the variation of

dx.doi.org/10.1021/jz5002264 | J. Phys. Chem. Lett. 2014, 5, 1341−1346

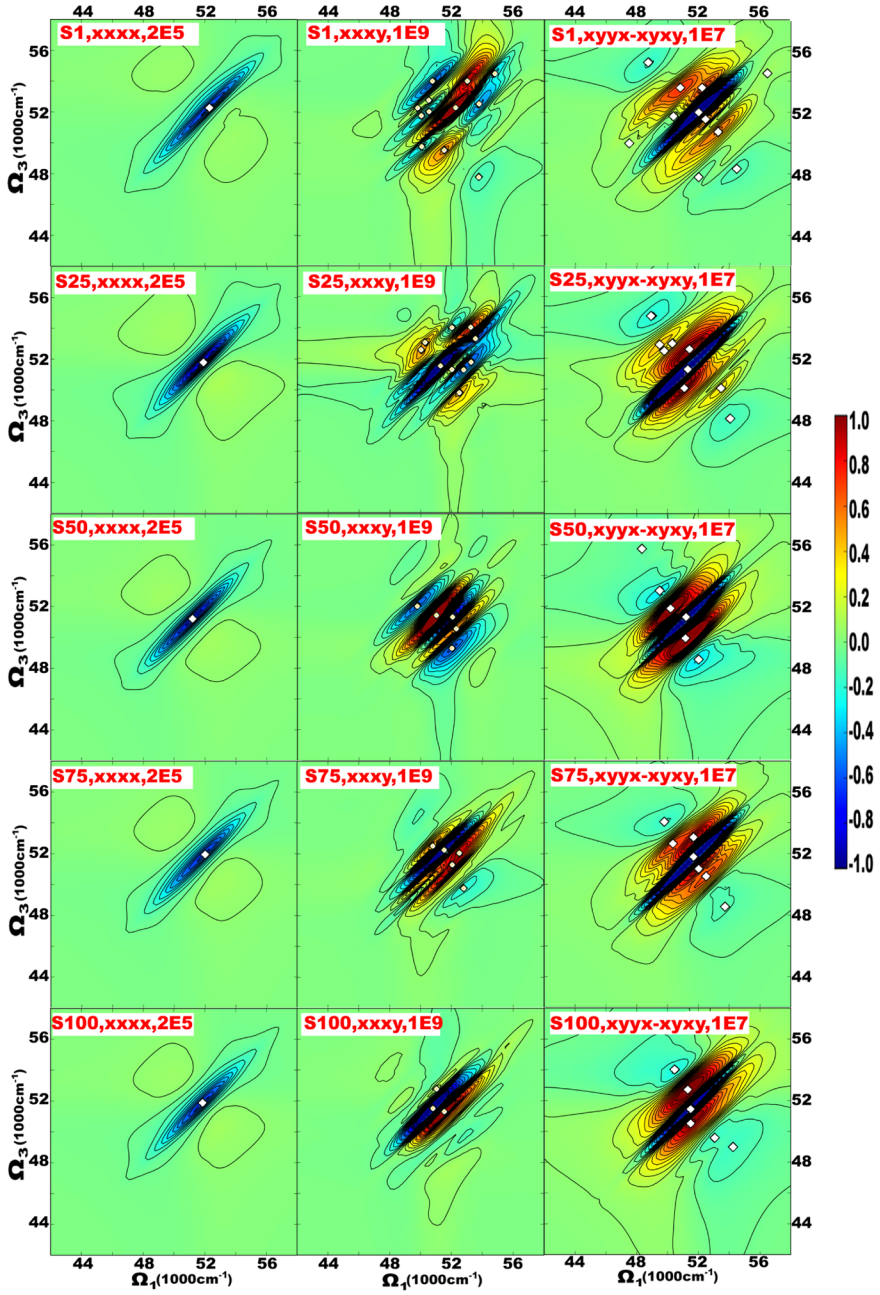


Figure 3. From left to right: 2DUV xxxx, xxxy, and xyyx-xyxy spectra of five states (from top to bottom:S1, S25, S50, S75, S100) along the Trp-cage folding process. Spectra are averaged over 100 MD snapshots for each state. The scale bar is plotted at the right top edge, and signal peaks are marked by white square dots.

secondary structural elements from state S1 to S100. The negative feature at ~56 000 cm^{-1}(~180 nm) and positive signals at ~43 000 cm^{-1}(~230 nm) marked 'RC' are typical of a random coil. These are seen in S1 and reversed in S75 and S100 following the decrease in random coil shown in Figure 1D. The helix structures increase from S1 to S100, so CDs from 53 000 to 58 000 cm^{-1} (~190 to 170 nm) marked 'H' change from negative to positive. However, because of the large inhomogeneous broadening due to geometric fluctuations, CD signatures of different structural elements are not well-resolved.

