

Disentangling Peptide Configurations via Two-Dimensional Electronic Spectroscopy: Ab Initio Simulations Beyond the Frenkel Exciton Hamiltonian

Artur Nenov,[†] Ivan Rivalta,^{‡,†} Giulio Cerullo,[§] Shaul Mukamel,^{*,||} and Marco Garavelli^{*,†,‡}

[†]Dipartimento di Chimica “G. Ciamician”, Università di Bologna, V. F. Selmi 2, 40126 Bologna, Italy

[‡]Laboratoire de Chimie, Ecole Normale Supérieure de Lyon, 46, allée d’Italie 69364 Lyon, France

[§]IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

^{||}Department of Chemistry, University of California, Irvine, California 92697-2025, United States

The functionality and aggregation state of a protein can be inferred from its secondary structure and the positions of key residues. Two-dimensional (2D) nuclear magnetic resonance (NMR) techniques are capable of resolving detailed structural information of proteins by addressing spin transitions. NMR is, however, a rather complex and time-consuming procedure involving several phases, such as data collection, resonance assignment, restraint generation and final structure calculation and refinement. Alternatively, rapidly screening different point mutations or solution conditions for a particular protein will be highly desirable. Recently, multidimensional spectroscopy has been transferred to the optical domain, starting from the infrared (IR) and evolving to the ultraviolet–visible (UV–vis) and the X-ray spectral regimes.^{1–3} While 2DIR is becoming a standard tool for probing vibrational modes of backbone amide groups,^{1,4–7} 2DIR experiments require intensive isotope labeling to resolve spectral congestion. Recent advances in ultrafast optics have led to the generation of broadband pulses in the UV, enabling to address electronic transitions in proteins.^{8–12} The $n\pi^*$ and $\pi\pi^*$ electronic transitions of the peptide backbone fall in the deep-UV (DUV) region and can be used as global probes of the protein secondary structure.¹³ Instead, the aromatic side chains of the amino acids phenylalanine (Phe), tyrosine (Tyr) and

tryptophane (Trp) absorb in the near-UV (NUV). As these amino acids are relatively rare in proteins (approximately 4% Phe, 3% Tyr, 1.5% Trp occurrence probability), they can be used as native probes for secondary structures.^{14,15} Therefore 2DUV holds the promise to provide fast structure determination without any isotopic labeling.

2D measurements of systems consisting of many coupled chromophores in solution have been traditionally described using the parametrized Frenkel exciton model,¹⁶ with applications in the IR (phospholipids),¹⁷ VIS (photosynthetic complexes)¹⁶ and UV (amyloid fibrils).^{14,15} Thereby, exciton parameters are obtained from ab initio calculations of the isolated chromophores in gas phase and can be used to model arbitrary size aggregates, while environmental fluctuations are accounted for on a grid using effective charges fitted from the electrostatic potential.^{14,15,18} This methodology offers a computationally affordable strategy for modeling extended systems but invokes several approximations. For instance, exciton Hamiltonian (EH) parameters based on single

