Ab Initio Simulations of Two-Dimensional Electronic Spectra: The SOS//QM/MM Approach

Ivan Rivalta,*[a] Artur Nenov,[a] Giulio Cerullo,[b] Shaul Mukamel,[c] and Marco Garavelli*[a,d]

Two-dimensional electronic spectroscopy (2DES) is a cutting-edge technique for investigating with high temporal resolution energy transfer, structure, and dynamics in a wide range of systems in physical chemistry, energy sciences, biophysics, and biocatalysis. However, the interpretation of 2DES is challenging and requires computational modeling. This perspective provides a roadmap for the development of computational tools that could be routinely applied to simulate 2DES spectra of multichromophoric systems active in the UV region (2DUV) using state-of-the-art *ab initio* electronic structure methods within a quatum mechanics/molecular mechanics (QM/MM) scheme and the sum-over-states (SOS) approach (here called SOS//QM/MM). Multiconfigurational and multireference perturbative methods, such as the complete active space self-consistent field and second-order multireference perturbation theory (CASPT2) techniques, can be applied to reliably calculate the electronic properties of multichromophoric systems. Hybrid QM/MM method and molecular dynamics techniques can be used to assess environmental and conformational effects, respectively, that shape the 2D electronic spectra. DNA and proteins are important biological targets containing UV chromophores. We report *ab initio* simulation of 2DUV spectra of a cyclic tetrapeptide containing two interacting aromatic side chains, a model system for the study of protein structure and dynamics by means of 2DUV spectroscopy. © 2013 Wiley Periodicals, Inc.

DOI: 10.1002/qua.24511

Introduction

Two-dimensional (2D) Fourier transform (FT) nuclear magnetic resonance (NMR) spectroscopic technique was introduced in the late 70s by Ernst,[1] revolutionizing the field of structural biology. In the last four decades, 2D-NMR has been source of inspiration for the development of other multidimensional spectroscopic techniques. The 2D-FT spectroscopy has been extended to the optical frequency domain using ultrashort light pulses,[2,3] with first applications in the infrared (IR) spectral range, resonant with vibrational transitions.[4] Today, 2DIR is a mature spectroscopic technique that provided transformative insights into structure and dynamics of complex molecules, liquids, surfaces, and proteins by direct mapping of vibrational couplings.[2,5–9] Recent advances in ultrafast optical techniques[10,11] allowed for the extension of 2D spectroscopy to the visible range, targeting electronic transitions in such frequency domain. In the past decade, applications of 2D electronic spectroscopy (2DES) in the visible yielded fundamental insights into energy-transfer processes in photosynthetic systems.[12–15] Extension of 2DES to the UV domain (2DUV) is extremely attractive, as many biomolecules display strong absorption bands in the UV; however, it has been so far hampered by technical difficulties, including attainment of interferometric-phase stability and sufficient laser bandwidth. After recent progresses,[16,17] a limited number of 2DUV experiments have been reported in the last two years,[18–20] investigating electronic excitations in DNA nucleobases[18,19] and chemical reaction dynamics.[20]

By spreading the information content of the nonlinear signal on two frequency axes, 2DES (UV or visible) provides a wealth of novel information on molecular structure and dynamics with respect to the signal collected in 1D pump-probe experiments. However, the interpretation of 2D electronic spectra is challenging and theoretical methods necessary to interpret experimental spectra and disentangle the information contained in the nonlinear optical response of the sample. In particular, theoretical methods based on exciton models (EMs) have been used to interpret 2DIR spectra[5] and 2DES spectra in the visible.[5,12,21] In such context, the development of computational strategies for simulations of 2DES spectra based on accurate characterization of the electronic structure of multichromophoric systems is crucial. *Ab initio* quatum chemistry

[a] *I. Rivalta, A. Nenov and M. Garavelli*
Dipartimento di Chimica "G. Ciamician", Università di Bologna, V. F. Selm i 2, Bologna, 40126, Italy
E-mail: ivan.rivalta2@unibo.it and marco.garavelli@unibo.it
[b] *G. Cerullo*
IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Piazza Leonardo Da Vinci 32, Milano, 20133, Italy
[c] *S. Mukamel*
Department of Chemistry, University of California, Irvine, California 92697-2025
[d] *M. Garavelli*
Laboratoire de Chimie, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364, Lyon, France

Contract grant sponsor: European Research Council Advanced Grant STRATUS (ERC-2011-AdG) (G.C. and M.G.); contract grant number: 291198.
Contract grant sponsor: National Institute of Health (S.M.); contract grant number: GM-59230.
Contract grant sponsor: National Science Foundation (S.M.); contract grant number: CHE-1058791.
Contract grant sponsors: Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, US Department of Energy (S.M.).

Ivan Rivalta received his PhD in Chemistry at Università della Calabria (Italy) in 2007, under supervision of Prof. Emilia Sicilia and Prof. Nino Russo. In 2009, he visited Prof. Michele Parrinello's group at ETH Zurich, working on hybrid QM/MM molecular dynamics simulations. In 2010, he joined Prof. Victor S. Batista's group at Yale University as associate research scientist until 2012, working on computational studies of allosteric enzymes, GPCRs and natural and artificial photosynthesis. In 2012, he started working with Prof. Marco Garavelli at University of Bologna on structure and dynamics of biomolecules and biocatalytic systems by two-dimensional electronic spectroscopy. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Artur Nenov received his PhD in Chemistry at Ludwig-Maximilians-Universität München (Germany) in 2012, under supervision of Prof. Regina de Vivie-Riedle. In the same year, he joined Prof. Marco Garavelli's group at University of Bologna to work on the development of *ab initio* techniques for simulating two-dimensional electronic spectroscopy. His research interests include computational photochemistry, role of geometrical and substituent effects on structure and energetics of conical intersections, mixed quatum-classical dynamics simulations, artificial chromophore systems for application as molecular switches. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Giulio Cerullo is full Professor of Physics at Politecnico di Milano. His research activity is focused on the generation of tunable femtosecond light pulses and on their application to ultrafast spectroscopy of molecules and solids. He is the recipient of an ERC Advanced Grant, together with Prof. Marco Garavelli, on the development of two-dimensional UV spectroscopy for the study of dynamics and structure in oligonucleotides and polypeptides. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Shaul Mukamel is a Chancellor Professor of Chemistry at UC Irvine. His research focuses on the design of ultrafast multidimensional coherent optical spectroscopies in molecules ranging from infrared to X-rays. He is the author of the textbook "Principles of Nonlinear Optical Spectroscopy" and is a member of the American Academy of Arts & Sciences and Senior Fellow of FRIAS and the recent recipient of the Lippincott award, the Plyler Award, Hamburg Prize for theoretical physics and the Lamb Award for Laser Science and Quatum Optics. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Marco Garavelli is full Professor of Theoretical Chemistry at ENS-Lyon (France) and Aggregate Professor at Bologna University (Italy). His main research activity focuses on the development and application of computational tools for modeling photoinduced events in complex molecular architectures. For his activity, he received several awards, including the *PRIMO LEVI Prize* and the *Younger European Chemists* commendation award. In 2012, he has been awarded, together with Prof. Giulio Cerullo, an ERC Advanced Grant for a project on bidimensional UV spectroscopy in biomolecules. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

