Photochemistry and Photobiology, 2017, 93: 1368–1380

On the Simulation of Two-dimensional Electronic Spectroscopy of Indole-containing Peptides†

Angelo Giussani[1], Jacopo Marcheselli[2], Shaul Mukamel[3], Marco Garavelli*[1,4] and Artur Nenov*[1,4]

[1]Dipartimento di Chimica "G. Ciamician", Università degli Studi di Bologna, Bologna, Italy
[2]International School for Advanced Studies (SISSA), Trieste, Italy
[3]Department of Chemistry, University of California, Irvine, CA
[4]Dipartimento di Chimica Industriale "Toso Montanari", Universita degli Studi di Bologna, Bologna, Italy

Received 31 January 2017, accepted 27 February 2017, DOI: 10.1111/php.12770

ABSTRACT

A benchmark study of low-cost multiconfigurational CASSCF/CASPT2 schemes for computing the electronic structure of indole is presented. This facilitates the simulation of near-ultraviolet (UV) pump visible (VIS) probe (*i.e.* two-color) two-dimensional electronic spectra (2DES) of homo- and hetero-aggregates as well as for processing of multiple snapshots from molecular dynamics simulations. Fingerprint excited-state absorption signatures of indole are identified in a broad spectral window between 10 and 25 k cm^{-1}. The 18–24 k cm^{-1} spectral window which has no absorption of the monomer and noninteracting aggregates is ideally suited to embed charge-transfer signatures in stacked aggregates. The small peptide Trp-cage, containing a tryptophan and a tyrosine amino acids, having indole and phenol as side chains, respectively, serves to prove the concept. Clear charge-transfer signatures are found in the proposed spectral window for an interchromophore distance of 5 Å making near-UV pump VIS probe 2DES a suitable technique for resolving closely packed aggregates. We demonstrate that 2DES utilizing ultra-short pulses has the potential to resolve the nature of the spectroscopically resolved electronic states and that the line shapes of the excited-state absorption signals can be correlated to the polarity of the relevant states.

INTRODUCTION

The function of a protein is determined not only by its amino acid constituents (primary structure), but also is based on its local structure (secondary structure), its topology (tertiary structure) and its conformation with respect to other proteins in protein complexes (quaternary structure). It is consequently not surprising that the study of protein composition and folding dynamics constitutes a very active field of research. While linear absorption (LA) and circular dichroism are powerful spectroscopic tools for differentiating secondary protein structures (1–3), they suffer from high spectral congestion and do not provide direct information on tertiary or quaternary structures.

The advent of two-dimensional (2D) spectroscopy (4–24), a nonlinear multipulse technique, introduced new potential tools for the determination of protein structures. The pioneering 2D technique, nuclear magnetic resonance (2DNMR) (8), is a key experimental methodology capable of resolving detailed structural information of proteins by detecting the couplings of spin transitions. The technique has, however, some drawbacks like the lack millisecond temporal resolution which impedes its application to protein folding process (9). 2D spectroscopy has been recently extended in the optical domain to the infrared (IR), visible (VIS) and ultraviolet (UV) regimes, thus addressing vibrational (2DIR) and electronic (2DES) transitions (9–24). 2DIR is nowadays a standard technique that allows insights into the conformational dynamics through probing vibrational modes of backbone amide groups. Theoretical simulations have demonstrated that the 2D peak line shapes can be used to distinguish secondary structures as α-helices and β-sheets (10). 2DIR is commonly used to study solvent reorganization dynamics (11,12). Broadband pulses in the UV have extended the applicability of 2D spectroscopy in two ways. First, electronic transitions in the aromatic side chains of some amino acids can be now addressed alongside transitions in the amide groups, profiting from the fact that aromatic residues are relatively rare in proteins and can be used as site-specific secondary and tertiary structure probes (19,20,23,24). Second, ultrafast processes and excited-state population relaxation can be temporally resolved (17,21,22).

Two-dimensional electronic spectra, like 2DNMR and 2DIR, is a nonlinear technique that measures the macroscopic polarization of a quantum system in third order with respect to the field–matter interaction. In the technique, the sample interacts with three pulses, whereas the experimental control knobs—pulse central frequency (VIS, UV spectroscopy), bandwidth (temporal and spectral resolution), wave vector direction and time order (rephasing, nonrephasing, quasi-absorptive techniques), time delay (temporal resolution), polarization (chiral and nonchiral techniques)—open intriguing possibilities for application. In the most common implementation of 2DES, the time delays between the first two (pump) pulses ($t1$) and between the third (probe) pulse and a local oscillator ($t3$), which heterodynes the field emitted by the sample in response to the perturbation by the three incident pulses (thus facilitating the extraction of the phase and time dependence of the field) is scanned, while the time delay between the second and third pulses ($t2$) is kept fixed.

*Corresponding authors' emails: marco.garavelli@unibo.it (Marco Garavelli) and artur.nenov@unibo.it (Artur Nenov)
†This article is a part of the Special Issue dedicated to Prof. Volker Buß on the occasion of his 75th birthday

Fourier transformation of the heterodyne signal along $t1$ and $t3$ provides a two-dimensional map for a fixed waiting time $t2$ in which all electronic transitions involved in the three field–matter interactions are resolved with highest spectral resolution. The signals can be classified into three groups: ground-state bleach signals (GSB), stimulated emission (SE) and excited-state absorptions (ESA). The latter arise from highly specific excitations falling within the envelope of the probe pulse which are accessible (bright) from the low-lying excited states populated with the pump-pulse pair. Hence, they can be considered as spectral fingerprints of the low-lying states. Using a model peptide containing the aromatic residues benzene and phenol, we demonstrated recently that ESA signals resolved through near-UV pump VIS probe 2DES provide a wealth of novel information and bear the potential to resolve aromatic stacking and conformational dynamics (23,24).

The application of near-UV pump 2D electronic spectroscopy utilizing near-UV pump pulses to proteins implies the excitation of the aromatic amino acids that can normally be found in a protein: phenylalanine (Phe), whose chromophore is benzene, tyrosine (Tyr), having a phenol residue, tryptophan (Trp), in which the chromophore indole is present, and to some extent histidine (His), whose aromatic group imidazole exhibits a tail in the near-UV spectral window. Among the aforementioned aromatic systems, indole is the protein chromophore with the highest oscillator strength in the near-UV (having a cross section about three times more intense than that of phenol and an order of magnitude more intense than benzene), and it undoubtedly constitutes the most important emissive source in polypeptides. LA spectra of tryptophan-containing proteins are dominated by the electronic structure of indole, while Tyr and Phe signatures remain hidden under the Trp envelope. Thanks to the higher spectral resolution and signal abundance, 2DES is a technique which could reveal insights about Tyr and Phe either in a direct way, through probing fingerprint signals in spectral regions where Trp does not exhibit bright ESA, or in an indirect way, through probing common transitions like charge-transfer (CT) states emerging in spatially close lying (*i.e.* stacked) chromophores. Knowing the electronic structure of the aromatic residues is, thus, imperative for interpreting 2DES spectra, as well as developing tailored experiments. In fact, the interpretation of a 2DES spectrum strongly relies on the theoretical level used to compute the system's transition energies and dipole moments (25–28). Employing UV and VIS radiations ideally allows the detection of all bright transitions among the electronic states placed up to 10 eV from the ground state. Already in dimeric and trimeric systems, over a hundred excited states may be encountered in this energy window, resulting in a large amount of information to be computed and analyzed. The latter task constitutes a considerable computational challenge related to both the methodological complexity of describing excited states of different nature (*e.g.* ionic *vs* covalent, singly *vs* multiply excited) on an equal footing and to the required computational effort (memory, CPU time, etc.). To meet these challenges, we developed a protocol based on the complete active space self-consistent field theory (CASSCF) and multiconfigurational second-order perturbation (CASPT2) (24–26) relying on their implementation in the quantum chemistry software Molcas 8 and profiting on recent progress in speeding up integral evaluation, parallelization and I/O efficiency. With this approach, we recently described the electronic structure of Trp, Tyr and Phe with unprecedented accuracy (23–25). The precise theoretical description of a chromophore's electronic structure requires very demanding computations due to the necessity to include significant part of the dynamic correlation already at the CASSCF level (25–27). Such expensive reference computations are consequently impractical for the description of multichromophore systems or for processing large number of structures as required for addressing the conformational freedom of a peptide or for tracking folding dynamics. It is then necessary to find a trade-off between accuracy and computational effort, that is, to determine a computationally affordable level of theory still able to describe the excited states of the system with satisfactory accuracy, thereby assessing the introduced error.

In the present contribution, the performance of different levels of theory for the description of indole's excited states is tested against the reference data published previously (25) and a computational strategy which provides qualitatively correct results at a computationally affordable cost is outlined. This strategy is then employed to simulate near-UV pump VIS probe (two-color) 2DES of the Trp-cage protein, which contains two aromatic amino acids, Trp and Tyr. Trp-cage has been chosen for different reasons: (1) It is a widely studied system with a wealth of experimental data available (19,20,29); (2) it is a very simple peptide, being formed of 20 amino acids, making it appealing for computational studies; (3) it contains two aromatic chromophores, consequently allowing us to address the applicability of 2DES for studying interchromophore interactions. In the present contribution, 2DES spectra of Trp-cage are simulated on the basis of the SOS//QM/MM approach developed in our group (28). It is a fully ab-initio-based approach combining hybrid quantum mechanics/molecular mechanics (QM/MM) with electrostatic embedding for computing electronic transitions and transition dipole moments to the sum-over-states approach (SOS) within the dipole approximation for simulating nonlinear spectroscopy in its different flavors.