Received: February 3, 2014

Accepted: February 7, 2014

Published: February 7, 2014

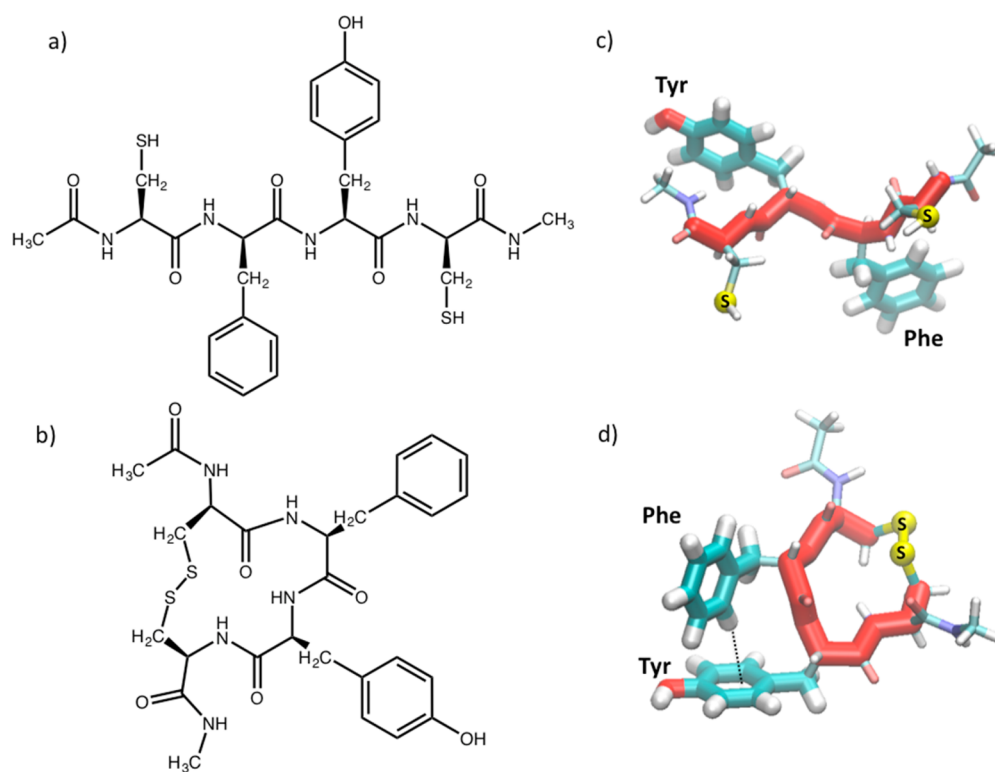


Figure 1. Schematic representations of the open (a) and closed (b) CFYC peptide configurations and selected conformations of the unstacked (c) and T-stacked (d) peptide.

chromophores cannot include charge transfer (CT) states and doubly excited states involving both chromophores. The basic assumption that the electronic (or vibrational) structure of each monomer is not significantly affected in the “interacting” dimer, is expected to break down more often in electronic than in vibrational spectra. Furthermore, only the lowest valence states of the chromophores are generally parametrized in exciton models, missing contributions from excitations to higher lying local states. Finally, inclusion of environmental electrostatic fluctuations (EF) in the EH models (EHEF) by parametrization with “map function” methods¹⁴ is less accurate than explicit calculations of electrostatics with QM/MM methods.¹⁹

Recently, we have demonstrated that using wave function based QM/MM method the aforementioned limitations of EHEF can be overcome for dimeric chromophoric systems.²⁰ Similar to the EHEF method, this approach relies on the accuracy of classical molecular dynamics (MD) simulations for the configurational space sampling of the peptide in solution. For selected geometries we refine the results with QM/MM calculations in order to obtain accurate excitation energies and transition dipole moments of the singly and doubly excited manifolds in solution, as required to simulate the 2DUV spectra. Multiconfigurational wave function techniques such as the complete active space self-consistent field (CASSCF) provide the required capabilities.²¹ The lack of dynamic correlations can be corrected using perturbation techniques (i.e., CASPT2).²² Finally, the 2D electronic spectra are calculated using the sum-over-states (SOS) approach for determining the nonlinear response of the system.²³ Our SOS//QM/MM protocol is capable of generating 2DUV spectra of aromatic residues with higher accuracy than any previously proposed EH approach.²⁰

In this Letter we use both the EHEF and the SOS//QM/MM approach to calculate 2DUV spectra for two different configurations of the small prototype polypeptide cysteine–phenylalanine–tyrosine–cysteine (CFYC, Figure 1a–d), an open configuration with unstacked aromatic side chains and a closed cyclic (through a disulfide-bridge) configuration with T-stacked arrangement (i.e., the preferred aromatic–aromatic interaction in proteins).²⁴ These two configurations represent the unfolded and folded structures of the CFYC peptide in solution. Distinguishing the two configurations by means of 2DUV will pave the way for conducting time-resolved folding/unfolding experiments. We demonstrate that the spatial proximity of the two chromophores clearly manifests itself in the SOS//QM/MM 2DUV spectra through several signals, most of which are missed in the EH approach. We further show that several features in our high-level calculations are missed by the simple Frenkel exciton model.

