

Two-Dimensional Electronic Spectroscopy of Benzene, Phenol, and Their Dimer: An Efficient First-Principles Simulation Protocol

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

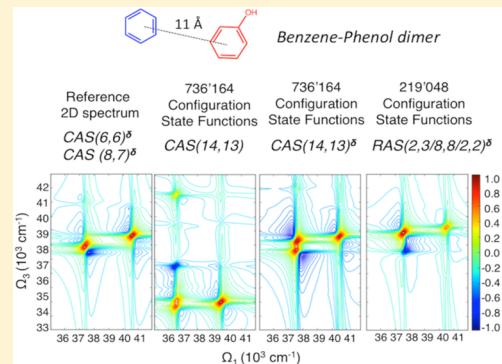
[†]Dipartimento di Chimica “G. Ciamician”, Università di Bologna, Via F. Selmi 2, 40126 Bologna, Italy

[‡]Department of Chemistry, University of California, Irvine, California 92697-2025, United States

[§]Université de Lyon, CNRS, Laboratoire de Chimie, École Normale Supérieure de Lyon, UMR 5182, 46 Allée d’Italie, 69364 Lyon, Cedex 07, France

Supporting Information

ABSTRACT: First-principles simulations of two-dimensional electronic spectroscopy in the ultraviolet region (2DUV) require computationally demanding multiconfigurational approaches that can resolve doubly excited and charge transfer states, the spectroscopic fingerprints of coupled UV-active chromophores. Here, we propose an efficient approach to reduce the computational cost of accurate simulations of 2DUV spectra of benzene, phenol, and their dimer (i.e., the minimal models for studying electronic coupling of UV-chromophores in proteins). We first establish the multiconfigurational recipe with the highest accuracy by comparison with experimental data, providing reference gas-phase transition energies and dipole moments that can be used to construct exciton Hamiltonians involving high-lying excited states. We show that by reducing the active spaces and the number of configuration state functions within restricted active space schemes, the computational cost can be significantly decreased without loss of accuracy in predicting 2DUV spectra. The proposed recipe has been successfully tested on a realistic model proteic system in water. Accounting for line broadening due to thermal and solvent-induced fluctuations allows for direct comparison with experiments.



INTRODUCTION

Benzene and phenol are aromatic chromophores present in the side chains of phenylalanine (Phe, F) and tyrosine (Tyr, Y) amino acid residues (Figure 1). Because of their characteristic near-ultraviolet (UV) absorption and the low occurrence of these residues in proteins (~3–4%),¹ they have been suggested to be targets of ultrafast nonlinear UV spectroscopy^{2,3} aimed at the determination of side chain interactions⁴ and eventually folding/unfolding pathways of proteins in solution.⁵ Linear absorption and circular dichroism spectroscopies, while being powerful tools for differentiating secondary protein structures, suffer from high spectral congestion and do not provide direct information on tertiary or quaternary structures of proteins. Two-dimensional (2D) electronic spectroscopy (2DES)^{3,6} is becoming an attractive tool for determining structural propensities with high spectral and temporal resolution. In 2DES experiments (Figure 1), coherent ultrashort pulses are used to correlate the wavelength of the pump pulses to the wavelength of the probed transitions, thereby adding a second dimension for manipulating the nonlinear signal and achieving unprecedented spectral resolution. Because of technical difficulties, practical extension of the 2DES technique to the high-energy UV spectral region (2DUV) has been achieved only recently.^{7–9} 2DUV signals contain information on electronic transitions (and their couplings) involving high-

energy excited states whose energies can lie beyond the ionization threshold of the target chromophores.

The complexity of the nonlinear response recorded in 2D electronic spectra has motivated the development of computational approaches to simulate and interpret 2DUV spectra.^{10–13} Two approaches have been proposed to tackle this complex issue: (i) parametrized Frenkel exciton Hamiltonian models that include electrostatic fluctuations, with parameters obtained from reference gas-phase quantum-mechanical (QM) calculations and intermolecular couplings calculated with a grid-based method;¹¹ and (ii) fully ab initio calculations using hybrid quantum mechanics/molecular mechanics (QM/MM) with electrostatic embedding in conjunction with the sum-over-states approach (i.e., the SOS//QM/MM).¹² Frenkel exciton Hamiltonians offer the great advantage of treating large chromophoric systems at low computational cost while taking into account signal broadening effects due to both environment and thermal fluctuations, making simulations of 2DES spectra of realistic systems computationally feasible. However, exciton Hamiltonians are generally constructed using the ground-state gas-phase equilibrium geometries of the chromophores and include just a few states within their singly excited state