methods allow for quatitative description of electronic energy levels in multichromophoric systems, going beyond approximate EMs. Time-dependent density functional theory and molecular dynamics (MD) have been used to simulate 2DUV spectra of amide and aromatic side chains models, providing a crude approximation of the coherent signals in which double-excitation states and environmental effects have been neglected.[22] State-of-the-art wavefunction methods, such as complete active space self-consistent field (CASSCF) and second-order multireference perturbation theory (CASPT2) techniques, allow for determination of single- and double-excitation states with high accuracy and reasonable computational cost. Accurate simulations of 2DES spectra requires inclusion of environmental effects (solvent, protein embedding, etc.) and quatitative reproduction of the signal broadening due to thermal fluctuations of the multichromophoric system.

Thus, advanced wavefunction methods have to be used in conjunction with a hybrid quatum mechanics/molecular mechanics (QM/MM) scheme for a realistic description of the multichromophoric system environment. Signal broadening and linewidths can be estimated by sampling the configurational space with MD techniques.

In this perspective, we show how first-principle simulations of 2DUV spectra of an oligopeptide containing two interacting aromatic side chains can be performed using state-of-the-art wavefunction methods within a hybrid QM/MM scheme. Only static environmental effects have been considered by selecting a representative configuration of the peptidic system, disregarding the effects of thermal fluctuations on signal broadening and linewidths. The results serve as proof of concept for extension of this methodology to larger proteic systems with inclusion of dynamical environmental effects and provide the first reported *ab initio* simulations of 2DUV spectra that include two-excitation states and static environmental effects.

Two-Dimensional Electronic Spectroscopy

Basic principles

Two-dimensional electronic spectroscopy emerged in the last decade as a nonlinear optical technique that provides fundamental insight into the interactions and dynamics of molecular aggregates.[23] 2DES measures the full nonlinear polarization of a quatum system in third order with respect to the field–matter interaction, which contains much more information than the linear polarization measured in standard (1D) absorption experiments.[2] Thus, 2DES is the "ultimate" time-resolved nonlinear optical experiment, as it provides the maximum amount of information that can be extracted from a system within third-order nonlinear spectroscopy. Figure 1 shows a typical 2DES experiment, where a sequence of three ultrashort laser pulses interact with the sample, and the signal field emitted in the phase-matched direction is detected as a function of the three controlled excitation-pulse time delays (t_1, t_2, and t_3; also named τ, T and t, respectively). By Fourier transforming with respect to t_1 and t_3 for a fixed value of the "waiting time" t_2 (or "population time"), the 2D spectrum is derived as a function of "excitation frequency" $\omega(t_1)$ and "detection frequency" $\omega(t_3)$. The nonlinear optical response of the system can be obtained by setting the time t_2 to zero, while varying the (excited state) population time t_2, it is possible to monitor the relaxation of the excited states. The diagonal peaks in the 2D traces are located at the linear absorption peaks, whereas off-diagonal contributions provide the coupling between electronic excitations (on the same or between different chromophores) for t_2 equal to zero and monitor energy (or coherence) transfer between or within chromophores, for $t_2 > 0$. Asymmetric off-diagonal peaks, with opposite sign with respect to diagonal peaks, are associated with absorption from populated excited states. The sign, intensity, and shape of the cross-peaks are sensitive to intermolecular electronic coupling, making possible to reveal spatial molecular configurations by probing electronic transitions (Fig. 1).

Two schemes have been successfully used to implement 2DES: the heterodyne detected three-pulse photon echo[10,24]

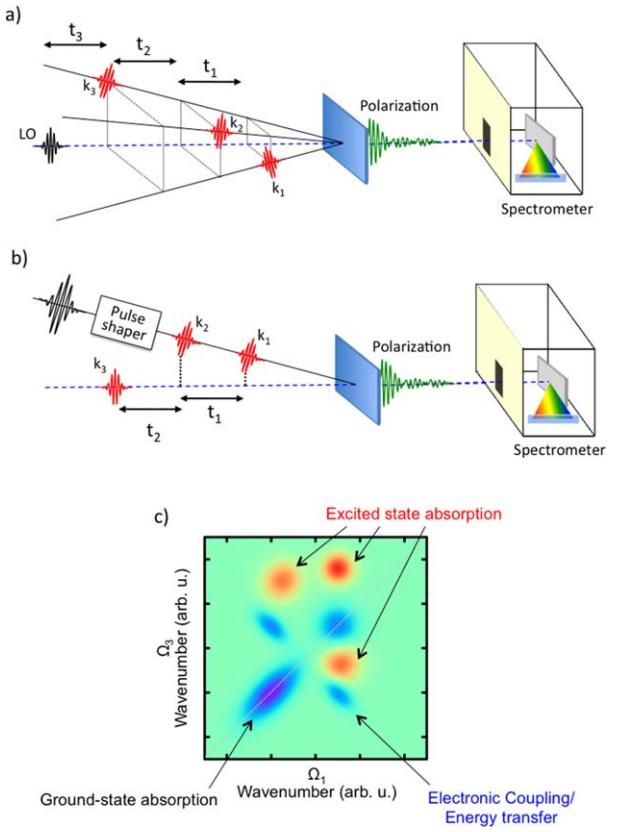

Figure 1. (a) Schematic representation of experimental setup for heterodyne detected three-pulse photon echo; LO: local oscillator, (b) schematic representation of experimental setup for partially collinear pump-probe geometry and (c) schematic representation of a 2D spectrum (at a fixed value of the waiting time t_2) showing different types, shapes, and signs of 2DES signals. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

(Fig. 1a) and the partially collinear pump-probe geometry (Fig. 1b).[25,26] Both approaches present advantages but also limitations and technical difficulties. The heterodyne-detected three-pulse photon echo scheme (Fig. 1a) has the advantage of having the emission of the four-wave-mixing signal in a background-free direction which is heterodyned by a local oscillator (LO) but requires interferometric stabilization of two pulse pairs (pulses 1-2 and 3-4, respectively) that is extremely demanding, especially for UV pulses. The partially collinear pump-probe geometry[20,25,26] uses a pulse shaper or a passive birefringent delay line[27] to generate two pump pulses with interferometric stability, whereas the nonlinear signal is self-heterodyned by the intrinsically phase-locked probe pulse; however, it has the drawback of strong background signal and mixing of many copropagating nonlinear signals. In this perspective, we will focus on simulations of 2DES spectra obtained with a three-pulse photon echo setup, as the corresponding 2D signals provide more information than in a partially collinear pump-probe geometry experiment.