The energetically lowest geometries were selected from two 40 ns classical MD trajectories of the open and cyclic CFYC oligopeptides in solution (Figure 1; simulation parameters provided in the Supporting Information, SI). Excited state calculations were then performed with MOLCAS 7.7²⁵ at CASSCF(14,13)/SS-CASPT2 level (including all valence π -electrons and π -orbitals of both chromophores in the active space) and the ANO-L(3s2p1d/2s1p) basis set²⁶ (see benchmarks in the SI). Rephasing 2DUV spectra, detected in a heterodyne manner in direction $\mathbf{k}_4 = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$, were generated with Spectron¹⁶ for the nonchiral xxx, xyxy, and xyxx polarization configurations. The 2D signals, which depend parametrically on the delay times t_1 , t_2 , and t_3 , were displayed in the frequency domain by 2D Fourier transformation along t_1 and t_3 , while t_2 was set to zero, thereby suppressing excited state dynamics. Spectra are plotted on a logarithmic scale to visualize

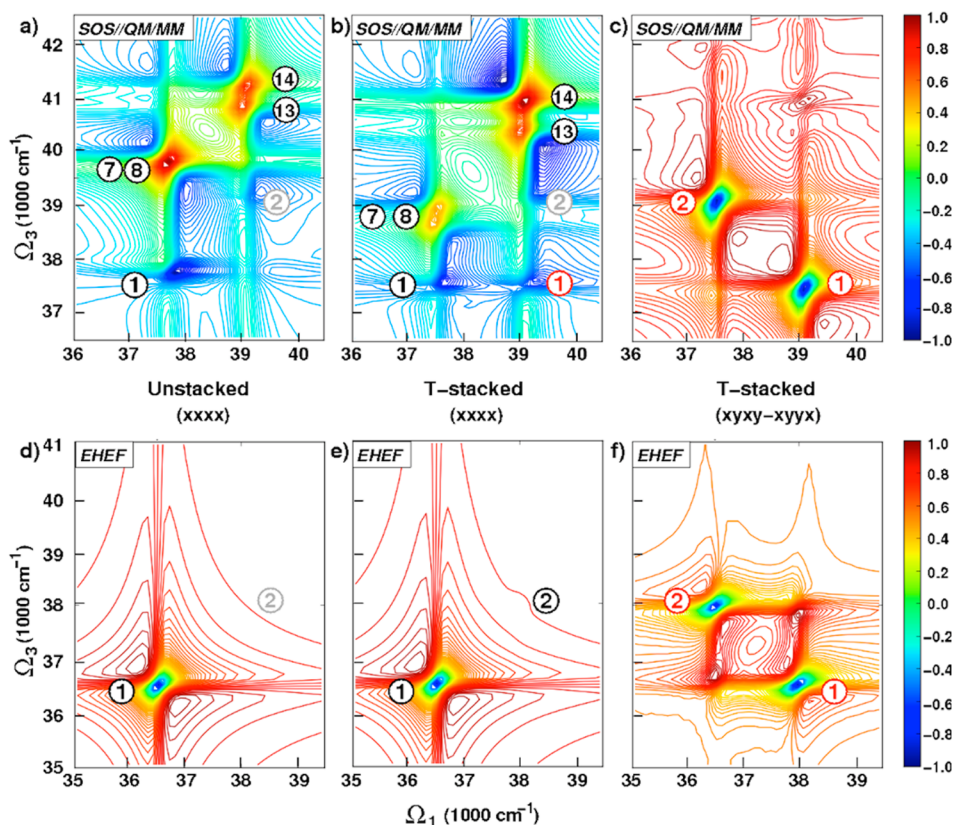


Figure 2. 2DUV spectra obtained with the SOS//QM/MM protocol (upper panel) and with the EHEF method (lower panel) with a one-color setup: (a,d) open CFYC (xxxx); (b,e) closed CFYC (xxxx); (c,f) closed CFYC (xyxy-xyyx). Peaks numeration refers to transitions shown in Figures S1 and S2.

both strong and weak absorptions. In addition, the EHEF method was applied to the MD snapshots to generate conventional spectra.