Received: May 13, 2015

Published: July 1, 2015



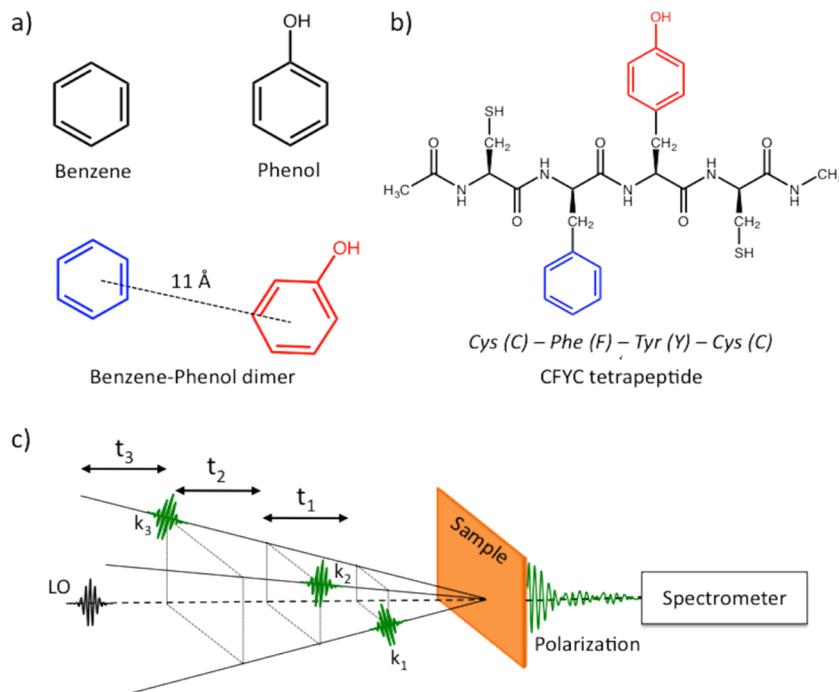


Figure 1. UV-active chromophoric systems including (a) benzene, phenol, and their dimer in the gas phase, and (b) the CFYC tetrapeptide (solvated in water) used as targets for (c) two-dimensional electronic spectroscopy. Schematic representation of (c) a heterodyne-detected three-pulse photon echo experiment (LO = local oscillator).

manifolds, neglecting charge transfer (CT) states. The modeling through simple exciton Hamiltonians thus provides a crude low-cost estimation of the nonlinear signals, as clearly revealed by comparisons with more accurate simulations based on the SOS//QM/MM method.^{4,12} On the other hand, fully ab initio calculations of multichromophoric systems require a proper description of high-lying excited states with different (covalent or ionic) characters involving single and double excitations (D) as well as CT states. Theoretical predictions of 2DUV spectra face the challenge of being able to characterize the transition energies of these high-lying excited states with errors well below the largest bandwidth of the ultrashort UV pulses within reach today (i.e., below $\sim 5000 \text{ cm}^{-1}$, which corresponds to 0.62 eV). These constraints impose the use of computationally very demanding multiconfigurational treatment for which errors lower than 0.2 eV can be expected when the highest levels of theory (with large basis sets and active spaces and affordable computational costs) are employed. Thus, the computational cost of SOS//QM/MM simulations restricts its application to relatively small molecular aggregates and to a limited number of selected structures. This hampers the full exploration of the configurational spaces of the multichromophoric systems and direct calculations of the 2D signal broadenings, forcing the use of constant broadenings for all electronic transitions.² The recent acquisition of 2DUV experimental data of photolabile biological samples^{7–9} is encouraging for the development of computational strategies that can provide accurate theoretical 2DUV spectra with realistic line widths and signal broadenings.

In this work, we propose a computational protocol that minimizes the computational cost of the SOS//QM/MM simulations by employing the restricted active space self-consistent field (RASSCF) multiconfigurational technique in conjunction with the PT2 multireference perturbation theory. We aim to provide a computational treatment that guarantees

the same accuracy of previously reported SOS//QM/MM simulations based on complete active space self-consistent field (CASSCF) and perturbation (CASPT2) techniques (i.e., CASSCF//CASPT2 approach) with computational savings that enable extensive SOS//QM/MM calculations of a large number of molecular structures.