Figure 1 shows the variety of peaks in a coherent 2D spectrum that need to be resolved by computer simulations. The line shape of the diagonal peaks (Fig. 1b) provides information on the relative contributions from the inhomogeneous and homogeneous linewidth broadening processes, with the extent

and slope of the elongation along the diagonal direction being time (t_2)-dependent. Excited-state absorption, ground-state bleaching, and stimulated emission optical responses contribute to sign and shape of the signals with signal overlapping leading to distortions in the line shapes. The intensities of the cross peaks are also time (t_2)-dependent, as they can originate from different processes such as excitation transfers between two different monomeric states or within different states of a single chromophore, chemical exchanges and reactions, coherence transfers, and structural transitions, depending on the chemical-physical processes involved.

2DES in the UV/visible: state-of-the-art

2DES is a powerful technique that has emerged in the last decade with applications to multichromophoric systems absorbing light in the visible region, with a special interest in the study of photosynthetic light-harvesting antenna systems and core complexes.[12–15,23] Figure 2 shows examples of application of 2DES in the visible. These studies provided fundamental insight into the degree of electronic coupling between chromophores in the light-harvesting complex (LH3) from photosynthetic purple bacteria[23] (Fig. 2a) and into the excitation energy flow in the Fenna−Matthews−Olson (FMO) photosynthetic complex from green sulfur bacteria (Fig. 2b).[12,23] Combination of experimental studies and theoretical modeling using EMs has also shown its potential in predicting pigment arrangement in the complex LH4 (Fig. 2c).[23]

Several technical issues have hampered so far extension of 2DES to the UV region but significant experimental progress has been reported very recently.[16–20] The first 2DUV experiments have been finally reported[18–20] and, therefore, 2DUV represents today the ultimate spectroscopy technique for studying systems whose lowest-frequency electronic resonances are found in the UV (e.g., DNA, proteins, cycloalkanes, quinones, etc.). The ability of 2DUV spectroscopy to track structure and dynamics of biological systems containing UV-active chormophores has been demonstrated by several

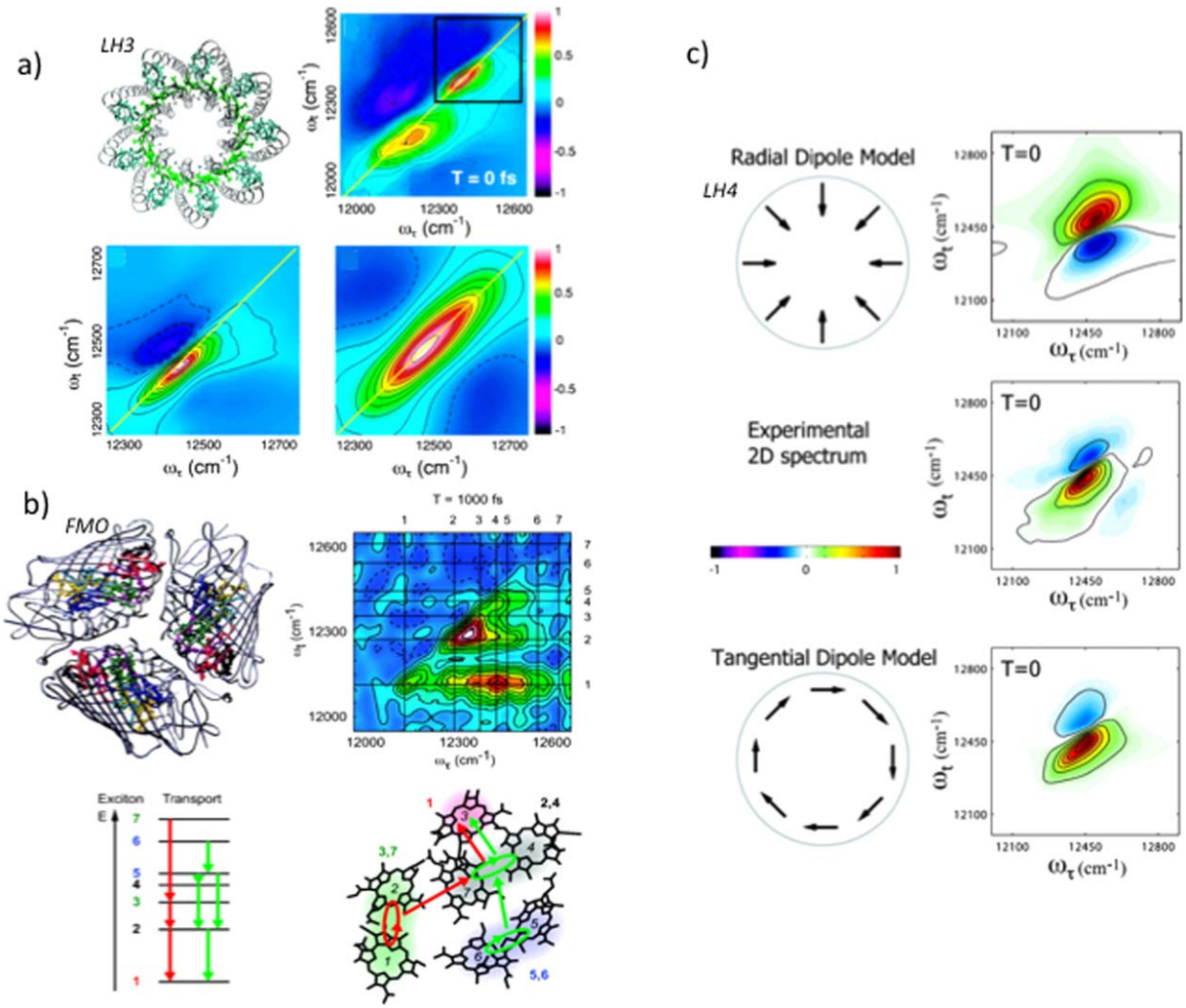

Figure 2. (a) B800 coupling in LH3 (top left): experimental 2D spectrum of B820 (top right) and B800 (boxed top right) bands (bottom left) and calculated B800 2D spectra for nearest-neighbor excitonic coupling (bottom right), (b) energy-transfer channels (bottom panels) in FMO complex (top left) as revealed from 2D spectrum (top right) and (c) modeling of chromophore configuration in LH4: calculated 2D spectra for radial (top) and tangential (bottom) dipole models compared with the experimentally obtained $T = 0$ spectrum (middle). Reproduced from Ref. [23]. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