The first excited states of Phe and Tyr with ${}^1B_{2u}$ symmetry are selected as targets for the incoming pulses (where the D_{6h} symmetry labeling of benzene is used also for phenol, see SI). These states have lower oscillator strengths in comparison to higher lying states. However they are spectrally well separated from other electronic transitions, while higher lying bright states as ${}^1B_{1u}$ overlap with the intense absorption of the backbone peptide bonds. Moreover, ultrashort broadband pulses in the NUV region are more easily within reach. Phe has a molar absorption coefficient much smaller than Tyr (195 vs 1405 $\text{M}^{-1}\text{cm}^{-1}$, respectively).²⁷ Setting the incoming pulses in resonance with the ${}^1B_{2u}$ transition of Tyr will cover any Phe signal. It is therefore essential to use narrowband k_1 and k_2 pulses centered near the frequency of the ${}^1B_{2u}$ transition of Phe to enhance it. It should be noted, though, that the overall signal intensity decreases (see Figures S4–S6 in the SI). On the other hand, broadband k_3 and k_4 pulses are required to cover a meaningful fraction of the 2D spectrum. Therefore, we used k_1 and k_2 pulses centered on Phe absorption (ca. 38500 cm^{-1}) with a full width at half-maximum (FWHM) bandwidth of 733 cm^{-1} (corresponding to a 20 fs pulse) and k_3 and k_4 pulses with a FWHM bandwidth of 2932 cm^{-1} (~ 5 fs pulse).

Figure 2 shows a comparison between the ab initio 2D spectra obtained with the explicit SOS//QM/MM approach (Figure 2a–c) and the EHEF method (Figure 2d–f) in the open and closed tetrapeptide, using all four pulses in resonance with the Phe absorption (one-color setup). The SOS//QM/

MM spectrum of the open peptide with xxxx polarization (Figure 2a) shows the diagonal (negative) bleach signal of the ${}^1A_{1g}$ (ground state) \rightarrow ${}^1B_{2u}$ transition in Tyr (peak 1) at ~ 37500 cm^{-1} . Moreover, the closely lying diagonal bleach signal of the ${}^1A_{1g} \rightarrow {}^1B_{2u}$ Phe transition (peak 2) is much weaker and not clearly visible. The spectrum is dominated by off-diagonal contributions at $\Omega_3 \approx 40000$ cm^{-1} (peaks 7,8) and $\Omega_3 \approx 41500$ cm^{-1} (peaks 13,14) arising from intense dipole allowed excitations out of the ${}^1B_{2u}$ states of Tyr and Phe into a pair of local superexcited (i.e., above the first ionization limit) valence states with 2^1E_{2g} symmetry. The 2^1E_{2g} pair of states represents the ionic counterparts of the covalent pair of states 1^1E_{2g} below the ionization potential. Our calculations are in agreement with experimental spectroscopy data postulating the existence of short-lived superexcited states in phenol and in benzene.^{28,29}

Due to the aromatic side chains spatial proximity and chromophore–chromophore interaction in the T-stacked configuration the ${}^1A_{1g} \rightarrow {}^1B_{2u}$ transitions (Figure 2b) are coupled and an off-diagonal peak at $\sim 39000/\sim 37500$ cm^{-1} is revealed (red peak 1, Figure 2b). However, the dominant (positive) signals of the open tetrapeptide spectrum (peaks 7–8,13–14) are preserved in the T-stacked configuration (Figure 2b) due to their local nature, partially covering both diagonal and off-diagonal negative signals. Remarkably, the aromatic stacking causes a red-shift and increases the splitting of the Tyr 2^1E_{2g} states by 300 cm^{-1} . More elaborate polarization schemes like xyxy or xyyx reduce the broadening of the signals significantly, thereby increasing the resolution (Figure S7 in the SI). Furthermore, the interchromophore couplings can be