MATERIALS AND METHODS

The computation of the vertical excitation energies of the electronic excited state of the different systems here studied has been performed employing the restricted active space self-consistent (RASSCF) method and its second-order perturbation theory extension, named RASPT2 (30,31). As in the CASSCF methodology, in the RASSCF method, the molecular orbitals of the system are subdivided into active, inactive and virtual orbitals and the electronic wave function is derived through a variational optimization of both the CI expansion coefficients and the orbitals. The difference with respect to the CASSCF methods resides in the fact that the active space in a RASSCF computation is subdivided in three subspaces: the RAS1 space, in which a maximum number of holes (*i.e.* the maximum number of electrons that can be simultaneously promoted to the remaining active orbitals) is defined; the RAS2 space, where any permutation of active electrons into the active orbitals is allowed; and the RAS3 space, in which a maximum number of excitation (*i.e.* the maximum number of electrons distributed among those orbitals) is defined. The wave function is then expanded in the basis of all the configuration state functions obtainable distributing the active electrons in the active space accordingly to the above-described restrictions. A RASSCF active space is normally defined with three couples of numbers: The first defines the number of holes and orbitals in RAS1; the second specifies the number of electrons and orbitals in RAS2; the last couple corresponds to the number of excitations and orbitals in RAS3. In the RASSCF computation, a state-averaged (SA) approach has been adopted, and the RASPT2 method has been employed in its single state (SS) variant due to limited active space sizes employed (32). In the RASPT2 computations, an IPEA shift of 0.0 a.u. has been used (33), a normal choice previously adopted in the description of systems of very diverse size and nature (34,35), while the imaginary shift parameter has been

scanned in the range 0.2–0.6 a.u. (36). The basis set of atomic natural orbitals (ANO-L) contracted to C,N [4s,3p,1d]/H[2s1p] has been employed (37). The Cholesky decomposition was used to speed up the calculation of two-electron integrals (38). The RASSCF and RASPT2 computations have been performed using the Molcas 8 software (39).

In the present contribution, three systems have been studied: the indole molecule; an ideal system composed by an indole and a phenol molecule separated by more than 10 Å (hereafter the noninteracting indole-phenol dimer); and the Trp-cage protein in water. The two former systems have been treated completely at a quantum mechanics level, while, due to its size, for the Trp-cage, a QM/MM approach has been adopted. The QM/MM computations have been performed using the COBRAMM interface developed in our group, whose details have been presented elsewhere (40). The QM part is composed by the two aromatic side chains of Trp and Tyr (*i.e.* indole and phenol), while the remaining part of the peptide, together with solvent compose the MM part, described with Amber's FF99SB force field (41). The structures for which the corresponding spectra have been simulated have been taken from a molecular dynamics simulation of the folding process previously published (19,20). The latter has been performed using an implicit model of water, while, due to COBRAMM's operational principle, an explicit water model had to be applied to the selected snapshots. For that reason, the snapshots were dissolved in a cubic box of water molecules of 15 Å length. A thermalization of the snapshots was then reached in a three-step procedure: First, the system's energy was minimized; second, the system was heated from 0 to 300 K in six subsequent 5-ps-long dynamics simulation runs in which the temperature was increased step-wise by 50 K; third, a 100-ps molecular dynamics at constant temperature (300 K) and pressure (1 atm) was performed. The solute's geometry was kept fixed throughout the procedure to preserve the original snapshot conformation. The final geometry of the MD simulation has then been processed through COBRAMM and refined by means of a QM/MM ground-state optimization, using the second-order Moller-Plesset (MP2) method for the QM region, thereby employing the Berny optimization algorithm implemented in Gaussian 03 through its interface with COBRAMM (40,42) During QM/MM optimizations, the degrees of freedom of the MM part were frozen. At the QM-MM interface, a hydrogen link-atom approach was adopted (40,43).

The SA-RASSCF computed transition dipole moments (TDMs) and SS-RASPT2 energies have been used to compute two-dimensional electronic spectra employing the Spectron 2.7 program (44), which simulates quasi-absorptive 2DES maps by the sum-over-state approach in the dipole approximation (28). Details on the working equations are presented in Appendix S1 (see section titled Computation of the 2D spectra). A static picture has been adopted to describe the coherence dynamics during time $t1$ and $t3$. In the calibration procedure of the computational protocol, signal broadening due to spectral diffusion and electronic dephasing caused by the coupling to the environment bath and by the finite excited-state lifetimes is represented through phenomenological state-specific dephasing rates. Our goal is to establish a low-cost protocol which most accurately reproduces the recently published computational reference of the electronic structure of indole and not to simulate realistic lineshapes. For the application to Trp-cage, inhomogeneous peak broadening due to the spread of transition energies in the bulk under environmental fluctuations is incorporated (see Results for details on the procedure). The waiting time $t2$ between the pump pulse pair and the probe pulse was set to zero. Within this framework, only computations at the Frank–Condon (FC) point are required to simulate the spectra. All calculated signals use the all-parallel pulse polarization configurations. Excited-state absorptions (ESA) appear as positive (red) peaks in the 2D spectra.

RESULTS AND DISCUSSION

Benchmarking reduced level of theory for the description of indole's excited states

In the following lines, the results of the RASPT2//RASSCF benchmark performed specifically for the indole system are presented. The electronic structure of indole is characterized by two low-lying singlet excited states, named, according to Platt's nomenclature, L_b and L_a. In the Franck–Condon (FC) region, L_b and L_a are, respectively, the S_1 and S_2 states both in the isolated system and in nonpolar solvent and the absorption bands are well separated, while in polar solvents, the L_a state exhibits pronounced stabilization making the state order dependent on the solvent's polarity (27,45). The L_b wave function is mainly composed of the negative linear combination of the HOMO (H) → LUMO (L) + 1 and H-1 → L configuration state functions (CSFs), while the L_a wave function is dominated by the H → L CSF (see Table 1). Using a valence bond terminology, the L_b state can be classified as a covalent, while the L_a as ionic. The transitions from the ground state to L_b and L_a have nonzero oscillator strengths, L_a being brighter than L_b. Together with the so called Bb and Ba states absorbing in the far-UV, these states give rise to the most intense LA bands below 6.5 eV.

While the aforementioned states have been the subject of various theoretical and experimental efforts (27–29,45), little is known about the higher excited states of indole. The latter are among the main protagonists in a 2DES experiment. To the best of our knowledge, the only study concerned with the manifold of higher lying excited states in indole comes from Serrano-Andrés and Roos and reports the lowest seven singlet $\pi \rightarrow \pi^*$ transitions up to ~6.5 eV (46). This was one of several seminal papers on the electronic structure of aromatic compounds which demonstrated the predictive power of the CASSCF/CASPT2 protocol and boosted momentum into its popularization. Recently, we extended the computational window up to 10 eV through a systematic increase in the level of theory within the CASSCF/CASPT2 framework, thereby expanding the full π-valence active space by adding virtual orbitals with higher orbital momenta (22). In that way, 25 states (hereafter referred to as the reference data set, see Table 1), which can be considered among the most reliable computations on the indole high-lying states, were obtained at the RAS(0.0|10.9|2.12)/SS-RASPT2 ANO-L[432/21] level of theory. The near-UV pump VIS probe quasi-absorptive nonchiral 2DES spectrum is presented in Fig. 1. It comprises the lowest ten excited states of indole, the remaining 15 excited states being accessible only with near-UV probe pulses. As expected due to the nature of the experiment, two traces appear in the spectrum, associated with the L_b and L_a transition, populated with the pump-pulse pair, at 35.1 and 39.6 k cm^{-1}, respectively. The L_b trace is characterized by two coincident peaks at around 17.0 k cm^{-1} due to the electronic promotion from the L_b to states 6 and 7 of indole. A second signature of the L_b trace, the ESA due to probing state 10 falls just outside of the probing window, nevertheless the tail of the peak can be spot around 26 k cm^{-1}. Along the L_a trace, two main signals are recognizable, one around 12.6 k cm^{-1}, associated with the electronic promotion into states 7 and 8, and a second one around 22.1 k cm^{-1} arising from probing states 10 and 11. A weaker signal at 14.7 k cm^{-1} is attributed to state 9. We note the existence of a nearly signal-free region (18–25 k cm^{-1} along the L_b trace, 16–22 k cm^{-1} along the L_a trace), bearing the potential to become a probe window for the detection of CT signals arising in aggregates.

Despite the valuable information given by the reference data, the high computational cost of the reference level of theory together with the intricate procedure for elimination of Rydberg orbitals imply that such a computational strategy is not feasible to operate in aggregates, where the presence of additional chromophores will force an unaffordable enlargement of the active space, or for processing multiple snapshots of molecular

Table 1. Vertical excitation energies (E_{VA}, eV) of the UV–vis reachable excited states of indole computed at the RAS(0.0|10.9|2.12) (reference data) and RAS(4.5|0.0|4.4) level of theory for the isolated molecule.