We push the computational limits of the multiconfigurational treatment to set the reference calculations for benzene and phenol chromophores with respect to experimental gas-phase data. These results could be used to build new and more extended exciton models for large-scale simulations of 2DUV spectra of proteic systems. Moreover, these reference calculations provide relevant information on the lowest excited states above the first ionization potentials of benzene and phenol, the so-called “superexcited” valence states,¹⁴ for which experimental transition energies are ambiguously assigned. As the cross-section of the direct ionization is found to be relatively small at energy slightly above the ionization limit, superexcited states can appear as discrete peaks in the ionization continuum,^{14–17} and thus contribute to 2DUV spectra. These valence states have very short lifetimes and may decay to a continuum of Rydberg states by autoionization or by defragmentation, giving rise to broad but structured absorption signals.

Wave function methods, like CASSCF//CASPT2, can be systematically improved, thus obtaining optimal values for transition energies and dipole moments that can be used, however, as reference rather than for practical applications. Bearing in mind that any reduction in the level of the wave function theory adopted will introduce errors in the wave function, we explore the possibility of decreasing the computational cost by reducing the active spaces and the basis sets adopted while preserving the accuracy. By monitoring the changes induced by any (basis set, active space, configuration state functions) truncation, we could play with the multi-

reference perturbation treatment to impose error compensation and calibrate the low cost results in a semiempirical fashion. The reliability of the calibration process is guaranteed by prior assessment of reference theoretical values against experimental data. In particular, we introduce an approach allowing the use of minimal active spaces (that include just the π -orbitals of the aromatic chromophores) while yielding simulated 2DUV spectra of benzene and phenol monomers that reproduce those with the highest level of theory. We focus on 2D spectra obtained by pumping in a spectral window where the lowest excited states of the aromatic side chains (i.e., the L_b states, following Platt's notation)¹⁸ are located at 4.5–5.0 eV, an optical region experimentally within reach and spectrally well-separated from the protein backbone absorptions (lying in the deeper UV, at energies >5.4 eV). The proposed approach is tested on the benzene-phenol dimeric aggregate in the gas phase, a model system with noninteracting chromophores that allows comparison with reference calculations and experimental data on monomers. The resulting good comparison encouraged attempts to further reduce the computational cost by employing less expensive RAS schemes. Finally, the best recipes are applied to a realistic model proteic system containing benzene and phenol chromophores as aromatic side chains (i.e., phenylalanine (Phe, F) and tyrosine (Tyr, Y) residues and the cysteine-phenylalanine-tyrosine-cysteine (CFYC) tetrapeptide) (Figure 1). We have previously shown that 2DUV spectra of CFYC contain enough structural information to clearly distinguish between polypeptide configurations with interacting and noninteracting aromatic residues,⁴ and how this information can be used to track the unfolding dynamics of proteic systems.⁵ However, because of the significant computational cost of the large active space CASSCF//CASPT2 calculations employed in these studies, only optimized ground-state structures and few snapshots along a biased unfolding molecular dynamics have been used for the simulations of the 2DES spectra. The present protocol allows for a major reduction in computer time for each electronic structure calculation with explicit computation of line broadening due to thermal fluctuations. This level of theory will be of vital importance for interpreting upcoming 2DUV experimental spectra.

■ COMPUTATIONAL DETAILS

Multiconfigurational calculations were performed using state-average (SA)-CASSCF or -RASSCF methodology as implemented in the Molcas 7.7 code,^{19,20} including up to 12 and 100 states in the state-averaging procedure for monomers and dimers, respectively. The choice of the number of states included in the SA was based on the number of excited states necessary to simulate 2DUV electronic spectra. SA-CASSCF (or SA-RASSCF) calculations were followed by single state (SS) CASPT2 calculations to account for dynamic correlation; hereafter, for simplicity, we avoid repeating that CASSCF or RASSCF calculations (with different active spaces) are followed by PT2 treatment and will just refer to the different multiconfigurational treatments. The ionization-potential-electron-affinity (IPEA) shift²¹ was set to 0.0, and an imaginary shift of 0.2 au was used.²² Transition dipole moments (TDMs) were calculated at SA-CASSCF (or SA-RASSCF) level using the RASSI routine of the Molcas code. The Cholesky decomposition approach²³ was used to accelerate the computations of two-electron integrals. The calculations of the monomers were conducted under C_s -symmetry. Never-

theless, the D_{6h} notation is used throughout, and thus, phenol transitions are referred to as benzene-like transitions.