theoretical studies carried out in the Mukamel's group.[22,28–37] Using a Frenkel exciton matrix model, 2DUV spectra of amyloid fibrils can be simulated and used to characterize their structures. Figure 3 shows three NMR structures of protein fibrils associated with Alzheimer's disease and their simulated 2DUV signals that originate from the backbone $n\pi^*$ and $\pi\pi^*$ transitions in the far-ultraviolet (2D-FUV, $\lambda = 190–250$ nm) and from aromatic side chains (Phenylalanine and Tyrosine residues) in the near-ultraviolet (2D-NUV, $\lambda \geq 250$ nm).[31]

Both 2DFUV and 2DNUV spectra show distinct cross-peak patterns that can serve as novel signatures for the secondary structure of proteins, a crucial information for understanding the aggregation mechanism of amyloid fibrils.[31]

Limitations and problems

As it must contend with several technical challenges, 2DES cannot be yet considered as a routine experimental tool. 2DES experiments require ultrabroad laser bandwidth and correspondingly ultrashort pulse durations (sub-20 fs) that can be obtained in the visible range, for example, using optical parametric amplifiers (OPAs)[38] but are less straightforward in the UV frequency domain. Ultrabroadband UV pulses have been obtained by frequency doubling of a visible OPA using achromatic phase-matching[39] or by four-wave mixing in a noble-gas-filled hollow waveguide.[18–20] Other common technical issues are the attainment of high interferometric-phase stability, the removal of undesired nonlinearities in the sample medium, and the photoionization of the sample. Successful examples of 2DES experiments reported so far have shown that the nonlinear response contains a wealth of information congested in the 2DES spectra, making sometimes their interpretation very difficult.[23] In this context, theoretical simulations represent a unique mean for the analysis and interpretation of 2D signals. Indeed, theoretical models provide the framework for direct interpretation of 2DES spectra in

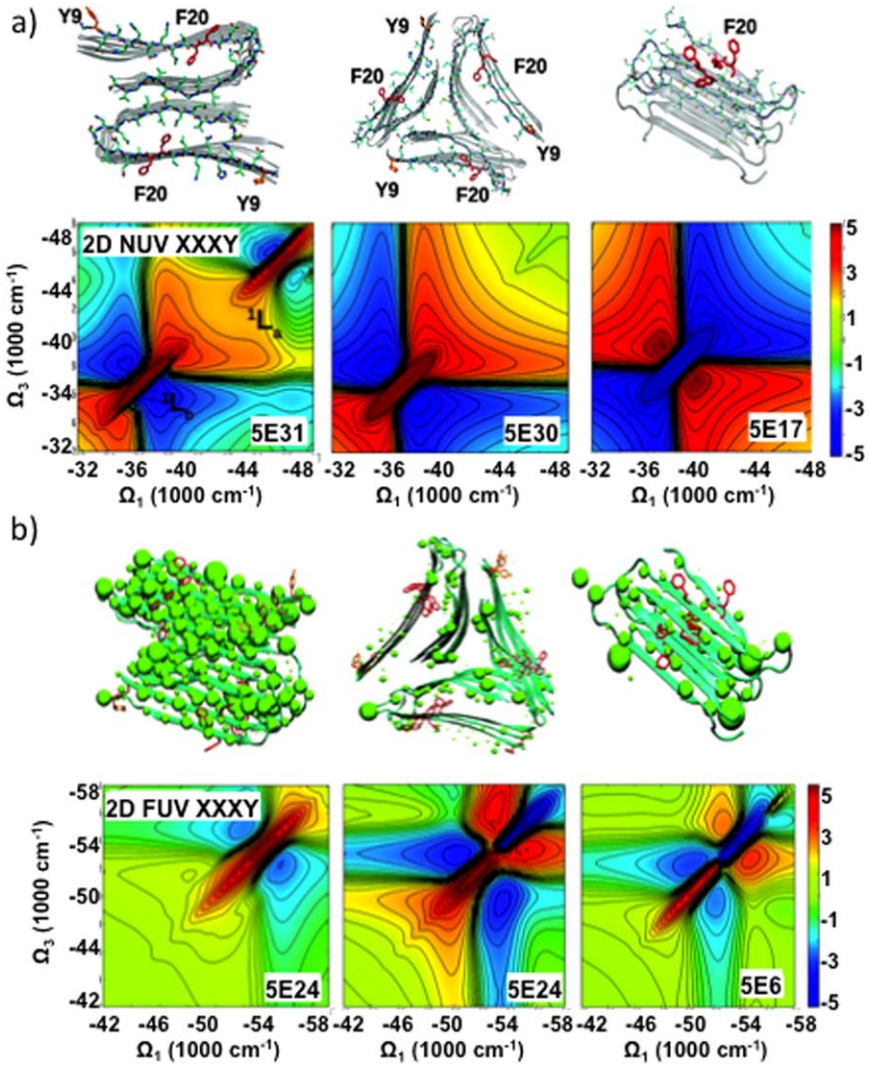

Figure 3. (a) Simulated 2DNUV photon echo signals of three amyloid fibrils models. Bottom row: Chirality-induced (*xxxy*) polarization configuration showing some differences between the three model structures originated by absorption of aromatic side chains. (b) Simulated 2DFUV photon echo signals. Top row: Transition populations for three FUV spectral regimes of the fibril models. Bottom row: Chirality-induced (*xxxy*) polarization spectra showing different patterns and intensities for the three model structures. Reproduced from Ref. [31]. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

terms of structural and dynamical changes of the system under investigation.[23]