	Reference data					RASSCF(4.5|0.0|4.4)					
State	E_{VA} RAS SCF	E_{VA} RAS PT2*	TDM L_b	TDM L_a	Main CSFs†	E_{VA} RAS SCF	E_{VA} RAS PT2‡	TDM L_b	TDM L_a	Main CSFs†	Error§
2 (L_b)	4.69	4.35 (4.37)	0.34¶	–	H-1 → L (0.45) H → L + 1 (0.26)	4.79	4.31 (4.38)	0.38¶	–	H-1 → L (0.46) H → L + 1 (0.27)	−0.04 (+0.03)
3 (L_a)	5.77	4.91 (4.77)	0.21	0.81¶	H → L (0.60) H-2 → L (0.06)	6.12	4.64 (4.76)	0.23	0.76¶	H → L (0.59) H-2 → L (0.08)	−0.27 (−0.15)
4	6.54	5.96	0.10	0.32	H-1 → L + 2 (0.14) H → L + 2 (0.12) H-2 → L + 1 (0.11) H-1 → L + 1 (0.07) H-2 → L (0.06) H → L (0.06)	6.86	5.84 (5.98)	0.14	0.44	H-2 → L + 1 (0.15) H-1 → L + 2 (0.13) H → L (0.09) H → L + 2 (0.09) H-1 → L + 1 (0.06)	−0.12 (+0.02)
5	6.96	6.00 (6.02)	0.09	0.72	H → L + 1 (0.20) H-2 → L (0.11) H-1 → L + 2 (0.11) H → L + 2 (0.09)	7.33	5.96 (6.12)	0.11	0.72	H-2 → L (0.15) H → L + 1 (0.12) H-1 → L + 2 (0.11) H → L + 2 (0.10)	−0.04 (+0.12)
6	7.43	6.45 (6.35)	0.45	0.16	H-1 → L + 1 (0.33) H-3 → L (0.09) H → L (0.09)	7.92	6.01 (6.28)	0.36	0.30	H-1 → L + 1 (0.46)	−0.44 (−0.17)
7	7.82	6.46	0.61	0.45	H → L + 1 (0.18) H-2 → L − (0.17) H-1 → L (0.14) H-1 → L + 2 (0.11)	8.32	6.17 (6.44)	0.54	0.44	H → L + 1 (0.24) H-1 → L (0.18) H-1 → L + 2 (0.09) H-2 → L (0.07)	−0.29 (−0.02)
8	7.22	6.48	0.15	0.41	H-3 → L (0.20) H-1 → L + 1 (0.14) H-1, H ⇉ L + 1 (0.07) H → L + 2 (0.07)	7.55	6.42 (6.58)	0.10	0.19	H-3 → L (0.25) H ⇉ L (0.11) H-1, H ⇉ L (0.08) H → L + 1 (0.08) H-1 → L (0.06) H → L + 3 (0.06)	−0.06 (+0.10)
9	8.00	6.74	0.26	0.23	H → L + 2 (0.25) H-2 → L (0.20)	8.65	6.55 (6.78)	0.11	0.29	H-2 → L (0.28) H → L + 2 (0.25) H-3 → L + 1 (0.10)	−0.19 (+0.04)
10	8.88	7.65	0.81	0.32	H-2 → L + 1 (0.23) H-1 → L + 2 (0.14) H-3 → L + 1 (0.07)	9.53	7.40 (7.65)	0.88	0.36	H-2 → L + 1 (0.24) H-1 → L + 2 (0.17)	−0.25 (0.00)
11	8.44	7.67	0.15	0.25	H-3 → L + 1 (0.17) H-2 → L + 1 (0.12) H-1 → L + 3 (0.08) H-1 ⇉ L + 1, L (0.06) H-1 ⇉ L + 1 (0.06)	8.79	7.39 (7.61)	0.17	0.27	H-3 → L + 1 (0.12) H-1 → L + 3 (0.11) H-1 ⇉ L + 1 (0.07) H-1 → L + 1 (0.07) H → L + 2 (0.06) H-1 ⇉ L + 1, L (0.06)	−0.28 (−0.06)

Transition dipole moments (TDM) from the L_b and L_a states, the main configuration state functions (CSFs) of each state, and RASPT2 energy difference between the two levels of theory are also reported. *In parentheses are reported, when available, the corresponding experimental data, adopted from Ref. (22). †In parentheses the weights are also reported. ‡The energies are reported for a value of the imaginary level shift parameter of 0.2 a.u. In parentheses are reported energies for a value of 0.5 a.u. §RASPT2 energy difference between the reference data and $RASSCF(4.5|0.0|4.4)_{0.2}$. In parentheses is reported the energy difference with respect to $RASSCF(4.5|0.0|4.4)_{0.5}$. ¶TDM from the ground state. →, denote single excitation CSFs; ⇉, denote double excitation CSFs.

dynamics simulations. For that reason, a benchmark study of computationally less demanding active spaces. Our computations on benzene and phenol (24), indole (25), pyrimidines (26) and purines (27) show that a full π-valence active space (specifically, ten electrons in nine orbitals for indole) is required to capture the multiconfigurational nature of the higher lying excited states, while further truncation leads to the omission of excited states (and, hence, ESA signals in the simulated spectra). This requirement makes CAS (10,9) and its truncated counterpart RAS(4.5|0.0|4.4) which have two orders of magnitude less CSFs than the reference RAS(0.0|10.9|2.12) reasonable candidates for benchmarking. Indeed, using the aforementioned truncated active spaces, we obtain all ten excited states characterizing the visible spectrum of indole. Wavefunction analysis shows that all states preserve their leading configurations and can be straightforwardly mapped to the reference data set (see Table 1). However, in a series of benchmark studies, we have established that π-valence active spaces lacking virtual orbitals with higher orbital momenta lead to a pronounced nonuniform underestimation of the excitation energies. In agreement, the computationally feasible active spaces as CAS (10,9) and RAS(4.5|0.0|4.4) exhibit an average red-shift of 0.2 eV (~1500 cm^{-1}) from the reference values, with individual states diverging by up to 0.5 eV (~4000 cm^{-1}) (see the section describing the calibration of the indole molecule in Appendix S1 for further details).

The underestimation of the excitation energies with the cheaper theoretical approach can be mainly ascribed to the documented overestimation of the CASPT2 correction for ionic states when insufficiently large active spaces have been used (22–24). States that are considered ionic in the sense of valence bond

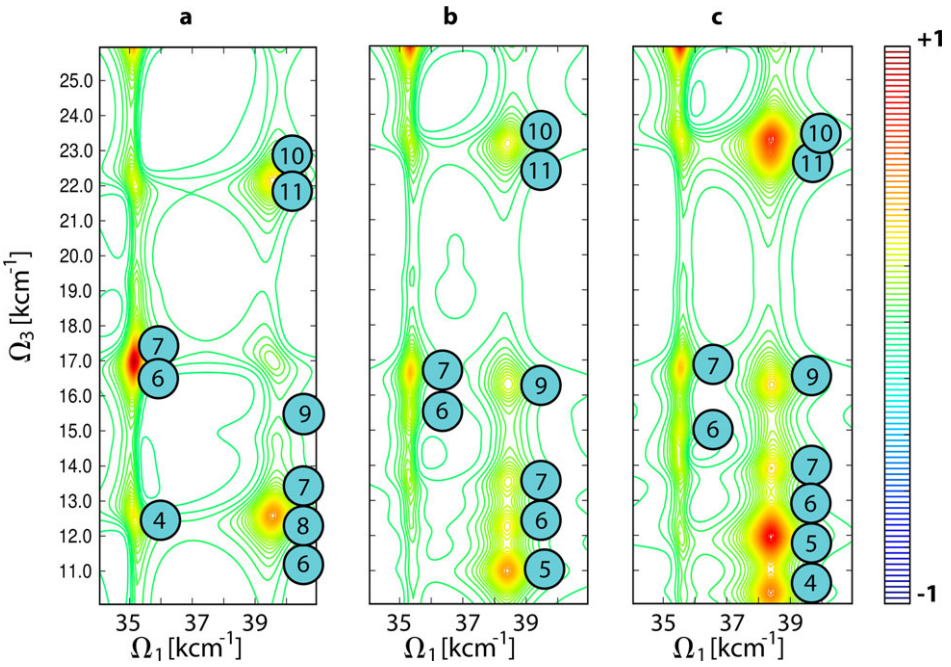

Figure 1. Near-UV pump VIS probe 2D electronic spectra of: (a) indole on the basis of RAS(0.0|10.9|2.12)$_{0.2}$ reference data; (b) indole on the basis of RAS(4.5|0.0|4.4)$_{0.5}$; (c) the noninteracting indole-phenol dimer on the basis of RAS(4.9|0.0|4.7)$_{0.4}$. Subscripts denote the value of the imaginary level shift parameter used in the computation. Labeling according to Table 1.

theory (*i.e.* with many ionic structures contributing to the overall wavefunction) exhibit a dynamic correlation correction much larger than that of (valence bond) covalent states, like the ground state, and their energies end up too low after CASPT2 correction (we remind the reader that the CASPT2 method is not variational). Following this trend, the L_a state, which is pronouncedly ionic (also recognized from the large permanent dipole moment (5.5 D) compared to the ground state (1.5 D), exhibits a red-shift, albeit "only" 0.27 eV, with respect to the reference data.

Another significant reason for the discrepancies between the reference data set and the results of the RAS(4.5|0.0|4.4) computation is the artificial state mixing in both, another consequence of the nonuniform description of covalent and ionic states at CASSCF level. Artificial state mixing is a subtle effect not linearly correlated to the employed level of theory but rather to the overall inability of CASSCF to describe states of different nature equally well and the density of states of the system under analysis. It is a well-known artifact of CASSCF that ionic and covalent states may swap order; however, in an unfortunate constellations, they may become energetically degenerate and mix their wavefunctions, thus distorting the computed transition energies and TDMs. We note that the multistate MS) CASPT2 approach was introduced precisely to correct problems arising from state mixing at CASSCF level and inherited by SS-CASPT2. While successfully un-mixing coupled states, in our experience MS-CASPT2 does not provide more accurate energies unless very large active spaces are used. The reason lies in the too large off-diagonal elements of the MS-CASPT2 Hamiltonian being often an order of magnitude larger than the threshold of 0.002 a.u. recommended in the seminal paper on MS-CASPT2 (32). An alternative way to distinguish artificial from authentic state mixing is to perform series of computations gradually expanding the active space as artificial mixing depends strongly on the correlation at CASSCF level.