Reference Calculations of Monomers. Reference calculations of benzene and phenol monomers in gas phase were performed at the RASSCF//SS-CASPT2 level using RAS(0,0/6,6/2,12) and RAS(0,0/8,7/2,12) large active spaces (IAS), respectively, where RAS nomenclature follows RAS (maximal number of holes in RAS1, number of RAS1 orbitals/number of electrons in RAS2, number of RAS2 orbitals/maximal number of electrons in RAS3, number of RAS3 orbitals). Twelve virtual orbitals were included in the RAS3 space, where only single and double transitions are allowed, adding up to the RAS2 space, including the “minimal” active spaces (*mAS*) of benzene and phenol, which involve the complete π -system and the phenol oxygen lone pair (i.e., the CAS(6,6) and CAS(8,7) spaces, respectively). Similar RAS schemes have been adopted for setting reference calculations of larger UV chromophores, such as indole¹³ and adenine.²⁴ We have also verified the reliability of these restricted active spaces by comparing with their corresponding complete active spaces (see Table S1, Supporting Information (SI)). The choice of reference calculations with IAS is based on the comparison with experimental data. We found that *mAS* does not provide accurate results; thus, it is essential to enlarge the active spaces to include enough correlation that provides correct state-ordering already at the CAS- or RAS-SCF level. In fact, the σ - π polarization effects, which are different for covalent and ionic states, and for singly and doubly excited states, are neglected when the active space is comprised only of π -orbitals.

The contracted ANO-L basis set was used with the following contraction scheme: C,O/[4s3p2d] and H/[2s1p]. In addition, the standard ANO basis sets were augmented with a set of 1s-, 1p- and 1d-contracted Rydberg-type basis functions positioned at the center of nuclear charge (namely ANO-L(432,21)-aug) to minimize Rydberg-valence mixing.²⁵ Despite the fact that with this augmented large basis set we would be able to describe both valence and Rydberg states with equal footing, in the present report, we focus exclusively on the valence states because in solution Rydberg states blue-shift, giving rise to broad absorption bands with low resolution. We follow an idea originally proposed by Roos et al.,^{26,27} namely, to improve the description of valence states through an increase in the active space with extra-valence orbitals while avoiding Rydberg-valence mixing by first optimizing and subsequently removing Rydberg orbitals from the molecular orbital set. This successful approach relies on the observation that well-described Rydberg orbitals (thus, the necessity of introducing diffuse basis functions) do not interact with valence orbitals.²⁶ In particular, three Rydberg orbitals with A'' symmetry (associated with the p_y , d_{xy} , and d_{yz} diffuse basis functions) were added to the *mAS* of benzene and phenol. Then, state-average CAS(6,9) and CAS(8,10) computations for benzene and phenol, respectively, were performed with a large enough number of excited states to incorporate at least three Rydberg excited states, each involving a single electron transition from occupied π -orbitals to one of the three Rydberg type virtual orbitals. In fact, once each Rydberg orbital is optimized in the state-average calculation, it could be excluded from the molecular orbital set. After removal of the three Rydberg orbitals with A'' symmetry, the active spaces were progressively increased by adding 4, 8, and 12 virtual extra-valence orbitals. The choice of orbitals was based on diagonal elements in the effective state-averaged Fock matrix. Despite the nature of the extravalue orbitals not being

crucial for the outcome of the procedure, the selected orbitals are expected to contribute the most to the dynamic correlation, helping convergence of PT2 transition energies to be achieved with a relatively small number of extravalence orbitals. In the course of increasing the progressive active space, we encountered three further orbitals, which gave rise to Rydberg transitions (a consequence of the large basis set). These were also excluded from the molecular basis set. Finally, in the perturbation procedure, all of the orbitals (including Rydberg orbitals with A' symmetry) were included, except for the three Rydberg orbitals with A'' symmetry initially removed.