The unified theory for nonlinear optical spectroscopy has been introduced by Mukamel,[40] formulating the nonlinear response representing the state of matter by the density matrix and following its evolution in the Liouville space. Practical applications of the nonlinear response theory for simulating the 2DES require a complete description of the electronic structure of the target system. This task is, in general, very challenging and becomes prohibitive for large systems of multiple interacting chromophores. Frenkel EMs are commonly used to calculate the nonlinear response of complex systems, making the simulations of 2DES spectra of realistic model systems computationally feasible. Here, energy levels of each isolated chromophore (and the corresponding transition dipoles) are preliminarily calculated and then parametrically used within the exciton Hamiltonian to estimate the electronic couplings of the real multichromophoric system. To partially account for signal broadening effects due to the environment, the so called efficient exciton Hamiltonian with electrostatic fluctuations (EHEF) has been developed.[36] However, it is apparent that such Frenkel EMs suffer from severe limitations, namely (i) a limited description of the chromophore excited state manifold is used (typically, no more than three energy levels are accounted for), (ii) the chromophore structure is always considered in its ground-state equilibrium geometry (making this model intrinsically unable to track intra-/inter-chromophore photoinduded dynamics and energy deactivation processes). Thus, for instance, a poor description is achieved for the double exciton manifolds, providing a crude approximation of the coherent signals. Here, we present an *ab initio* approach for simulations of 2D electronic spectra that aims at overcoming some limitations of the Frenkel EMs. Electrostatic effects on the electronic structure of the multichromophoric system are accounted for using a hybrid QM/MM approach that, in principle, can be extended to model 2DES of complex photoinduced events and radiationless decay processes of multichromophoric systems.

Ab Initio Simulations of 2DUV Spectra

Methods

In this perspective, we compare 2DUV spectra of a cyclic Cysteine-Phenylalanine-Tyrosine-Cysteine (CFYC) tetrapeptide (Fig. 4a) calculated with two different approaches, the EM based on the same Frenkel exciton matrix used in the EHEF method (see eq. (1) in Ref. [36]) and the *ab initio* model based

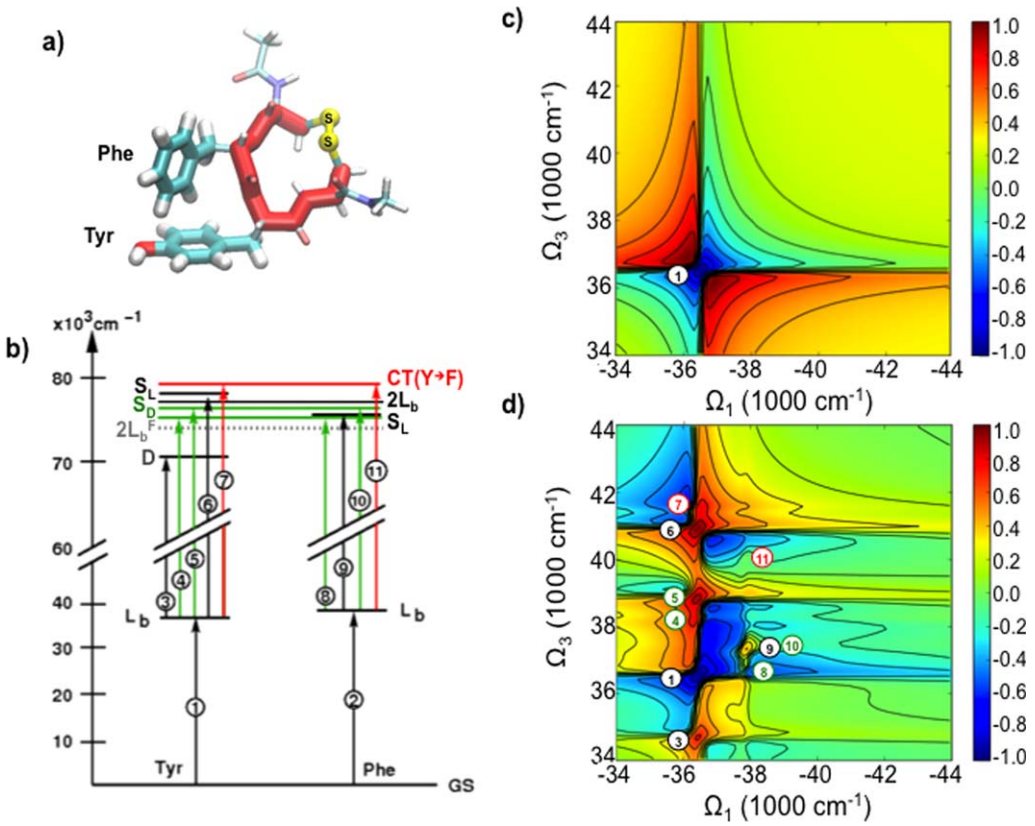

Figure 4. Simulated 2DUV spectra of an UV-active heterodimer (the Tyr-Phe dimer of the cyclic CYFC tetrapeptide, panel a). Single- and double-excitation manifolds (panel b) derived from *ab initio* (CASSCF/CASPT2//ANO-L) calculations, including ground-state (GS) and excited states (bold lines) energy levels, doubly excited state (D), high-lying singly excited states (S_L, localized in black and S_D, delocalized in green), CT (from Tyr to Phenylalanine, Y→F, in red), and electronic transitions (vertical arrows) that contribute to 2DUV signals are depicted (panel b). For reference, the level lying at the energy sum of L_b states of Tyr and Phe ($2L_b^F$, dashed line in gray) and the corresponding energy level obtained at the *ab initio* level ($2L_b$, bold line in black) are depicted. 2DUV signals obtained with the EM (panel c) and the *ab initio* SOS//QM/MM methods (panel d). [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

on the sum-over-states (SOS) approach and QM/MM (CASSCF/CASPT2//Amber) calculations (named SOS//QM/MM, or simply *ab initio*, thereafter). In the EM, the single excitation energies are calculated for isolated chromophores (by means of gas-phase *ab initio* calculations) and parametrically used in the exciton Hamiltonian, whereas the couplings between two excitons (within or between chromophores) are estimated using a quasiparticle approach,[41] with mixed doubly excited states located at energies that are the sum of the corresponding single excitation energies. In the *ab initio* description of the electronic structure, the double-exciton manifold is constituted by different localized doubly excited states, high-lying single-excited states, and delocalized doubly excited states that are located at an energy different from the exact sum of the single excitation energies (due to the quartic coupling, δ). As the nonlinear response detected by a 2D experiment is represented by the evolution of the density matrix in the Liouville space,[40] a difference in the energy levels distribution gives rise to differences in the calculated 2D electronic spectra.