In order to improve the results obtained with the low-cost method, the most straightforward strategy would be to include more dynamic correlation at CASSCF level, either through enlarging the active space or allowing for a larger number of simultaneous excitations in the construction of the CSFs. All these solutions would in principle improve their description; however, they imply an increased computational effort. An alternative solution, proven successful in the benchmarking of adenine (24), benzene and phenol (21), is the parametric use of the level shift parameters (real and imaginary) available in the RASPT2 protocol. Despite introduced with a different purpose, namely to cure intruder state problems (47), the real-shift parameter essentially comprises in its core an increase in the energy gap between the CASPT2 first-order wave function configurations and the reference CASSCF wave function (*i.e.* a nonzero value of the real-shift parameter introduces a positive real number in the denominators of the PT2 correction terms) with the effect of a nonuniform decrease in the CASPT2 correlation contribution for every state. Thus, once semi-empirically fitted with respect to the reference data set, level shift parameters can be utilized to damp the systematic overestimation of the dynamic correlation when the underlying RASSCF wavefunction is constructed from a limited active space. In Appendix S1 (see Tables S2–S5 and the accompanying text in the section describing the use of level shift parameters for semi-empirical correction of the overestimated dynamic correlation with limited active spaces), we thoroughly demonstrate that the use of real level shift introduces a significant improvement in the description of ionic states at the price of an acceptable overestimation of the energies of covalent states and explain why it is necessary to use values considerably higher than the ones (in the range of 0.1–0.2 a.u.) normally applied in conjunction with intruder states. Furthermore, in an exhaustive comparison, we show that the more robust imaginary level shift can be used instead of the real

one. The reader may also refer to previous discussions on the parametrization in Refs. (24,27). In a nutshell, a value for the imaginary level shift parameter between 0.50 and 0.55 a.u. gives the smallest average deviation with respect to the reference (0.08 eV, whereas we remind the reader that the reference data set was obtained with an imaginary level shift of 0.2 a.u., and in that case, the shift parameter was used in its classical scope, namely to cure intruder states problems).

In summary, the RAS(4.5|0.0|4.4)$_{0.5}$ level of theory (the subscript denotes the value of the imaginary level shift) reproduces the indole's electronic structure up to 8 eV with a satisfactory accuracy, however, at a substantially reduced computational cost compared to the reference data set (1 022 868 *vs* 2751 CSFs in the reference data and in the RAS(4.5|0.0|4.4)$_{0.5}$ computation, respectively). The 2DES spectrum (see Fig. 1c) qualitatively reproduces the above-described main features, although a 1000 cm^{-1} blue-shift is observed for most signals along the L_a trace, a consequence of the overstabilization of the L_a state at this level of theory by 0.15 eV compared to the reference. This leads to the appearance of an additional ESA, peak 5, in the low-energy part of the spectrum, which falls outside of the probing window in the reference spectrum. The L_b trace is satisfactorily reproduced but we note the reduced intensity of the 17 k cm^{-1} ESA. The overall agreement with the reference computation makes the method attractive for application to indole-containing systems. As a first application, we choose to compute 2D spectra of two Trp-cage conformations along the folding pathway. Trp-cage is composed by 20 amino acids, among which the aromatic Trp (containing an indole molecule) and Tyr (containing a phenol molecule). Consequently, the L_b/L_a excited states of the Trp's indole (referred to as Trp_{Lb}, Trp_{La}) and the L_b state of Tyr's phenol (referred to as Tyr_{Lb}) are expected to contribute to the 2D electronic spectrum when near-UV pump pulses are used. Before addressing the peptide, we tested the stability of indole's electronic structure computation in an ideal system composed by the indole and phenol molecules separated by about 10 Å under active space expansion, necessary to describe both chromophores in the aggregate, and under the consideration of additional states in the range of interest like mixed doubly excited and CT states (see the section on calibrating the noninteraction indole-phenol dimer in Appendix S1 for further details). Overall, we observe a good agreement with the results of RAS (4.5|0.0|4.4)$_{0.5}$ (Fig. 1c), noting a slight increase in the intensity of the L_a trace.

Application to Trp-cage

The application of 2DES as a tool for probing structural variations during protein folding has been recently investigated on a theoretical level by Mukamel and co-workers in two contributions (19,20) on the example of Trp-cage assessing the near- and far-UV spectral range for characteristic signatures. Relying on a classical molecular dynamics simulation of the folding process and clustering analysis for structure selection, a parametrized Frenkel exciton Hamiltonian (19,20) is built which employs the lowest excited states of the separated isolated chromophores: (1) the L_b and L_a states of indole and the L_b state of phenol in the case of near-UV 2DES; (2) the $n\pi^*$ and $\pi\pi^*$ state of the peptide bond in the case of far-UV 2DES. Thermal fluctuations of the peptide are incorporated through averaging over a large number of snapshots, interchromophore and interactions with the environment are accounted for on a grid using effective charges fitted from the electrostatic potentials of the isolated molecules computed at QM level. The far-UV study reveals a correlation between the signal complexity and the conformational entropy of the peptide and can consequently be used to evaluate the folding status of Trp-cage. According to the authors, near-UV 2DES captures important structural information regarding π-stacking and reveals exciton transport signatures.

Despite the valuable insights into the capabilities of 2DES provided by those studies, the adopted theoretical framework inherently implies several approximations. Most prominently, signals due to absorbing higher lying states in the chromophores, that is ESA features, have been completely neglected so far. Furthermore, the use of parameters obtained on the separated chromophores prevents the description of potentially important states localized on more chromophores such as CT states, whose presence and oscillator strength is expected to be strongly correlated with the distance and orientation of the interacting chromophores. To address these limitations of the exciton models, we computed the 2D spectrum of the Trp-cage using the SOS//QM/MM approach developed in our group for treating aggregates of few (two to three) monomers, thereby applying the computational methodology for indole benchmarked in the previous section. In this first application to indole-containing systems, we focus on two snapshots of the folded and unfolded Trp-cage with both aromatic residues being 5 Å apart and π-stacked in the folded snapshot and separated by more than 10 Å and noninteracting in the unfolded snapshot (see Fig. 2). We argue that ESA signatures bear a wealth of information that deserves detailed analysis. Furthermore, we highlight the new possibilities two-color 2DES (*i.e.* pumping in the near-UV, probing ESA in the VIS) offers compared to one-color experiments restricted to the near-UV and focusing mainly on the GSB features. Thus, the results presented in the following sections supplement the intriguing perspectives addressed by Mukamel and co-workers.

Comparison between unfolded and folded conformation based on single snapshots out of molecular dynamics simulations. Trp-cage snapshots were obtained from a molecular dynamics simulation of the folding process (19,20). The selection of the two structures has been dictated on the basis of the distance between the center of mass of the two chromophores, so to describe an

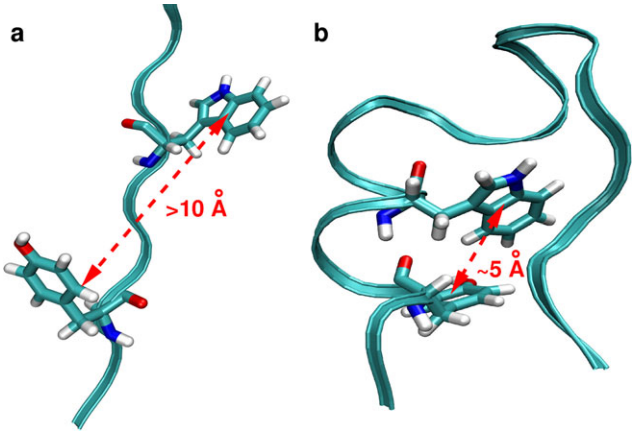

Figure 2. Structures of the unfolded (a) and folded (b) snapshots of the Trp-cage used in the simulations of the spectra.

unfolded structure in which the indole and phenol can be considered noninteracting (*i.e.* separated by a distance of the order of 10 Å), and a folded geometry where the two molecules are as close as possible (see Fig. 2). At a sufficiently reduced distance, the two chromophores are expected to interact and low-lying states involving both molecules (as for example CT states) should become visible in the corresponding 2D UV–vis spectra, constituting consequently a detectable fingerprint of the proximity of the two molecules. As the selected structures are the results of a molecular dynamics simulation, a QM/MM optimization of the structures has been operated so to refine the geometrical parameters (see Materials and Methods). Such a step is required because it has been documented that bond lengths and angles in the aromatic moiety can significantly influence the excited-state energies of the system and it is consequently important to describe them quantum mechanically (27). In the present contribution, we have employed the second-order Møller-Plesset (MP2) method for the ground-state structural refinement, because MP2 describes well weak dispersive interactions in stacked systems, although it is known to overestimate π-stacking (48). The MP2 optimization relaxes the indole and phenol intramolecular bond lengths to values comparable to the ones characterizing the two isolated molecules minimized with the same method (see Figure S4). The distance between the center of mass of the two chromophores is larger than 10 Å in the unfolded structure, while in the folded structure, it is equal to 5.21 Å, which is slightly larger than the distance of 5.097 Å for the unrefined snapshot.

Computing the excited states in the unfolded and folded structures, first important observation is the significantly decreased energy gap between the L_b (Trp_{Lb}) and L_a (Trp_{La}) states of 0.15 eV, being equal to 0.38 eV in the noninteracting indole-phenol dimer. Similar trend has been obtained in different polar solvents (49,50). The reported decrease in the energy gap is partially due to the more pronounced stabilization of the Trp_{La} in polar solvent compared to the Trp_{Lb}. This trend can be easily rationalized by comparing the permanent dipole moments of both states to that of the ground state (see Fig. 3). Trp_{La} has a characteristically strong dipole moment of 5.5 D, while Trp_{Lb} and ground state have comparable moments of 1.6 D and 1.5 D, respectively. The pronounced polarity of Trp_{La} makes it much more sensitive to the solvent. An additional source for the stabilization of Trp_{La} is the fact that the chromophore is linked to the protein through a methyl group which has been shown to further stabilize the L_a state (46). Thanks to the solvatochromic shift in water, the Trp_{La} absorption band falls on top of the L_b band of the phenol side chain of Tyr (Tyr_{Lb}) and is thus expected to cover it due to the much stronger oscillator strength of Trp. This constellation, although unfortunate, agrees with the LA spectra of Trp and Tyr (see Fig. 4a); hence, the selected snapshots can be considered representative of the situation in water.