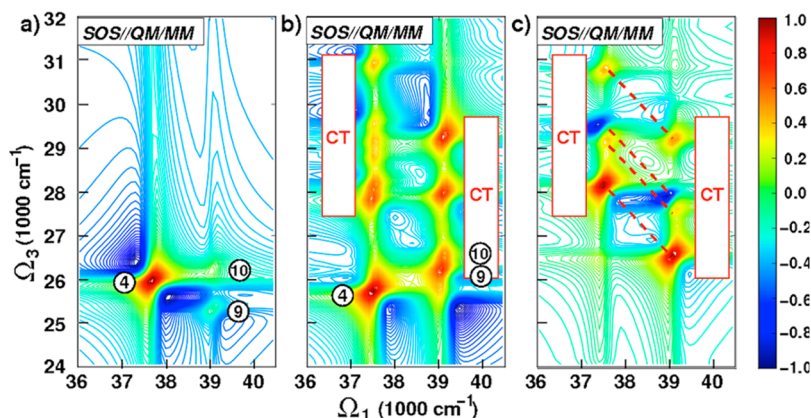


Figure 3. 2DUV spectra obtained with the SOS//QM/MM protocol with a two-color setup: (a) open CFYC (xxxx); (b) closed CFYC (xxxx); (c) closed CFYC (xyxy - yyyy).

selectively enhanced in 2D spectra obtained by recording the difference between spectra with two polarizations (e.g., xyxy - yyyy), see Figure 2c. Diagonal bleach and local absorption features are removed, leaving clearer off-diagonal bleach signals as well as signals coming from excitations to bright common excited states. In the unstacked configuration, the 2D difference spectra are signal-free (data not reported). Instead, in the T-stacked configuration symmetric off-diagonal peaks appear clearly in the SOS//QM/MM spectrum (red peaks 1 and 2 in Figure 2c).

The differences between SOS//QM/MM and EHEF spectra are clearly deducible from the comparison shown in Figure 2. Excitations to high lying local states (peaks 7–8, 13–14 in Figure 2a-b) are missed in the EHEF spectra of both open and closed tetrapeptides (Figure 2d,e). Therefore, the $^1B_{2u}$ bleaching of Tyr is the only signal visible in the EHEF spectra (Figure 2d) of the unstacked tetrapeptide, while only a weak enhancement of the Phe bleaching signal is emerging in the EHEF spectrum upon chromophore interaction. Notably, the interchromophore couplings observed in the SOS//QM/MM difference spectrum (Figure 2c) are also visible, but with lower intensities, in the EHEF spectrum of the T-stacked configuration (Figure 2f).

It is clear from the comparison of the unstacked and T-stacked SOS//QM/MM spectra (Figure 2a versus 2b) that considering line broadening due to thermal fluctuations it will be difficult to discriminate between the two configurations. In fact, the presence of intense off-diagonal local absorptions and contributions from ionization and fragmentation to the background are likely to cover the weak couplings. We therefore propose to shift the probing window to the red (two-color setup) between 24000 cm^{-1} and 32000 cm^{-1} and, hence, below the ionization limit (k_3 and k_4 pulses centered at 28000 cm^{-1}), where only the signals originating from the $^1B_{2u} \rightarrow ^1E_{2g}$ transitions are encountered in the unstacked tetrapeptide (peak 4 for Tyr and peaks 9 and 10 for Phe; Figure 3a,b). A number of CT states exhibit a red shift of around 15000 cm^{-1} upon stacking, as shown in Figure 3b (“CT” region). Notably, their oscillator strength increases by several orders of magnitude with respect to the unstacked configuration. Although the overall signal intensity obtained with the two-color setup is weaker, the lack of intense valence transitions and its background-free nature should facilitate the detection of CT states already with the xxxx polarization. Again, a clearer signature of the CT states can be obtained from the

difference spectrum only in the T-stacked configuration (Figure 3c). The current EHEF model is not adequate in this region for two reasons: first, CT states are completely neglected; second, higher valence states have not yet been parametrized for Frenkel excitons.