Optimal Setup for Dimer Calculations. In an effort to reduce the computational cost of *mAS* calculations, which are too expensive for treating dimeric systems, we examined the possibility of reducing the basis set and the active space sizes, aiming at optimal computational setups of monomers that minimize the mean deviation from the reference calculations to be exploited for dimeric systems. Because the ANO-L(432,21) and ANO-L(321,21) basis sets used in conjunction with the CAS(6,6) and CAS(8,7) *mAS* do not provide adequate results for the transition energies of monomers, we have introduced a protocol that allows the use of *mAS* and the small ANO-L(321,21) basis set without loss of accuracy. The protocol simply consists of the selection (or localization) and removal of extravalence virtual π^* -orbitals (with high angular momentum) from the set of CAS molecular orbitals in order to properly account for the overestimation of dynamic correlation introduced with the perturbation treatment of the *mAS* calculations in a similar way to how real and imaginary shifts are more commonly used in PT2 treatment (see Results and Discussion).

In the monomers, the selection of the extravalence virtual π^* -orbitals is facilitated by the C_S symmetry. In dimers, the symmetry is lost, and strong mixing of π^* -orbitals with σ^* -orbitals can complicate the virtual orbital selection procedure. For instance, in strongly interacting dimers (such as stacked chromophores), discriminating the virtual π^* -orbitals becomes virtually impossible due to the strong orbital mixing/delocalization. Therefore, we have devised a general protocol for localization and removal of π^* -orbitals from the virtual orbitals manifold through the following steps: (i) a standard minimal active space SA-CASSCF calculation (e.g., CAS(14,13) for the benzene-phenol dimer) is performed; (ii) the Pipek-Mezey²⁸ localization procedure is used to localize the virtual orbitals (although the Pipek-Mezey localization procedure is meant for occupied orbitals, and the resulting orbitals are semilocalized at most, we found that orbitals perpendicular to the aromatic plane can be clearly distinguished from orbitals that lie in the plane); (iii) localized orbitals are allowed to “re-delocalize” in a SA-CASSCF calculation, where the converged CI-coefficients of step i are used and, more importantly, rotations between the localized π^* -orbitals of the aromatic rings and the remaining virtual orbitals are forbidden using the subsymmetry approach. Because the CASSCF energy is invariant with respect to virtual–virtual rotations, the latter calculation provides exactly the same outcome as step i but with the π^* -orbitals separated from the other virtual orbitals, allowing easy removal of extravalence π^* -orbitals from the set of molecular orbitals.

The proposed localization/delocalization procedure was tested with the generally contracted ANO-type basis sets and was found to work with contraction schemes having up to one set of d functions on the nonhydrogen atoms (e.g., 3s2p1d or

4s3p1d) for both stacked and unstacked configurations of the benzene-phenol dimer. However, it fails with larger contraction schemes due to mixing of p- and d-type orbitals. Nonetheless, if more than one set of d functions is required, a detour first separates the π^* -orbitals via the procedure described above using a smaller contraction scheme, and subsequently, the basis set is expanded to the desired size using the EXPBAS routine in Molcas. This protocol was found to introduce an overall <0.05 eV error to the absolute values of the CASSCF (or RASSCF) energies.

Selected Structures. The gas-phase benzene and phenol geometries were taken from Rogers et al.²⁹ The noninteracting benzene-phenol dimer geometry (reported in the SI) was constructed by placing the two chromophores arbitrarily in space (thus, without symmetry constraints) at a distance of ~11 Å. The cysteine-phenylalanine-tyrosine-cysteine (CFYC) tetrapeptide was taken as the model proteic system,⁴ and two configurations with interacting and noninteracting aromatic chromophores were selected from previously reported unfolding dynamics.⁵ The two selected configurations correspond to snapshots number 14 (folded configuration) and 31 (unfolded configuration), whose structures are reported in ref 5.

Simulations of 2D Electronic Spectra. In a heterodyne detected three-pulse photon echo 2DUV experiment^{30,31} three laser pulses with wavevectors k_1 , k_2 , and k_3 interact with the sample, giving rise to a third-order nonlinear polarization that emits a signal along a given phase-matched direction, which is superimposed with a fourth pulse known as the local oscillator (LO) (see Figure 1). The nonlinear polarization is related to the nonlinear response of the system $R^{(3)}(t_1, t_2, t_3)$ ^{30–32} generated by the multiple field-matter interactions

$$R^{(3)}(t_1, t_2, t_3) = \left(\frac{i}{\hbar}\right)^3 \text{Tr}[\hat{\mu} G(t_3) \mathbf{V}G(t_2) \mathbf{V}G(t_1) \mathbf{V}\hat{\rho}(0)] \quad (1)$$