In order to generate the 2D spectra, a reasonable estimation of the spectral dephasings, that is the reciprocal lifetimes with which the coherences between the ground-state and the near-UV manifold of states Trp_{Lb}, Trp_{La} and Tyr_{Lb}, created with the first pump pulse and evolving during the delay time $t1$, as well as the coherences created between the near-UV and far-UV manifolds responsible for the ESA features and evolving during the delay time $t3$, decay due to the coupling to the low-frequency degrees of the solvent as well due to the internal vibrational dynamics of the system and finite excited-state lifetimes. 2D electronic spectra

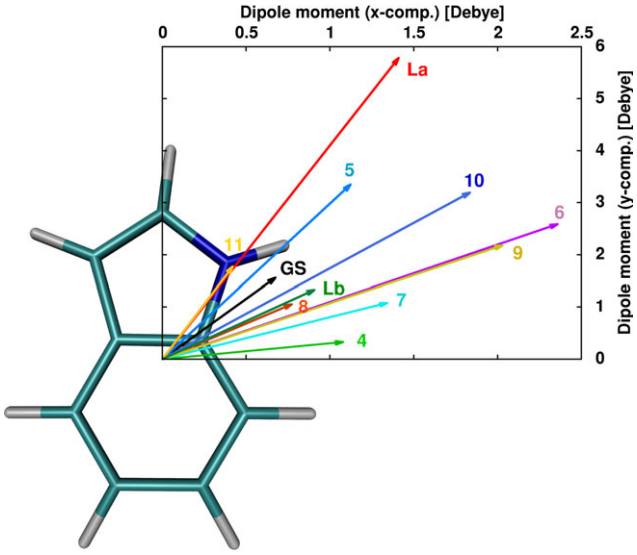

Figure 3. Permanent dipole moments of indole superimposed over the molecular structure. Origin (0,0) of the graph set at the origin of the Cartesian coordinates.

recorded in solution commonly feature broad unstructured peaks implying a very fast decay of the coherences so that any vibrational substructure is hidden underneath. We, therefore, chose to model signal broadening via homogeneous state-specific dephasing rates. This strategy has been applied also in previous 2DES simulations on the Trp-cage (19,20). Theoretical studies document that the L_a/L_b order upon excited-state geometrical relaxation depends on the polarity of the solvent. According to a CASSCF/CASPT2 study by Serrano-Andrés and co-workers, the states become nearly iso-energetic in water (46), while Drew and colleagues argue, based on ADC computations, that an inversion of the state order occurs (49). These findings stand out in comparison with the non-negligible gap of 0.2 eV both states have in gas-phase and nonpolar media with L_b being always the lower lying state, even after geometry relaxation (46,49,50). Due to the similar fate of both excited states, we use similar dephasing constants of 700 cm^{-1} (full-width-half-maximum, FWHM) for describing the ground-state-$Trp_{Lb/a}$ coherences. A somewhat smaller value of 400 cm^{-1} FWHM was used for the ground-state-Tyr_{Lb} coherence. Larger dephasing constants of 1000 cm^{-1} FWHM were applied to coherences between excited states in order to account for the shorter lifetimes of higher lying excited states. Applying the above-defined dephasing constants assumes immediate loss of correlation between the pump and probe frequency (loss of memory) and leads to spectral signals elongated along the probe frequency. A refined model, taking into account static disorder and thus introducing memory effects, is discussed in the next section.

Another important point that deserves attention is the necessity to increase the magnitude of the transition dipole moments of all transitions involving Tyr_{Lb} by a factor of 2 to account for the increase in the oscillator strength due to Herzberg-Teller vibrational coupling. While negligible for bright strong transitions, it is particularly important for (nearly) symmetry-forbidden ones.

The near-UV pump VIS probe nonchiral quasi-absorptive 2DES spectra of the unfolded and folded Trp-cage snapshots are

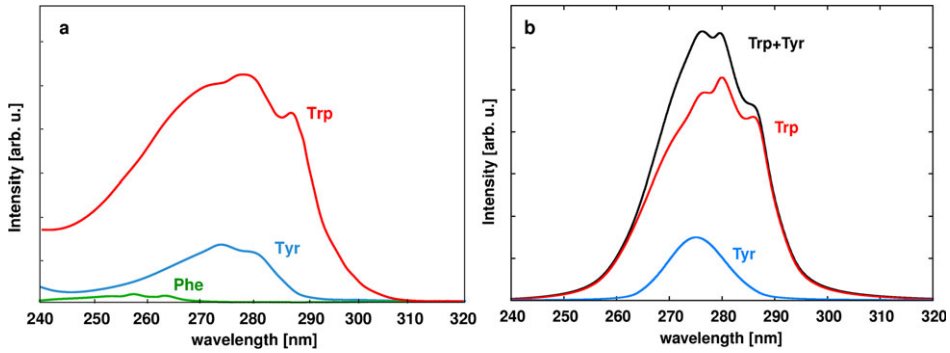

Figure 4. Experimental (a) and theoretical (b) linear absorption spectra of Trp (red), Tyr (blue), Phe (green) and a noninteracting Trp-Tyr dimer (black). The LA experimental spectra of Trp, Tyr, and Phe are based on the spectra reported in Ref. (58).

shown in Fig. 5. All parameters (energies and transition dipole moments) used to generate the spectra are listed in Table S11 in Appendix S1. Due to the energetic proximity of all three transitions, only the Trp_{La} band (Ω1-frequency ~36.3 k cm^{-1}) is clearly recognized with multiple characteristic signatures in the low-energy window up to 17 k cm^{-1}, as well as a strong contribution around 24 k cm^{-1}. The Trp_{Lb} band (Ω1-frequency ~35.5 cm^{-1}) shows an intense fingerprint ESA around 25 k cm^{-1} and a weaker ESA around 17 k cm^{-1} The only resolvable ESA which can be associated with Tyr_{Lb} belongs to an excitation of the 1E2 g state which appears as a weak signal at 19.5 k cm^{-1}. As anticipated from the gas-phase spectra, a broad signal-free region between 18 and 24 k cm^{-1} is evident in the unfolded Trp-cage. To our delight, we encounter a number of CT and mixed states signatures that absorb exactly in this range in the folded Trp-cage, giving rise to a broad unstructured ESA signal. While being two orders of magnitude less intense than the fingerprint signals of indole, they are clearly recognized and can be selectively enhanced with probe pulses centered at 21 k cm^{-1} having FWHM bandwidths of 4 k cm^{-1} or less.

Comparison between unfolded and folded conformation accounting for static disorder. While the 2DES spectra based on single snapshots give a promising first insights into the capabilities of multidimensional electronic spectroscopy, it must be recognized that the spectra lack one significant feature which could obscure the fingerprints and, thus, render questionable the applicability of 2DES for the purposes addressed in the present paper, namely a proper description of the static disorder in the system. Looking at the LA spectra of Trp and Tyr (see Fig. 4a), it immediately stands to one's attention that the absorption bands of the states are very broad and overlap. As discussed above, Trp_{La} shows a particularly strong dependence on the solvent's polarity and its energy fluctuates between 255 nm (4.85 eV) and 290 nm (4.30 eV) (here we neglect contributions to the high-energy range above 260 nm, presumably coming from L_a overtones). Similarly, Tyr_{Lb} shows a broad absorption window ranging from 265 nm (4.70 eV) to 285 nm (4.35 eV). In contrast, Trp_{Lb} characterizes with a rather narrow band peaking at 287 nm (4.32 eV) with just about 0.10 eV broadening. A justified question is how the higher lying excited states are affected by the polarity of the solvent. It is, therefore, imperative for producing realistic spectral line shapes to incorporate the inhomogeneous peak broadening due to solvation and conformational freedom of the peptide (interactions of the aromatic residues with the peptide skeleton in the unfolded system and chromophore–chromophore interactions in the folded snapshots). Mukamel and co-workers demonstrated how accounting for these interactions within the exciton

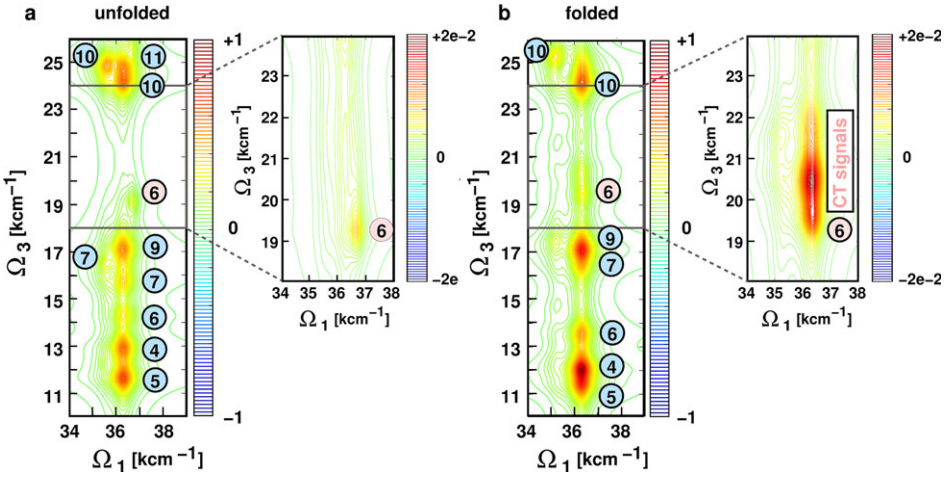

Figure 5. 2D UV pump VIS probe spectra of the unfolded (a) and folded (b) Trp-cage snapshots extracted from the molecular dynamics simulations. Excited-state absorptions (ESA) associated with the indole side chain of Trp (light blue) or with the phenol side chain of Tyr (light red) are labeled. Spectral region of 18–24 k cm^{-1} is enhanced to highlight the weak signatures of charge-transfer states emerging in the stacked conformation.

Hamiltonian model with electrostatic fluctuations (EHEF) (51) leads to a broad GSB signal in nonchiral Trp-cage spectra elongated along the diagonal at waiting times $t2$ close to zero making discrimination of contribution from the individual states hardly possible (19,20). While accurate conformational analysis within the SOS//QM/MM framework is beyond the scope of this paper, we incorporate static disorder in our spectra following the parametrization procedure outlined in the next paragraph.