In conclusion, we present the 2DUV spectra obtained from state-of-the-art ab initio simulations, combining QM/MM methods based on wave function techniques and nonlinear response theory. By applying the SOS//QM/MM protocol, we have demonstrated that NUV transitions in aromatic residues can be utilized to discriminate between different polypeptide configurations. The 2DUV electronic spectra are extremely rich. Some information is completely neglected in conventional exciton Hamiltonian models. The proposed SOS//QM/MM protocol requires further extension to include spectral fluctuations and line broadening and it is limited to relatively small aggregates. However, our results can be straightforwardly used to improve current exciton models by providing standard parameters for local excitations and to construct new exciton models that include CT states. Moreover, the SOS//QM/MM method can provide accurate nonlinear response for complex systems involving multiple monomeric (or dimeric) isolated species. Finally, the present study should serve as a guide for conducting 2D experiments on polypeptides. A combination of pulses with different bandwidth is proposed for disentangling the $^1B_{2u}$ bands. Furthermore, experimental setups to reduce undesired signal broadening and background contributions and to isolate and enhance specific electronic couplings are discussed, including a two-color scheme. These results indicate that 2D electronic spectroscopy holds the promise to become a novel diagnostic tool to track the folding state and dynamics of unlabeled peptides, and be complementary to well-established multidimensional NMR and 2DIR techniques.

ASSOCIATED CONTENT

Supporting Information

Computational details, reference calculations for benzene and phenol in the gas phase, additional spectra, level schemes, CASSCF(14,13) orbitals, and Cartesian coordinates. This material is available free of charge via Internet.

AUTHOR INFORMATION

Corresponding Authors

*(S.M.) E-mail: smukamel@uci.edu. Mailing address: Department of Chemistry, University of California, Irvine, California 92697-2025, United States.

*(M.G.) E-mail: marco.garavelli@unibo.it. Mailing address: Dipartimento di Chimica "G. Ciamician", Università di Bologna, V. F. Selmi 2, 40126 Bologna, Italy. Phone: +39 051 20 9 9476.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

G.C. and M.G. acknowledge support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). S.M. gratefully acknowledges the support of the National Institute of Health Grant No. GM-59230 and the National Science Foundation through Grant No. CHE-1058791.

REFERENCES

(1) Mukamel, S. Multidimensional Femtosecond Correlation Spectroscopies of Electronic and Vibrational Excitations. *Annu. Rev. Phys. Chem.* **2000**, *51*, 691–729.

(2) Hamm, P.; Zanni, M., *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: Cambridge, U.K.; 2011.

(3) Mukamel, S.; Healion, D.; Zhang, Y.; Biggs, J. D. Multidimensional Attosecond Resonant X-ray Spectroscopy of Molecules: Lessons from the Optical Regime. *Annu. Rev. Phys. Chem.* **2013**, *64*, 101–127.

(4) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. Watching Hydrogen-Bond Dynamics in a Beta-Turn by Transient Two-Dimensional Infrared Spectroscopy. *Nature* **2006**, *444*, 469–472.

(5) Zhuang, W.; Abramavicius, D.; Hayashi, T.; Mukamel, S. Simulation Protocols for Coherent Femtosecond Vibrational Spectra of Peptides. *J. Phys. Chem. B* **2006**, *110*, 3362–3374.

(6) 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*, 6123–6138.

(7) Mukherjee, P.; Kass, I.; Arkin, I.; Zanni, M. T. Picosecond Dynamics of a Membrane Protein Revealed by 2D IR. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3528–3533.

(8) Ajdarzadeh Oskouei, A.; Bram, O.; Cannizzo, A.; van Mourik, F.; Tortschanoff, A.; Chergui, M. Photon Echo Peak Shift Experiments in the UV: *p*-Terphenyl in Different Solvents. *J. Mol. Liq.* **2008**, *141*, 118–123.

(9) 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.

(10) West, B. A.; Moran, A. M. Two-Dimensional Electronic Spectroscopy in the Ultraviolet Wavelength Range. *J. Phys. Chem. Lett.* **2012**, *3*, 2575–2581.

(11) Tseng, C.-h.; Matsika, S.; Weinacht, T. C. Two-Dimensional Ultrafast Fourier Transform Spectroscopy in the Deep Ultraviolet. *Opt. Express* **2009**, *17*, 18788–18793.