The 2D photoecho signal is strongly affected by the couplings between electronic transitions and structural variations. 2DUV xxxx (nonchiral) and xxxy (chiral) spectra of our chosen five states are displayed in the right and middle columns of Figure 3, respectively. The xxxx spectra of the five states are very similar, dominated by the negative diagonal ~52 000 cm^{-1} peak accompanied by two positive side bands. In contrast, the xxxy signals vary significantly as we move from S1 to S100. For instance, the unfolded states S1 and S25 have negative diagonal peaks at from 48 000 to 56 000 cm^{-1}, which are typical for random-coil and strand-structural elements.[24] A helical structure normally produces positive diagonal signals in that region.[24] Therefore, moving from S50 to S75 and S100, the increased helical structure reduces the negative signals and induces additional positive peaks at the diagonal part.

The xxxy chiral signals also reflect the tertiary structure. As expected from the decrease in conformational entropy in the folded structure, the xxxy spectral pattern becomes more

dx.doi.org/10.1021/jz5002264 | J. Phys. Chem. Lett. 2014, 5, 1341−1346

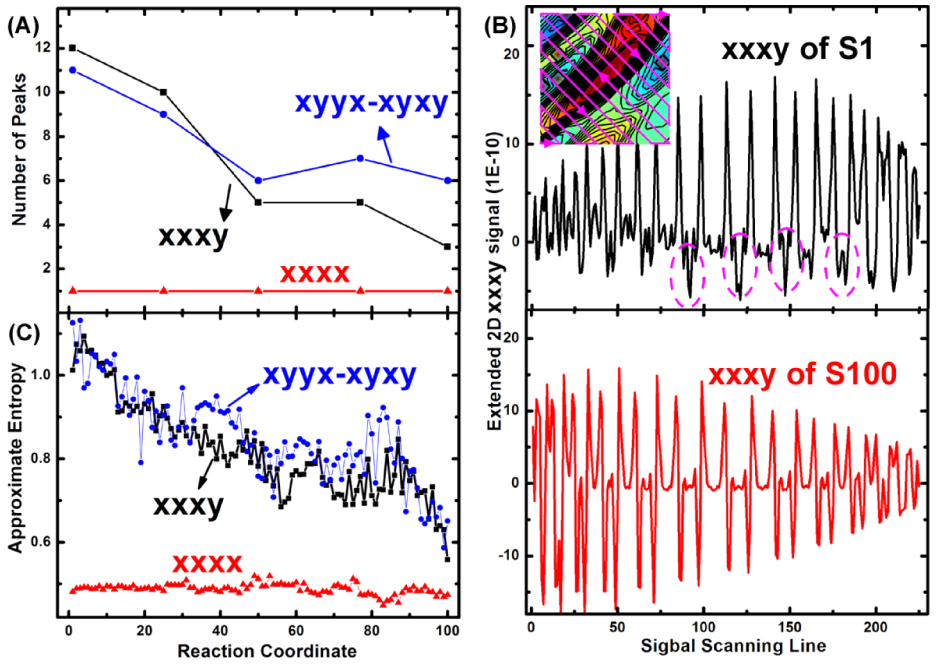


Figure 4. Evolution of the number of 2DUV peaks (A) and ApEn (B) during the Trp-cage folding process. Spectra are averaged over 100 MD snapshots for each state. (C) 2DUV xxxy signal evolution curves of states S1 and S100 along the scanning line given in the inset of panel C. Purple circles highlight the multiple-peak patterns. Spectra are averaged over 100 MD snapshots for each state.