From the LA spectra of Trp and Tyr (see Fig. 4a), we estimate roughly the broadening of the absorption bands of Trp_{Lb}, Trp_{La} and Tyr_{Lb} to 1.0 k cm^{-1} (0.12 eV), 4.5 k cm^{-1} (0.54 eV) and 3.0 k cm^{-1} (0.36 eV) and associate each value with the distribution of transition energies caused by environmental fluctuations. Thus, due to the thermal motion in the solvated system, an instantaneous solvent arrangement around the chromophores varies the energies of the absorbing states Trp_{Lb}, Trp_{La} and Tyr_{Lb} around their maxima 287 nm (34.8 k cm^{-1}), 272 nm (36.8 k cm^{-1}) and 275 nm (36.3 k cm^{-1}) within the spectral windows defined above. Homogeneous signal broadening due to vibrationally induced energy fluctuations (spectral diffusion) and excited-state lifetimes is accounted for by constant dephasing parameters. Following the considerations of the previous section, we recognize that in the red-edge of the LA spectrum, Trp_{La} and Trp_{Lb} are nearly iso-energetic already in the FC region; hence, an inversion of the gas-phase state order is expected to occur upon geometry relaxation in the excited states. Thus, the ground-state Trp_{Lb} coherence is expected to decay on a sub-100 fs timescale due to population relaxation from Trp_{Lb} to Trp_{La}. Therefore, in the red-edge of the spectrum, we use dephasing constants 1000 cm^{-1} for Trp_{Lb} and 400 cm^{-1} for Trp_{La}. On the contrary, in the blue-edge of the LA spectrum, the picture resembles the one in vacuo as the L_a and L_b states are energetically well separated. After excitation and initial relaxation, population transfer is, thus, expected to occur from Trp_{La} to Trp_{Lb}. Thus, in the blue-edge of the spectrum, we use dephasing constants 400 cm^{-1} for Trp_{Lb} and 1000 cm^{-1} for Trp_{La}. For the center of the LA spectrum, the picture is similar to the one in the single snapshots, discussed earlier; hence, intermediate dephasing values of 700 cm^{-1} are used for both states. A unique constant of 400 cm^{-1} is used for Tyr_{Lb}. All coherences between excited states are set to 1000 cm^{-1}. While the spectra presented below rely on the aforementioned assumption of transition energy distribution induced by environmental fluctuations, we note that estimation of the fluctuations of the Trp_{La}, Trp_{Lb} and Tyr_{Lb} transition energies in Trp-cage based on classical molecular dynamics simulations made by Mukamel and co-workers (17) suggests a narrower distribution of oscillators. For example, while confirmed to have the broadest distribution pattern, Trp_{La} is predicted to span a 2 k cm^{-1} range. In the section of Appendix S1 describing an alternative set of parameters accounting for static disorder we present LA and 2DES spectra based on these findings.

A total of 100 individual LA computations were performed in which the energies of Trp_{Lb}, Trp_{La} and Tyr_{Lb} were scanned in equidistant steps in order to cover the estimated bandwidths (*i.e.* 1.0, 4.5, 3.0 k cm^{-1}), whereas one of the three sets of dephasing parameters discussed above was used in each computation depending on whether the energies are closer to the blue-edge, center or red-edge of the LA spectrum. Subsequently, the 100 spectra were added, thereby weighting each one according to a Gaussian distribution for each of the singly excited states peaking at the maxima (*i.e.* 287 nm for Trp_{Lb}, 272 nm for Trp_{La} and 275 nm for Tyr_{Lb}) with a standard deviation proportional to the estimated bandwidths. Thus, "snapshots" (*i.e.* computations in which the scanned energies are) closer to the experimental peak maxima contribute more to the LA spectrum.

It should be pointed out that we assume that the energies of the three absorbing states are correlated, that is a solvent arrangement favoring (stabilizing) Trp_{La} also favors the remaining two states. This is justified for Trp as Trp_{Lb} and Trp_{La} have their permanent dipole moments pointing in the same direction (see Fig. 3). The correlation between the states of Trp and Tyr is also justified for the unfolded systems where the spectra of the noninteracting monomers can be also computed separately. In the case of the folded Trp-cage, we do not exclude that this correlation can be violated due to, albeit weak, chromophore–chromophore interactions and the fact that in the folded peptide, the chromophores may not be fully exposed to the environment.

The LA spectra of water-solvated Trp, Tyr and of a mixture of solvated noninteracting Trp-Tyr dimer are presented in Fig. 4b. We encounter a surprisingly good agreement in the energy range 265–300 nm considering the simplicity of the applied model. The nonzero cross section in the high-energy window above 265 nm is most probably traced back to Trp_{La} overtones which are not considered.

Next, we estimate the dependence of the higher lying excited states on the solvation. We assume that the energies obtained for the indole-phenol dimer in vacuo and for the Trp-cage in solution (noninteracting dimer), computed with the same methodology $RAS(4.9|0.0|4.7)_{0.4}$, correspond to two solvent arrangements. The gas-phase one corresponds to an unfavorable water arrangement which blue-shifts the absorption energy of Trp_{La} in the blue-edge of the LA spectrum up to 4.76 eV. The relaxed unfolded snapshot corresponds to a representative favorable water orientation which stabilizes Trp_{La} down to 4.50 eV, thus, close to the center of the LA spectrum. Using these two reference points, we apply a linear fit to all excited-state energies which allows to extrapolate their energies throughout the whole LA window 250–300 nm (see Fig. 6). In support, we correlate the estimated dependencies to the dipole moments (see Fig. 3). As elucidated earlier, due to its large dipole moment (5.5 D), Trp_{La} exhibits the strongest dependence on the solvation. States 5, 9 and 10 of Trp show notable stabilization upon solvation in agreement with their intermediate dipole moments. Only state 6 does not exhibit the stabilization expected due to its dipole moment strength. On the other hand, states L_b, 7 and 11 of Trp show negligible dependence on the solvation, while states 4 and 8 are destabilized upon solvation as their dipole moments are smaller than that of the ground state.

A total of 100 individual near-UV pump VIS probe nonchiral quasi-absorptive 2D spectra were computed, representative of the unfolded and folded ensembles in which the energies of all states were scanned in equidistant steps. The energies of the CT and mixed states emerging in the folded (stacked) conformation were not modified. Further, we assume that the transition dipole moments of the unfolded and folded geometries, discussed in the previous section, are representative for the entire ensemble. As done in the treatment of the single snapshots, the magnitude of the transition dipole moments of all transitions involving Tyr_{Lb} was increased by a factor of 2. Subsequently, the 100 spectra were added, thereby weighting each according to a Gaussian distribution for each of the singly excited states as elucidated above.

We note that the adopted procedure does not take into account effects like secondary splitting that stacking could have on the excited-state energies.

The 2D spectra of the unfolded and folded Trp-cage ensembles for a waiting time $t2 = 0$ are shown in Fig. 7. It can be seen that the ESA signals are no longer circular but rather elongated. This is related to the relative stabilization of the singly and higher lying excited states in water. As the Trp_{La} state exhibits the most pronounced stabilization, all ESA peaks along its trace (now spanning the pump-range 35–38 k cm^{-1}) exhibit similar and pronounced tilt downward (denoted with dashed lines in Fig. 7). The more pronounced the covalent character of the excited states, the stronger the tilt (*e.g.* peak 4 *vs* peak 5) along the Trp_{La} trace. Similar is the line shape of the only visible signal of Tyr absorbing near 19 k cm^{-1} as the excited state associated with the ESA (state 6) is of covalent nature. On the contrary, states along the Trp_{Lb} trace, distinguishable from the Trp_{La} only through its fingerprint signals around 18 k cm^{-1} (peak 9) and 26 k cm^{-1} (peak 10), are tilted upward. We emphasize that the line shapes depend strongly on the relative magnitudes of homogeneous and inhomogeneous broadening. In the section of Appendix S1 describing an alternative set of parameters accounting for static disorder we demonstrate how larger constant dephasing parameters (implying broad unstructured signals due to the short correlation times and immediate decay of the generated coherences) accompanied by a narrow distribution of transition energies (implying a less pronounced effect of the fluctuating environment on the transition energies of the system) can conceal the information about the nature of the spectroscopically resolved electronic states imprinted in the above-mentioned tilt. Peak broadening makes the discrimination of the individual signals a hard task, particularly in the low-energy window between 11 and 15 k cm^{-1}. The signal enhancement in the 18–24 k cm^{-1} window in the folded Trp-cage, a clear signature of π-stacking, is still observed, whereas it deserves attention that upon introducing static disorder in the spectra, the fingerprint signals of the interchromophore interaction become more intense than reported previously on the basis of the single snapshots and are now only three times less intense than the local ESA of the monomers.

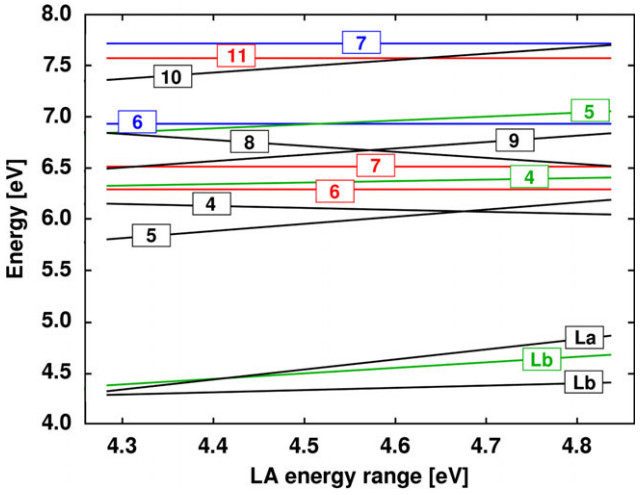

Figure 6. Electronic structure of Trp and Tyr as a function of instantaneous solvation in the polar solvent water. Red-edge (295 nm) and blue-edge (260 nm) correspond to solvent arrangements favoring and disfavoring, respectively, ionic configurations. Trends are determined through a linear interpolation between the energies of the unfolded solvated Trp-cage snapshot refined at QM/MM level and of the noninteracting indole-phenol dimer, whereas the electronic structure of the noninteracting dimer is assumed to be representative for an unfavorable water arrangement. Ionic states (*e.g.* Trp_{La}, Trp_5) show a pronounced dependence on the solvation, covalent states are affected to a less extent (*e.g.* Trp_{Lb}), not affected (*e.g.* Trp_7) or show opposite trend (*e.g.* Trp_4). Legend: Trp and Tyr excited states exhibiting solvatochromic shifts are given in black and green, respectively. Trp and Tyr excited state which are not affected by the solvent are given in red and blue, respectively. Labeling follows notation from Tables 1 and S1.