(12) Beutler, M.; Ghotbi, M.; Noack, F.; Brida, D.; Manzoni, C.; Cerullo, G. Generation of High-Energy Sub-20 fs Pulses Tunable in the 250–310 nm Region by Frequency Doubling of a High-Power Noncollinear Optical Parametric Amplifier. *Opt. Lett.* **2009**, *34*, 710–712.

(13) Lai, Z.; Preketes, N. K.; Mukamel, S.; Wang, J. Monitoring the Folding of Trp-Cage Peptide by Two-Dimensional Infrared (2DIR) Spectroscopy. *J. Phys. Chem. B* **2013**, *117*, 4661–9.

(14) 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.

(15) Jiang, J.; Mukamel, S. Two-Dimensional Near-Ultraviolet Spectroscopy of Aromatic Residues in Amyloid Fibrils: A First Principles Study. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2394–2400.

(16) Abramavicius, D.; Palmieri, B.; Voronine, D. V.; Sanda, F.; Mukamel, S. Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle versus Supermolecule Perspectives. *Chem. Rev.* **2009**, *109*, 2350–2408.

(17) Zhuang, W.; Hayashi, T.; Mukamel, S. Coherent Multidimensional Vibrational Spectroscopy of Biomolecules: Concepts, Simulations, and Challenges. *Angew. Chem., Int. Ed.* **2009**, *48*, 3750–3781.

(18) Jiang, J.; Mukamel, S. Probing Amyloid Fibril Growth by Two-Dimensional Near-Ultraviolet Spectroscopy. *J. Phys. Chem. B* **2011**, *115*, 6321–6328.

(19) Altoe, P.; Stenta, M.; Bottoni, A.; Garavelli, M. A Tunable QM/MM Approach to Chemical Reactivity, Structure and Physico-Chemical Properties Prediction. *Theor. Chem. Acc.* **2007**, *118*, 219–240.

(20) Rivalta, I.; Nenov, A.; Cerullo, G.; Mukamel, S.; Garavelli, M. Ab Initio Simulations of Two-Dimensional Electronic Spectra: The SOS//QM/MM Approach. *Int. J. Quantum Chem.* **2014**, *114*, 85–93.

(21) Roos, B. O. *Ab Initio Methods in Quantum Chemistry: Part II*; Wiley: Chichester, U.K., 1987.

(22) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. 2nd-Order Perturbation-Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(23) Chen, G. H.; Mukamel, S.; Beljonne, D.; Bredas, J. L. The Coupled Electronic Oscillators vs the Sum-over-States Pictures for the Optical Response of Octatetraene. *J. Chem. Phys.* **1996**, *104*, 5406–5414.

(24) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction - A Mechanism of Protein-Structure Stabilization. *Science* **1985**, *229*, 23–28.

(25) Aquilante, F.; De Vico, L.; Ferre, N.; Ghigo, G.; Malmqvist, P. A.; Neogrady, P.; Pedersen, T. B.; Pitonak, M.; Reiher, M.; Roos, B. O.; et al. Software News and Update MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224–247.

(26) Widmark, P. O.; Malmqvist, P. A.; Roos, B. O. Density-Matrix Averaged Atomic Natural Orbital (ANO) Basis-Sets for Correlated Molecular Wave-Functions 0.1. 1st Row Atoms. *Theor. Chim. Acta* **1990**, *77*, 291–306.

(27) Fasman, G. D., *CRC Hand Book of Biochemistry and Molecular Biology Proteins*; CRC Press: Boca Raton, FL, 1976.

(28) Schick, C. P.; Weber, P. M. Ultrafast Dynamics in Superexcited States of Phenol. *J. Phys. Chem. A* **2001**, *105*, 3725–3734.

(29) Nakashima, N.; Sumitani, M.; Ohmine, I.; Yoshihara, K. Nanosecond Laser Photolysis of the Benzene Monomer and Eximer. *J. Chem. Phys.* **1980**, *72*, 2226–2230.