where $\hat{\rho}(0) = |g\rangle\langle g|$ is the density matrix of the system in the ground state equilibrium (g) (i.e., before interacting with the first pulse), $\mathbf{V}\hat{\rho}(0) \equiv [\hat{\mu}, \hat{\rho}]$ is the dipole superoperator, $G(t)$ is the retarded Green's function (forward propagator) describing the free evolution of the density matrix between interaction events

$$G(t)\hat{\rho}(0) = \Theta(t)e^{-(\frac{i}{\hbar})\hat{H}(t)}\hat{\rho}(0)e^{(\frac{i}{\hbar})\hat{H}(t)} \quad (2)$$

with $\Theta(t) = \int_{-\infty}^t d\tau \delta(\tau)$, the Heaviside step-function ensuring causality, and \hat{H} the vibronic Hamiltonian. The coordinate-dependent transition dipole moment $\hat{\mu} = \sum_{ij}|i\rangle\langle j|\mu_{ij}$ couples each pair of states i and j and commutes with the density matrix operator, giving rise to different interaction sequences of the incident electric fields with the system (i.e., the Feynman pathways). For the rephasing k_1 ($k_{\text{LO}} = -k_1 + k_2 + k_3$) and nonrephasing k_{II} ($k_{\text{LO}} = +k_1 - k_2 + k_3$) phase-matching conditions, the Feynman pathways include ground state bleach (GSB), excited state absorption (ESA), and stimulated emission (SE). A suitably positioned detector can be used to select pathways subject to a given phase-matching condition. For instance, for the rephasing $k_1 = -k_1 + k_2 + k_3$ phase-matching condition, the GSB, ESA, and SE pathways are simultaneously recorded in the response function

$$R_{k_1}^{(3)}(t_1, t_2, t_3) = \sum_{i=\text{GSB,ESA,SE}} R_{k_1,i}^{(3)}(t_1, t_2, t_3) \quad (3)$$

In the limit of slow dynamics, where all dephasing processes occur on a time scale much slower than the duration of the experiment (the “snapshot” approximation), the GSB, SE, and ESA response functions can be written as^{2,33}

$$R_{k_1, \text{GSB}}^{(3)}(t_1, t_2, t_3) = \left(\frac{i}{\hbar}\right)^3 \Theta(t_1)\Theta(t_2)\Theta(t_3) \sum_{ee'} \mu_{ge'}\mu_{eg}\mu_{e'g} e^{-i(\omega_{e'g}-i\gamma)t_3-i(\omega_{eg}-i\gamma)t_1} \quad (4.1)$$

$$R_{k_1, \text{SE}}^{(3)}(t_1, t_2, t_3) = \left(\frac{i}{\hbar}\right)^3 \Theta(t_1)\Theta(t_2)\Theta(t_3) \sum_{ee'f} \mu_{ge'}\mu_{eg}\mu_{e'g}\mu_{ge} e^{-i(\omega_{e'g}-i\gamma)t_3-i(\omega_{e'e}-i\gamma)t_2-i(\omega_{eg}-i\gamma)t_1} \quad (4.2)$$

$$R_{k_1, \text{ESA}}^{(3)}(t_1, t_2, t_3) = -\left(\frac{i}{\hbar}\right)^3 \Theta(t_1)\Theta(t_2)\Theta(t_3) \sum_{ef} \mu_{ef}\mu_{fe}\mu_{e'g}\mu_{ge} e^{-i(\omega_{fe}-i\gamma)t_3-i(\omega_{e'e}-i\gamma)t_2-i(\omega_{eg}-i\gamma)t_1} \quad (4.3)$$