CONCLUSION

The present contribution continues the series of papers devoted to the benchmarking of low-cost multiconfigurational RASSCF/

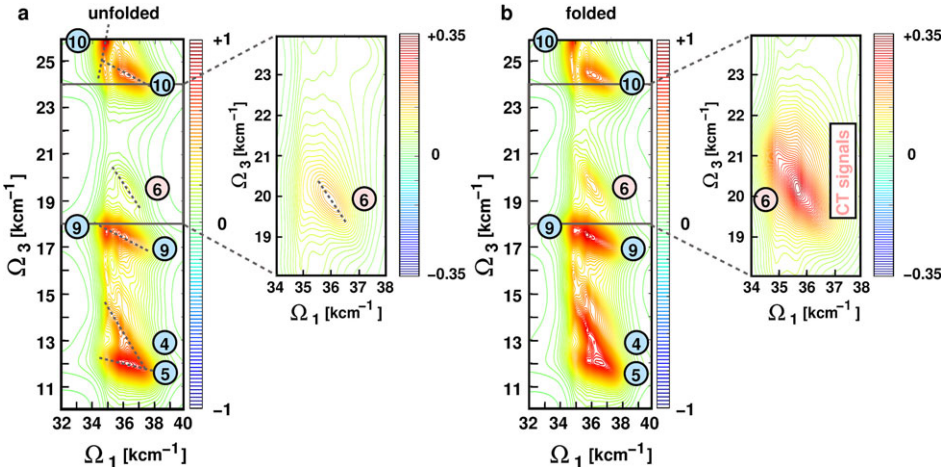

Figure 7. 2D UV pump VIS probe spectra of the unfolded (a) and folded (b) Trp-cage considering peak broadening due to static disorder. Excited-state absorptions (ESA) associated with the indole side chain of Trp (light blue) or with the phenol side chain of Tyr (light red) are labeled. Dashed lines indicate the tilt of the ESA and highlight the correlation between the solvatochromic shifts of the higher lying excited state and the singly excited states Trp_{Lb}, Trp_{La} and Tyr_{Lb}. Spectral region of 18–24 k cm^{-1} is enhanced to highlight the weak signatures of charge-transfer states emerging in the stacked conformation.

RASPT2 schemes for computing the electronic structure of some biologically relevant aromatic systems in nucleotides (26,27) and amino acids (24,25). Following our recent study on phenol and benzene, the aromatic side chains of Tyr and Phe, here we focus on indole, the chromophore of Trp. Its two states absorbing in the near-UV, of essentially opposite character—L_b is covalent, apolar, with a low oscillator strength, L_a is ionic, polar and bright—present a more intricate challenge compared to benzene and phenol that characterize with a single covalent absorption in the near-UV. Understandably, more compromises had to be made in the calibration procedure. Nevertheless, the RAS(4.5|0.0|4.4)$_{0.5}$ level of theory, implying a full π-valence active space with a restriction on the maximal number of simultaneously excited electrons during the generation of the configurational space, followed by a single state RASPT2 energy correction utilizing an imaginary level shift parameter of 0.5 a.u., shows a reasonable agreement with the reference data for states below 8 eV at significantly reduced computational effort. This computational strategy, applied recently also to adenine, facilitates the simulation of near-UV pump VIS probe (*i.e.* two-color) 2D electronic spectra of aggregates of several (two, three) chromophores as well the processing of multiple snapshots coming, for example, from molecular dynamics simulations. We analyzed critically the effect of poor dynamic correction description at RASSCF level leading to spurious state mixing which can introduce subtle errors in the simulations, discussed their detection and alleviation.

Beyond the purely technical side of the report, we characterized fingerprint signals of Trp for both its states Trp_{Lb} and Trp_{La}. Selectively addressing these fingerprints could allow to monitor the fate of the L_b and L_a states after excitation. Furthermore, we bring forward the idea of applying two-color 2DES to detect ESA signatures of chromophore–chromophore interactions, providing also an indirect way to detect Tyr and Phe, whose absorption bands remain hidden under the intense envelope of Trp in linear absorption experiments. A signal-free spectral window between 18 and 24 k cm^{-1} along the Trp traces seems to facilitate this undertaking. In a first application, the small Trp- and Tyr-containing peptide Trp-cage is used to prove the concept, thereby resolving weak CT signals in a folded Trp-cage snapshot having both chromophores about 5 Å apart. We recognize that near-UV pump VIS probe 2DES is not suitable for tracking the folding process in Trp-cage as already the geometry with the shortest interchromophore distance along the folding pathway shows only weak signatures of stacking. Nevertheless, we can claim with certainty that an experimentally detected enhancement of the ESA signal in peptides in the otherwise signal-free 18–24 k cm^{-1} spectral window should be regarded as a clear signature of chromophore–chromophore proximity of 5 Å or below.

Finally, we demonstrated that when static disorder is properly accounted for 2DES line shapes bear valuable information about the nature of the spectroscopically resolved electronic states and that the tilt of the ESA signals can be correlated to the polarity of the states. We remind that this information can be recovered experimentally only with ultra-short pulses for waiting times $t2$ closer to zero as solvent relaxation dynamics would quickly lead to loss of memory and broadening of the ESA. The spectra were simulated using many approximations; however, the outlined cost-efficient computational protocol facilitates the applications to a large number of snapshots, thus explicitly incorporating oscillating transition energies and dipole moments under environment fluctuations in the simulation. Furthermore, our approach can be interfaced with more sophisticated protocols that go beyond homogeneous dephasing rate constants to describe electronic dephasing and spectral diffusion (52–56). Following a recent proposal of Mukamel and co-workers that 2D double coherence spectroscopy can be used to probe electron correlation (57), this marks another fascinating feature of 2DES that can be utilized to gain insights into the electronic structure of complex molecules.

Acknowledgements—A.G. and A.N. acknowledge Jin Jiang for providing the Trp-cage molecular dynamics simulation and for his helpful comments on its analysis. A.G. acknowledges the support of Project No. CTQ2014-58624-P of the Spanish MINECO (Ministerio de Economía y Competitividad) and Project. No. GV2015-057 of the Generalitat Valenciana. S. M. gratefully acknowledges the support of the National Science Foundation (grant CHE-1361516). M.G. acknowledges support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198) and of the French Agence Nationale de la Recherche (FEMTO-2DNA, ANR-15-CE29-0010).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. (a) Details on the working equations for simulation of 2DES; (b) additional information on the calibration of the reduced level of theory for the description of indole's excited states (Figure S1, Table S1); (c) 2DES spectra with an alternative set of parameters accounting for homogeneous broadening and static disorder (Figures S2 and S3); (d) benchmarking of the real and imaginary shift parameters against the reference data set for benzene and indole (Tables S2–S5); (e) electronic structures of isolated (gas-phase) indole and phenol computed at CASSCF/CASPT2 level with various parameters (Tables S6–S9); electronic structures of solvated (as a part of Trp-cage) indole and phenol computed at SA-30-RAS(4.9|0.0|4.7)/SS-CASPT2 ANO-L[431,21] level with imaginary shift parameter 0.4 a.u. (Tables S10 and S11); and (f) characteristic bond lengths of the indole geometries (gas-phase and inside Trp-cage) used in this study.

REFERENCES

1. Kennedy, D. F., M. Crisma, C. Toniolo and D. Chapman (1991) Studies of peptides forming 310-helices and α-helices and β-bend ribbon structures in organic solution and in model biomembranes by Fourier-transform infrared-spectroscopy. *Biochemistry* **30**, 6541–6548.
2. Keiderling, T. A. (2002) Protein and peptide secondary structure and conformational determination with vibrational circular dichroism. *Opin. Chem. Biol.* **6**, 682–688.
3. Johnson, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structure. *Proteins* **35**, 307–312.
4. Cho, M. H. (2008) Coherent two-dimensional optical spectroscopy. *Chem. Rev.* **108**, 1331–1418.
5. Mukamel, S. (2000) Multidimensional Femtosecond correlation spectroscopies of electronic and vibrational excitations. *Annu. Rev. Phys. Chem.* **51**, 691–729.
6. Mukamel, S. and D. Abramavicius (2004) Many-Body Approaches for simulating coherent nonlinear spectroscopies of electronic and vibrational excitons. *Chem. Rev.* **104**, 2073–2098.
7. Mukamel, S. (1995) *Principles of Nonlinear Optical Spectroscopy*. Oxford University Press, New York, NY.