Simulations of 2DUV spectra using the EM method were performed using the computational protocol previously described.[36] The computational procedure for simulating 2DUV from *ab initio* calculations includes the following steps: (i) configurational space sampling with classical MD simulations, (ii) selection of MD snapshots, (iii) refinement of selected geometries at the QM/MM level, (iv) calculation and collection of energies and transition dipole moments, and (v) calculation of 2DUV signals. While in this perspective, we consider only one selected structure of the biological target, focusing on the comparison of 2DUV spectra obtained with a Frenkel excition method or with first-principle calculations, inclusion of thermal fluctuations (that shape 2D electronic peaks and are disregarded here) is straightforward and will be documented soon.

Classical MD simulation of the cyclic CFYC oligopeptide was carried out for 40 ns after pre-equilibration, using a 2 fs time step. The CFYC molecule was capped with acetyl and N-methylamine protecting groups and solvated in a box of TIP3P water molecules,[42] using cubic periodic boundary conditions as implemented in the Amber 11 tools,[43] and the standard ff10 Amber force field. The particle-mesh Ewald approach was applied to treat long-range electrostatic interactions, with a cut-off of 12 Å for nonbonding interactions. After initial relaxation, heating in six stages of 50 K using Langevin thermostat was applied to achieve a constant temperature of 300 K. Subsequently, the system was equilibrated for 5 ns at 1 atm constant pressure. Cluster analysis of the 40 ns trajectory indicates that the aromatic rings are oriented in a T-stacked conformation in ~75% of the frames, with a shallow barrier allowing for the short-time population of unstacked conformations. The energetically lowest geometry was selected as a representative structure of the equilibrium dynamics of the cyclic CFYC oligopeptide in solution, having a T-stacked conformation (Fig. 4a). The selected geometry was refined in an optimization at QM/MM level performed with the Cobramm package,[44] using Molpro 2010[45] for state-of-the-art QM calculations, and an electrostatic embedding scheme for describing the electrostatic QM/MM interactions. The link-atom technique[46] and redistribution of residual charges among nearest neighbors were used, with both aromatic side chains included in the QM layer and the remaining atoms treated classically. The H-atom link was located along the C_α-C_β bond axis of the aromatic side chains. The peptide and the water molecules participating in hydrogen bonds with CFYC were allowed to move during the optimization, whereas the remaining bulk waters were kept frozen. The structure was optimized at the CASSCF[47] level with an active space of eight electrons in eight aromatic orbitals (four electrons and four orbitals per each chromophore). The ANO-L(4s3p2d/2s)[48] basis set was used to account for both polarization and dispersion. To collect the information on the single- and double-excitation manifolds required for calculation of 2DUV signals, a single point state-average CASSCF(12,12)/ANO-L(4s3p2d/2s) calculation with 120 roots, followed by an independent single-state CASPT2 correction for the lowest 70 roots was performed on top of the CASSCF optimized structure (the so called CASPT2//CASSCF approach)[49] using Molcas 7.7.[50] The number of roots comprised in the CASSCF and the CASPT2 calculations are chosen to ensure that all the excitations of the double-exciton manifold lying, upon PT2 correction, in the energy range reported in the 2DUV spectra are included. An imaginary shift[51] of 0.2 was used, and the Ionization Potential Electron Affinity (IPEA) shift[52] was set to zero. Transition dipole moments were calculated at the CASSCF level.

For the generation of the 2D-NUV rephasing signal (K_I), heterodyne detected by superimposing it with an LO in direction $k_I = -k_1 + k_2 + k_3$ (Fig. 1), we assume four equivalent short Gaussian laser pulses with central frequency at 38,000 cm^{-1} and a full width at half maximum 5864 cm^{-1} (corresponding to a Fourier limited pulse of ~2.5 fs). Long laser pulses (15–20 fs) can be used in realistic experiments, whereas considering that the corresponding pulse frequency narrowing implies weakening (or loss) of some 2DUV signals. The ideal central frequency and the duration of the laser pulses are, thus, dependent on the type of signals that are investigated. We focus on the NUV region between 34,000 and 44,000 cm^{-1}, showing clear signatures of the L_b signals. This pulse spectrum does not cover the DUV region above 45,000 cm^{-1} where the very intense B_a, B_b, and backbone amide absorptions are expected to obscure the L_a signals. A constant broadening of 200 cm^{-1} was used throughout. Calculations were performed with Spectron 2.7[21] for the nonchiral xxxx polarization configuration. The 2D signals, which are functions of the two frequencies Ω_1 and Ω_3, were calculated by 2D Fourier transformation along t_1 and t_3, whereas t_2 was set to zero, thereby suppressing coherent excited state dynamics.

2DUV spectra of a cyclic tetrapeptide

Figure 4 shows the comparison between the simulated 2D spectra in the NUV region (i.e., 34,000–44,000 cm^{-1}) of the solvated CFYC oligopeptide in a T-stacked conformation, obtained with the EM and the *ab initio* approaches. A single geometrical configuration of the cyclic tetrapeptide, that is representative of

the most populated cluster, has been harvested from the classical MD simulation and structurally refined at the QM/MM level (Fig. 4a). The unconstrained geometry optimization led to planarization of the two chromophores, while keeping the relative distance between them and the conformation of the peptide unchanged with respect to the initial structure.

The EM spectrum (Fig. 4c) is dominated by the diagonal L_b absorption of Tyrosine (Tyr) at ~36,500 cm^{-1}, which is close to the central frequency set for the four laser pulses (i.e., 38,000 cm^{-1}). The closely lying L_b absorption of Phenylalanine (Phe) around 38,000 cm^{-1} has a molar absorption coefficient much smaller than Tyr (195 vs. 1405 M^{-1} cm^{-1}, respectively)[53] and, hence, not clearly visible in the spectrum.

Within the *ab initio* SOS//QM/MM approach, excitation energies and transition dipole moments of the singly and doubly excited manifolds lying in a large energy window are needed to simulate the 2DUV spectrum. Figure 4b shows the level scheme obtained from CASPT2//CASSCF calculations, indicating the energy levels that are involved in bright transitions among the singly and doubly excited manifolds. CASPT2//CASSCF calculations indicate the presence of several excited states lying in the region between 70,000 and 80,000 cm^{-1} (i.e., between 34,000 and 44,000 cm^{-1} from the L_b manifold), with nonvanishing transition dipole moments, out of the singly excited (L_b) manifold (36,500–38,000 cm^{-1}).