compact and simple, lowering the signal complexity. The number of peaks (marked with white dots in Figure 2) in the xxxx and xxxy spectra of the five states are plotted in Figure 4A. The number of xxxx peaks remains fixed during the folding. The xxxy peak number decreases from S1 to S50 and remains flat from S50 to S100, implying a trend similar to the variation of R_g and the inverse of packing density. Our previous study[17] showed that the approximate entropy (ApEn)[25] provides a good measure for the complexity of 2D signals. In the inset of Figure 4B, we display a scanning line perpendicular to the diagonal of the 2D contour map, starting from the bottom left (lower energy) to the upper right (higher energy) corner. The projections of the xxxy signals of state S1 and S100 along this line are depicted in Figure 4B, showing that S1 signals have richer structures (such as more peaks) than S100. Figure 4C depicts the variation of ApEn values of xxxy spectra during the folding process. The xxxy ApEn decreases considerably as we move from S1 to S50 and does not vary from S50 to S100, consistent with the evolution of R_g and the inverse of packing density (conformational entropy) shown in Figure 1C,D. Spectra were obtained by averaging over 100 MD snapshots for each state point. As 2DUV signals are becoming feasible,[12−15] protein folding can be measured by the ApEn value of 2DUV chiral signals.

Chiral signals are harder to measure due to their very weak intensities. The technique of difference spectroscopy between two nonchiral spectra with different polarizations can cancel the single exciton contributions, while the correlations of transitions are retained and better resolved. The computed 2DUV xyyx-xyxy difference spectra of our five states are displayed in the right column of Figure 3. The signal complexity is reduced as we move from S1 to S100. This may also be seen from the number of 2DUV xyyx-xyxy spectral peaks (marked with white dots in Figure 3), and the ApEn values are shown in Figure 4A,C. The change of the complexities of difference spectra thus also provides a quantitative measure of the decrease in protein conformational entropy during the folding process. To check the convergence of the signal evolutions, we have also computed 2DUV xxxy and xyyx-xyxy signals directly on every single MD snapshot and found nearly the same evolution behaviors for their ApEn values (see Figure S2 in the Supporting Information) as those from averaged inhomogeneous spectra.

In summary, we have employed a QM/MM protocol to simulate the 2DUV spectra of the folding of a peptide. We demonstrated that 2DUV signals are sensitive to the change of peptide secondary and tertiary structure and especially useful in probing the global structural changes in comparing with IR and Raman tools. The complexity of 2DUV spectra of peptide backbone as measured by their ApEn is a good marker for the conformational entropy and provides a quantitative index of folding status with the same accuracy as the calculated RMSD and R_g values. The RMSD and R_g can only be extracted from the known atomic structures at every time point, which require huge computation resources for most proteins and are not accessible to any existing experimental techniques. 2DUV, in contrast, can offer a fast experimental measurement and theoretical verification of the protein folding state and may thus provide new insights into the protein evolution mechanism and function.

■ ASSOCIATED CONTENT

Supporting Information

MD simulations. QM/MM simulations of electronic transitions producing UV signals. Calculation of ApEn. Te evolution of the ApEn value of 2DUV spectra based on every single MD snapshot during the Trp-cage folding process. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*J.J.: E-mail: jiangj1@uci.edu.

*S.M.: E-mail: smukamel@uci.edu.
*J.W.: E-mail: jin.wang.1@stonybrook.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Institutes of Health (Grant GM059230 and GM091364), the National Science Foundation (Grant CHE-1058791), and National Natural Science Foundation of China (Grant 91221104).