8. Aue, W. P., E. Bartholdi and R. R. Ernst (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229.
9. Hamm, P. and M. Zanni (2011) *Concepts and Methods of 2D Infrared Spectroscopy*. Cambridge University Press, Cambridge, UK.
10. Buchanan, L. E., J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. de Pablo, J. L. Skinner and M. T. Zanni (2014) Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc. Natl Acad. Sci. USA*, **111**, 5796–5801.
11. Tucker, M. J., M. Abdo, J. R. Courter, J. Chen, S. P. Brown, A. B. Smith III and R. M. Hochstrasser (2013) Nonequilibrium dynamics of helix reorganization observed by transient 2DIR spectroscopy. *Proc. Natl Acad. Sci. USA* **110**, 17314–17319.
12. Kiefer, L. M. and K. J. Kubarych (2015) Solvent-dependent dynamics of a series of rhenium photoactivated catalysts measured with ultrafast 2DIR. *J. Phys. Chem. A* **119**, 959–965.
13. Zhuang, W., T. Hayashi and S. Mukamel (2009) Coherent multidimensional vibrational spectroscopy of biomolecules: Concepts, simulations, and challenges. *Angew. Chem. Int. Ed.* **48**, 3750.
14. Duan, H. G., V. I. Prokhorenko, R. Cogdell, K. Ashraf, A. L. Stevens, M. Thorwart and R. J. D. Miller (2016) Nature does not rely on long-lived electronic quantum coherence for photosynthetic energy transfer. arXiv:1610.08425.
15. Duan, H. G., A. L. Stevens, P. Nalbach, M. Thorwart, V. I. Prokhorenko and R. J. D. Miller (2015) Two-dimensional electronic spectroscopy of light-harvesting complex II at ambient temperature: A joint experimental and theoretical study. *J. Phys. Chem. B* **119**, 12017–12027.
16. Duan, H. G., P. Nalbach, V. I. Prokhorenko, S. Mukamel and M. Thorwart (2015) On the origin of oscillations in two-dimensional spectra of excitonically-coupled molecular systems. *New J. Phys.* **17**, 072002.
17. Tseng, C.-h., P. Sándor, M. Kotur, T. Weinacht and S. Matsika (2012) Two-dimensional Fourier transform spectroscopy of adenine and uracil using shaped ultrafast laser pulses in the deep UV. *J. Phys. Chem. A* **116**, 2654–2661.
18. West, B. A. and A. M. Moran (2012) Two-dimensional electronic spectroscopy in the ultraviolet wavelength range. *J. Phys. Chem. Lett.* **3**, 2575–2581.
19. Jiang, J., Z. Lai, J. Wang and S. Mukamel (2014) Signatures of the protein folding pathway in two-dimensional ultraviolet spectroscopy. *J. Phys. Chem. Lett.* **5**, 1341–1346.
20. Li, J., M. Deng, D. V. Voronine, S. Mukamel and J. Jiang (2015) Two-dimensional near ultraviolet (2DNUV) spectroscopic probe of structural-dependent exciton dynamics in a protein. *J. Phys. Chem. B* **119**, 1314–1322.
21. Li, Q., A. Giussani, J. Segarra-Martí, A. Nenov, I. Rivalta, A. A. Voityuk, S. Mukamel, D. Roca-Sanjuán, M. Garavelli and L. Blancafort (2016) Multiple decay mechanisms and 2D-UV spectroscopic fingerprints of singlet excited solvated adenine-uracil monophosphate. *Chem. Eur. J.* **22**, 7497–7509.
22. Nenov, A., J. Segarra-Martí, A. Giussani, I. Conti, I. Rivalta, E. Dumont, V. K. Jaiswal, S. Altavilla, S. Mukamel and M. Garavelli (2015) Probing deactivation pathways of DNA nucleobases by two-dimensional electronic spectroscopy: Simulations form first principles. *Faraday Discuss.* **177**, 345.
23. Nenov, A., I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli (2014) Disentangling peptide configurations via two-dimensional electronic spectroscopy: Ab initio simulations beyond the Frenkel exciton hamiltonian. *J. Phys. Chem. Lett.* **5**, 767–771.
24. Nenov, A., S. Mukamel, M. Garavelli and I. Rivalta (2015) Two-dimensional electronic spectroscopy of benzene, phenol, and their dimer: An efficient first-principles simulation protocol. *J. Chem. Theory Comput.* **11**, 3755–3771.
25. Nenov, A., I. Rivalta, S. Mukamel and M. Garavelli (2014) Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles. *Comput. Theor. Chem.* **1040–1041**, 295–303.
26. Giussani, A., J. Segarra-Martí, A. Nenov, I. Rivalta, A. Tolomelli, S. Mukamel and M. Garavelli (2016) Spectroscopic fingerprints of DNA/RNA pyrimidine nucleobases in third-order nonlinear electronic spectra. *Theor. Chem. Acc.* **135**, 121.
27. Nenov, A., A. Giussani, J. Segarra-Marti, V. K. Jaiswal, I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli (2015) Modelling the high energy electronic state manifold of adenine: Calibration for nonlinear electronic spectroscopy. *J. Chem. Phys.* **142**, 212443–212459.
28. Rivalta, I., A. Nenov, G. Cerullo, S. Mukamel and M. Garavelli (2014) Ab initio simulations of two-dimensional electronic spectra: The SOS//QM/MM approach. *Int. J. Quantum Chem.* **114**, 85–93.
29. Neidigh, J., R. M. Fesinmeyer and N. H. Andersen (2002) Designing A20-residue protein. *Nat. Struct. Biol.* **9**, 425–430.
30. Malmqvist, P. Å., A. Rendell and B. O. Roos (1990) The restricted active space self-consistent-field method, implementation with a spit graph unitary group approach. *J. Phys. Chem.* **94**, 5477–5482.
31. Malmqvist, P. Å., K. Pierloot, A. R. M. Shahi, C. J. Cramer and L. Gagliardi (2008) The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO_2 and Cu_2O_2 systems. *J Chem. Phys.* **128**, 204109.
32. Serrano-Andrés, L., M. Merchán and R. Lindh (2005) Computation of conical intersections by using perturbation techniques. *J. Chem. Phys.* **122**, 104107.
33. Ghigo, G., B. O. Roos and P.-Å. Malmqvist (2004) A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chem. Phys. Lett.* **396**, 142–149.
34. Giussani, A., R. Pou-Amérigo, L. Serrano-Andrés, A. Freire-Corbacho, C. Martínez-García, M. A. P. Fernández, M. Sarakha, M. L. Canle and J. A. Santaballa (2013) Combined theoretical and experimental study of the photophysics of asulam. *J. Phys. Chem. A* **117**, 2125.
35. Zobel, J. P., J. J. Nogueira and L. Gonzalez (2017) The IPEA dilemma in CASPT2. *Chem. Sci.* **8**, 1482–1499.
36. Forsberg, N. and P.-Å. Malmqvist (1997) Multiconfiguration perturbation theory with imaginary level shift. *Chem. Phys. Lett.* **274**, 196–204.
37. Widmark, P. O., B. J. Persson and B. O. Roos (1991) Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. II. Second row atoms. *Theor. Chim. Acta* **79**, 419–432.
38. Aquilante, F., T. B. Pedersen and R. Lindh (2009) Density fitting with auxiliary basis sets from Cholesky decompositions. *Theor. Chem. Acc.* **124**, 1–10.
39. Aquilante, F., J. Autschbach, R. Carlson, L. Chibotaru, M. G. Delcey, L. De Vico, I. Fernández Galvan, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. Hoyer, G. Li Manni, H. Lischka, D. Ma, P.-Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. Vysotskiy, O. Weingart, F. Zapata and R. Lindh (2016) Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* **37**, 506–541.
40. Altoè, P., M. Stenta, A. Bottoni and M. Garavelli (2007) A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. *Theor. Chem. Acc.* **118**, 219–240.
41. Salomon-Ferrer, R., D. A. Case and R. C. Walker (2013) An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198–210.
42. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, C. Y. Al-Laham, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople (2003). Gaussian Inc., Pittsburgh, PA.

43. Lin, H. and D. G. Truhlar (2006) QM/MM: What have we learned, where are we, and where do we go from here. *Theor. Chem. Acc.* **117**, 185–199.
44. Abramavicius, D., B. Palmieri, D. V. Voronine, F. Sanda and S. Mukamel (2009) Coherent multidimensional optical spectroscopy of excitons in molecular aggregates, quasiparticle versus supermolecule perspectives. *Chem. Rev.* **109**, 2350–2408.
45. Ilich, P. (1987) Lowest singlet states in isolated indoles. *Can. J. Spectrosc.* **67**, 3274.
46. Serrano-Andrés, L. and B. O. Roos (1996) Theoretical study of the absorption and emission spectra of indole in the gas phase and in a solvent. *J. Am. Chem. Soc.* **118**, 185–195.
47. Roos, B. O. and K. Anderson (1995) Multiconfigurational perturbation theory with level shift – the Cr2 potential revisited. *Chem. Phys. Lett.* **245**, 215.
48. Černý, J. and P. Hobza (2007) Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **9**, 5291–5303.
49. Brisker-Klaiman, D. and A. Dreuw (2015) Explaining level inversion of the La and Lb states of indole and indole derivatives in polar solvents. *ChemPhysChem* **16**, 1695–1702.
50. Giussani, A., M. Merchán, D. Roca-Sanjuán and R. Lindh (2011) Essential on the photophysics and photochemistry of the indole chromophore by using a totally unconstrained theoretical approach. *J. Chem. Theory Comput.* **7**, 4088–4096.
51. Jiang, J. and S. Mukamel (2011) Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: A first principles study. *Phys. Chem. Chem. Phys.* **13**, 2394–2400.
52. van der Vegte, C. P., A. G. Dijkstra, J. Knoester and T. L. C. Jansen (2013) Calculating two-dimensional spectra with the mixed quantum-classical Ehrenfest method. *J. Phys. Chem. A* **117**, 5970–5980.
53. Tempelaar, R., C. P. van der Vegte, J. Knoester and T. L. C. Jansen (2013) Surface hopping modeling of two-dimensional spectra. *J. Chem. Phys.* **138**, 164106-1–164106-10.
54. Nenov, A., A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel and M. Garavelli (2015) Spectral lineshapes in nonlinear electronic spectroscopy. *Phys. Chem. Chem. Phys.* **17**, 30925–30936.
55. Duan, H. G., A. G. Dijkstra, P. Nalbach and M. Thorwart (2015) Efficient tool to calculate two-dimensional optical spectra for photoactive molecular complexes. *Phys. Rev. E* **92**, 042708.
56. Richter, M. and B. P. Fingerhut (2016) Simulation of multi-dimensional in the optical domain: Quantum-classical feedback in nonlinear exciton propagation. *J. Chem. Theory Comput.* **12**, 3284–3292.
57. Li, Z., D. Abramavicius and S. Mukamel (2008) Probing electron correlations in molecules by two-dimensional coherent optical spectroscopy. *J. Am. Chem. Soc.* **130**, 3509–3515.
58. Lakowicz, J. R. (2006) Protein fluorescence. In *Principles of Fluorescence Spectroscopy*, 3rd edn (Edited by J. R. Lakowicz), pp. 529–575. Springer, US.