As for the spectrum obtained at the EM level, the calculated SOS 2DUV spectrum is dominated by the L_b absorption of Tyr. However, the latter spectrum is much richer than the EM one, due to the presence of resolved bright excitations to double and high-lying single excited states appearing in the spectrum as off-diagonal peaks. Part of the off-diagonal contributions to the nonlinear signal arises from excitations to local doubly excited states that can be accessed from the singly excited manifold of the chromophores. In particular, an excitation out of the L_b state of Tyr to the localized doubly excited state (D) appears in the *ab initio* spectrum at Ω_1 = 36,516 cm^{-1} and Ω_3 = 34,625 cm^{-1} (peak 3). Due to its local nature, this signal appears independently of the presence of a chromophore–chromophore interaction, however, its spectral position and relative intensity depends on the immediate local environment and, thus, contains implicit structural information.

Excitations to local doubly excited states are usually neglected in the Frenkel EMs, where only the double excitations associated with combination of electronic states in the single-exciton manifold are considered. In particular, in the cyclic tetrapeptide, a mixed doubly excited state resulting from combination of the L_b states of the two chromophores ($2L_b$) can be accessed with excitation out of the L_b states of the chromophores, either Tyr or Phe. In the Frenkel EM, the energy of the $2L_b$ state is approximated to be the exact sum of the energies of the two L_b states ($2L_b^F$, see Fig. 4b), leading to two symmetric off-diagonal peaks in the 2D map. However, in the CFYC tetrapeptide, the excitations from the L_b states to the $2L_b$ state have small oscillator strengths and thus, they are not visible in the EM spectrum (Fig. 4c). While the *ab initio* SOS//QM/MM results confirm that the $2L_b$ off-diagonal peaks are covered by the intense L_b bleach signal of Tyr, they also indicate the presence of quartic coupling shifting the energy of the $2L_b$ state by 1359 cm^{-1} higher than the sum of the energies of the two L_b states (Fig. 4b). This outcome implies that the *ab initio* SOS//QM/MM approach allows assessment of the quartic couplings between electronic states of the single-exciton manifold, determining more accurately the positions of the positive signals associated with mixed doubly excited states.

Other important off-diagonal contributions to the 2DUV spectrum are caused by excitations from local excited states (such as the L_b states) to high-lying singly excited states (S) that could have localized, delocalized (S_L and S_D, black and green lines in Fig. 4b, respectively) or charge-transfer (CT) character, depending on the nature of the molecular orbitals involved in the excitation out of the L_b states and the permanent dipole moment of the final excited state. For example, excitation out of the L_b state of Tyr to a high-lying singly excited state localized on the same chromophore gives rise to a strong positive peak at Ω_1 = 36,516 cm^{-1} and Ω_3 = 40,912 cm^{-1} (peak 6 in Fig. 4d), whereas the analogous transition from the L_b state of Phe originates a strong positive signal at Ω_1 = 37,938 cm^{-1} and Ω_3 = 37,257 cm^{-1} (peak 9), which overlaps with the tail of the main diagonal (negative) peak of the Tyr L_b state. Moreover, excitations from the L_b states of Tyr and Phe to two delocalized high-lying singly excited states (S_D) give rise to two strong positive peaks at Ω_1 = 36,516 cm^{-1} and Ω_3 = 38,186, 38,831 cm^{-1} (peaks 4 and 5, respectively) and two strong signals at Ω_1 = 37,938 cm^{-1} and Ω_3 = 36,764, 37,409 cm^{-1} (peaks 8 and 10, respectively). Again, the excitations out the L_b state of Phe (peaks 8 and 10) overlap with the tail of the Tyr L_b state absorption, showing weak positive peaks in the 2D map. The excitations out the L_b states of both chromophores can also involve occupation of CT states, high-lying singly excited states with permanent dipole moment different from the ground state. In particular, transitions to the excited state characterized by transfer of charge from Tyr to Phe, namely CT(Y==>F), contribute to the 2D spectrum with two positive signals at Ω_1 = 36,516 cm^{-1} and Ω_3 = 41,517 cm^{-1} (peak 7) and Ω_1 = 37,938 cm^{-1} and Ω_3 = 40,095 cm^{-1} (peak 11), corresponding to excitations from the L_b states of Tyr and Phe, respectively.

The positions and the intensities of the signals involving delocalized high-lying singly excited states depend explicitly on the electronic coupling between the two aromatic residues. In the T-stacked conformation of the CFYC tetrapeptide, we observed several bright excitations to such delocalized excited states (peaks 4-5, 7-8, 10), including excitation to a CT state (Fig. 4d). These signals are neglected in the Frenkel exciton approach (Fig. 4c), implying that the EM signal is missing some of the structural information and part of the electronic couplings contained in the 2D map, which can be instead revealed by simulations based on first-principle calculations.

Conclusions

Two-dimensional electronic spectroscopy holds great potential for studying structure, dynamics, and electronic excitations in biological systems whose lowest-frequency electronic resonances

are found in the UV, such as proteins and DNA (or RNA). Accurate simulations of 2DUV signals are crucial for the analysis and the interpretation of complex 2D spectra. We present a computational strategy for first-principle simulations of 2DUV signals using QM state-of-the-art (CASPT2//CASSCF) wavefunctions within a QM/MM scheme and combined with the SOS approach. Its application to a model proteic system, a small oligopeptide containing two UV-active aromatic side chains, is shown. The proposed *ab initio* simulations are able to resolve off-diagonal peaks in the 2D spectra that contain structural information of the oligopeptide and chromophore–chromophore electronic coupling between aromatic residues, beyond standard Frenkel EMs. Future integration of thermal fluctuation effects on 2DUV spectra is necessary for proper characterization of signal broadenings and linewidths, allowing for suitable comparison with experimental data.

Acknowledgments

I.R. and A.N. thank N. Preketes, D. Healion, and J. Jiang for very useful discussions.