■ REFERENCES

(1) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science* **1995**, *267*, 1619−1620.
(2) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884−890.
(3) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(4) Bowman, G.; Volez, V.; Pande, V. S. Taming the Complexity of Protein Folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.
(5) Wasmer, C.; Lange, A.; van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. Amyloid Fibrils of the het-s(218−289) Prion Form A Beta Solenoid with A Triangular Hydrophobic Core. *Science* **2008**, *319*, 1523−1526.
(6) Kelly, S. M.; Jess, T. J.; Price, N. C. How to Study Proteins by Circular Dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, *1751*, 119−139.
(7) Ostroumov, E. E.; Mulvaney, R. M.; Cogdell, R. J.; Scholes, G. D. Broadband 2d Electronic Spectroscopy Reveals A Carotenoid Dark State in Purple Bacteria. *Science* **2013**, *340*, 52.
(8) Mukamel, S.; Abramavicius, D.; Yang, L.; Zhuang, W.; Schweigert, I. V.; Voronine, D. Coherent Multidimensional Optical Probes for Electron Correlations and Exciton Dynamics: From NMR to x-rays. *Acc. Chem. Res.* **2009**, *42*, 553−562.
(9) Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A. Transient 2D IR Spectroscopy of Ubiquitin Unfolding Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14237−14242.
(10) Lai, Z.; Preketes, N.; Jiang, J.; Mukamel, S.; Wang, J. Monitoring the Folding of Trp-Cage Peptide by Two-Dimensional Infrared (2dir) Spectroscopy. *J. Phys. Chem. Lett.* **2013**, *4*, 1913−1917.
(11) Ren, H.; Lai, Z.; Biggs, J.; Wang, J.; Mukamel, S. Two-Dimensional Stimulated Resonance Raman Spectroscopy Study of the Trp-Cage Peptide Folding. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19457−19464.
(12) West, B. A.; Womick, J. M.; Moran, A. M. Probing Ultrafast Dynamics in Adenine with Mid-Uv Four-Wave Mixing Spectroscopies. *J. Phys. Chem. A* **2011**, *115*, 8630−8637.
(13) Tseng, C.; Matsika, S.; Weinacht, T. Two-Dimensional Ultrafast Fourier Transform Spectroscopy in the Deep Ultraviolet. *Opt. Express* **2009**, *17*, 18788−18793.
(14) Nuernberger, P.; Selle, R.; Langhojer, F.; Dimler, F.; Fechner, S.; Gerber, G.; Brixner, T. Polarization-Shaped Femtosecond Laser Pulses in the Ultraviolet. *J. Opt. A: Pure Appl. Opt.* **2009**, *11*, 085202.
(15) Consani, C.; Aubock, G.; van Mourik, F.; Chergui, M. Ultrafast Tryptophan-To-Heme Electron Transfer in Myoglobins Revealed by UV 2d Spectroscopy. *Science* **2013**, *339*, 1586.
(16) Jiang, J.; Mukamel, S. Two Dimensional Ultraviolet (2duv) Spectroscopic Tools for Identifying Fibrillation Propensity of Protein Residue Sequences. *Agnew. Chem., Int. Ed.* **2010**, *49*, 9666−9669.
(17) Jiang, J.; Golchert, K. J.; Kingsley, C. N.; Brubaker, W. D.; Martin, R. W.; Mukamel, S. Exploring the Aggregation Propensity of Γ S-Crystallin Protein Variants Using Two Dimensional Spectroscopic Tools. *J. Phys. Chem. B.* **2013**, *117*, 14294.
(18) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Design A 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(19) Liao, H.; Yeh, W.; Chiang, D.; Jernigan, R.; Lustig, B. Protein Sequence Entropy Is Closely Related to Packing Density and Hydrophobicity. *Protein Eng., Des. Sel.* **2005**, *18*, 59−64.
(20) Hirst, J. D. Improving Protein Circular Dichroism Calculations in the Far Ultraviolet Through Reparametrizing the Amide Chromophore. *J. Chem. Phys.* **1998**, *109*, 782−788.
(21) Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. Ultraviolet Spectroscopy of Protein Backbone Transitions in Aqueous Solution: Combined QM and MM Simulations. *J. Phys. Chem. B* **2010**, *114*, 8270−8277.
(22) Abramavicius, D.; Palmieri, B.; Voronine, D. V.; Šanda, F.; Mukamel, S. Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle, Vs. Supermolecule Perspectives. *Chem. Rev.* **2009**, *109*, 2350−2408.
(23) Adams, C. M.; Kjeldsen, F.; Patriksson, A.; van der Spoel, D.; Graslund, A.; Papadopoulos, E.; Zubarev, R. A. Probing Solution- and Gas-Phase Structures of Trp-Cage Cations by Chiral Substitution and Spectroscopic Techniques. *J. Mass Spectrom.* **2006**, *253*, 263−273.
(24) Abramavicius, D.; Jiang, J.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. Simulation Study of Chiral Two Dimensional Ultraviolet (2DUV) Spectroscopy of the Protein Backbone. *J. Am. Chem. Soc.* **2010**, *132*, 7769−7775.
(25) Pincus, S. M. Approximate Entropy as A Measure of System Complexity. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2297−2301.