Keywords: two-dimensional spectroscopy · electronic coupling · wavefunction methods · protein structure · 2DUV spectra

How to cite this article: I. Rivalta, A. Nenov, G. Cerullo, S. Mukamel, M. Garavelli, *Int. J. Quantum Chem*. **2014**, *114*, 85–93. DOI: 10.1002/qua.24511

[1] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, *64*, 2229.
[2] S. Mukamel, *Annu. Rev. Phys. Chem.* **2000**, *51*, 691.
[3] D. M. Jonas, *Annu. Rev. Phys. Chem.* **2003**, *54*, 425.
[4] P. Hamm, M. H. Lim, R. M. Hochstrasser, *J. Phys. Chem. B* **1998**, *102*, 6123.
[5] W. Zhuang, T. Hayashi, S. Mukamel, *Angew. Chem. Int. Ed.* **2009**, *48*, 3750.
[6] J. Bredenbeck, P. Hamm, *Chimia* **2007**, *61*, 45.
[7] M. T. Zanni, R. M. Hochstrasser, *Curr. Opin. Struct. Biol.* **2001**, *11*, 516.
[8] J. Bredenbeck, A. Ghosh, H. K. Nienhuys, M. Bonn, *Acc. Chem. Res.* **2009**, *42*, 1332.
[9] M. H. Cho, *Chem. Rev.* **2008**, *108*, 1331.
[10] M. L. Cowan, J. P. Ogilvie, R. J. D. Miller, *Chem. Phys. Lett.* **2004**, *386*, 184.
[11] T. Brixner, T. Mancal, I. V. Stiopkin, G. R. Fleming, *J. Chem. Phys.* **2004**, *121*, 4221.
[12] T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, G. R. Fleming, *Nature* **2005**, *434*, 625.
[13] E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer, G. D. Scholes, *Nature* **2010**, *463*, 644.
[14] G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship, G. S. Engel, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12766.
[15] G. S. Engel, T. R. Calhoun, E. L. Read, T.- K. Ahn, T. Mancal, Y.- C. Cheng, R. E. Blankenship, G. R. Fleming, *Nature* **2007**, *446*, 782.
[16] U. Selig, C.- F. Schleussner, M. Foerster, F. Langhojer, P. Nuernberger, T. Brixner, *Opt. Lett.* **2010**, *35*, 4178.
[17] C.- H. Tseng, S. Matsika, T. C. Weinacht, *Opt. Express* **2009**, *17*, 1888.
[18] C.- H. Tseng, P. Sandor, M. Kotur, T. C. Weinacht, S. Matsika, *J. Phys. Chem. A* **2012**, *116*, 2654.
[19] B. A. West, J. M. Womick, A. M. Moran, *J. Phys. Chem. A* **2011**, *115*, 8630.
[20] B. A. West, A. M. Moran, *J. Phys. Chem. Lett.* **2012**, *3*, 2575.
[21] D. Abramavicius, B. Palmieri, D. V. Voronine, F. Sanda, S. Mukamel, *Chem. Rev.* **2009**, *109*, 2350.
[22] Z. Li, S. Mukamel, *J. Phys. Chem. A* **2007**, *111*, 11579.
[23] N. S. Ginsberg, Y.- C. Cheng, G. R. Fleming, *Acc. Chem. Res.* **2009**, *42*, 1352.
[24] T. Brixner, I. V. Stiopkin, G. R. Fleming, *Opt. Lett.* **2004**, *29*, 884.
[25] P. F. Tian, D. Keusters, Y. Suzaki, W. S. Warren, *Science* **2003**, *300*, 1553.
[26] E. M. Grumstrup, S.- H. Shim, M. A. Montgomery, N. H. Damrauer, M. T. Zanni, *Opt. Express* **2007**, *15*, 16681.
[27] D. Brida, C. Manzoni, G. Cerullo, *Opt. Lett.* **2012**, *37*, 3027.
[28] J. J. Rodriguez, S. Mukamel, *J. Phys. Chem. B* **2012**, *116*, 8830.
[29] A. R. Lam, J. Jiang, S. Mukamel, *Biochemistry* **2012**, *51*, 6262.
[30] H. Ren, J. Jiang, S. Mukamel, *J. Phys. Chem. B* **2011**, *115*, 13955.
[31] A. R. Lam, J. Jiang, S. Mukamel, *Biochemistry* **2011**, *50*, 9809.
[32] J. Jiang, S. Mukamel, *PCCP* **2011**, *13*, 2394.
[33] J. Jiang, S. Mukamel, *J. Phys. Chem. B* **2011**, *115*, 6321.
[34] J. Jiang, S. Mukamel, *Angew. Chem. Int. Ed.* **2010**, *49*, 9666.
[35] J. Jiang, D. Abramavicius, C. Falvo, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Phys. Chem. B* **2010**, *114*, 12150.
[36] J. Jiang, D. Abramavicius, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Phys. Chem. B* **2010**, *114*, 8270.
[37] D. Abramavicius, J. Jiang, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Am. Chem. Soc.* **2010**, *132*, 7769.
[38] D. Brida, C. Manzoni, G. Cirmi, M. Marangoni, S. Bonora, P. Villoresi, S. De Silvestri, G. Cerullo, *J. Opt.* **2010**, *12*, 013001.
[39] P. Baum, S. Lochbrunner, E. Riedle, *Opt. Lett.* **2004**, *29*, 1686.
[40] S. Mukamel, Principles of Nonlinear Optical Spectroscopy; O.U.P.: New York, **1995**.
[41] S. Mukamel, Y. Nagata, In 22nd Solvay Conference on Chemistry: Quatum Effects in Chemistry and Biology; G. R. Fleming, G. D. Scholes, A. DeWit, Eds., Elsevier: Brussels, BELGIUM, Procedia Chemistry, **2011**.
[42] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
[43] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668.
[44] P. Altoe, M. Stenta, A. Bottoni, M. Garavelli, *Theor. Chem. Acc.* **2007**, *118*, 219.
[45] H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schutz, *WIREs Comput. Mol. Sci.* **2012**, *2*, 242.
[46] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718.
[47] B. O. Roos, Ab Initio Methods in Quatum Chemistry: Part II; Wiley: Chicester, UK, **1987**.
[48] P. O. Widmark, P. A. Malmqvist, B. O. Roos, *Theor. Chim. Acta* **1990**, *77*, 291.
[49] M. Garavelli, F. Bernardi, A. Cembran, In Computational Photochemistry; M. Olivucci, Ed.; Elsevier: Amsterdam, **2005**; pp. 191–223.
[50] F. Aquilante, L. De Vico, N. Ferre, G. Ghigo, P. A. Malmqvist, P. Neogrady, T. B. Pedersen, M. Pitonak, M. Reiher, B. O. Roos, L. Serrano-Andres, M. Urban, V. Veryazov, R. Lindh, *J. Comput. Chem.* **2010**, *31*, 224.
[51] N. Forsberg, P. A. Malmqvist, *Chem. Phys. Lett.* **1997**, *274*, 196.
[52] G. Ghigo, B. O. Roos, P. A. Malmqvist, *Chem. Phys. Lett.* **2004**, *396*, 142.
[53] G. D. Fasman, Hand Book of Biochemistry and Molecular Biology, **1976**; CRC Press: Cleveland, Ohio, **1976**; Vol. 1.

Received: 24 May 2013
Revised: 21 June 2013
Accepted: 26 June 2013
Published online 10 July 2013