During t_1 , the system density matrix is in an optical coherence with a characteristic frequency ω_{eg} given by the energy difference between states e and g at the time of the interaction with the laser pulse with e representing states of the singly excited state manifold. During the interval t_2 , in the case of SE and ESA, the system evolves in the single-exciton manifold, being either in a coherence state ($e' \neq e$) or in a population state ($e = e'$); whereas in the case of GSB, the system is in the ground state g . Population transfer between states from the singly excited state manifold, as well as population relaxation, back to the ground state during t_2 are not considered in eqs 4.1–4.3 but can be introduced.^{2,33} During the third interval (t_3), coherence between the ground and singly excited state manifold (for GSB and SE) or between the singly and doubly excited state manifolds (for ESA) is created, which oscillates with frequency $\omega_{e'g}$ or ω_{fe} , respectively. Signal lineshapes are modeled with a phenomenological dephasing rate constant γ . Eqs 4.1–4.3 can be easily reformulated for the k_{II} nonrephasing phase-matching condition, generating the set of working equations that we used to calculate the nonlinear response functions in this work. Within the sum-over-states (SOS) approximation, transition energies and transition dipole moments among the manifolds of excited states are all that is required to compute the nonlinear response functions according to eqs 4.1–4.3.

The third-order signal generated in the k_S direction (here, we considered $k_S = k_1$ or $k_1 + k_{II}$, where $k_{II} = k_1 - k_2 + k_3$)

$$S_{k_S}^{(3)}(t_1, t_2, t_3) = \int_0^\infty dt \int_0^\infty dt_3 \int_0^\infty dt_2 \int_0^\infty dt_1 R_{k_S}^{(3)}(t_1, t_2, t_3) \times E(r, t)E(r, t-t_3)E(r, t-t_3-t_2)E(r, t-t_3-t_2-t_1) \quad (5)$$

is a convolution of the wavevector-dependent response function and four laser pulses ($E(r, t) = E(t) e^{ikr-i\omega(t)}$) with central frequency ω and complex envelope $E(t)$.

Two dimensional signals are obtained using three temporally well-separated ultrashort laser pulses and scanning the delay times t_1 and t_3 for a given delay time t_2 . In the present work, the waiting time t_2 is set to zero (i.e., coherent excited state dynamics are eliminated). The bidimensional electronic spectra were obtained by a 2D Fourier transform along t_1 and t_3 .

$$S_{k_S}^{(3)}(\Omega_1, t_2, \Omega_3) = \int_0^\infty dt_3 \int_0^\infty dt_1 S_{k_S}^{(3)}(t_1, t_2, t_3) e^{i\Omega_1 t_1 + i\Omega_3 t_3} \quad (6)$$

The transition dipole moments and energies were obtained at the multiconfigurational (CASSCF or RASSCF) and multireference level (PT2), respectively. The SOS//QM/MM approach has been used for the CFYC tetrapeptide solvated in water, as previously reported.^{4,12} The rephasing (k_1) signals have been simulated for the gas-phase species (i.e., the monomers and the noninteracting dimer model). For the realistic proteic model (i.e., the solvated tetrapeptide), we have simulated the combined rephasing and nonrephasing ($k_1 + k_{II}$) signals, which possess absorptive features and could be directly compared to experimental 2D spectra collected with both the heterodyne detected three-pulse photon echo and the collinear pump–probe geometry setups.^{32,34} A fixed dephasing of 200 cm⁻¹ has been employed for all transitions. The laser pulses have finite transform-limited Gaussian envelopes and different bandwidths and central frequencies, as specified in the text where necessary. All spectra refer to the nonchiral xxxx polarization configuration.³⁵ Calculations were performed with Spectron 2.7³³ and 2D spectra are plotted on a logarithmic scale. Nonlinear negative signals due to GSB, overlapping with SE, appear along the diagonal of the 2D map as negative peaks, whereas ESAs from populated excited states give rise to off-diagonal positive peaks.

RESULTS AND DISCUSSION

Reference Computations of Benzene and Phenol Monomers. In this section, we compare the experimental electronic spectra of benzene and phenol monomers below and above their ionization limits (up to 11 eV) with large active space (IAS) calculations at the RAS(0,0/6,6/2,12)/ANO-L(432,21)-aug and RAS(0,0/8,7/2,12)/ANO-L(432,21)-aug levels, respectively, hereafter named “reference calculations”. These RAS schemes have been shown to yield transition energies that are within 0.06 eV of the corresponding CAS calculations (see Table S1 in the SI) (i.e., CAS(6,18) and CAS(8,19) for benzene and phenol, respectively), providing a consistent gain of computational cost. The inclusion of occupied and virtual σ -orbitals in the reference calculations of the benzene monomer (see Table S2 in the SI) does not affect the energies of covalent states, whereas the ionic states are blue-shifted, and transition energies of superexcited valence states become significantly overestimated with respect to the experimental data.

The valence states below the first ionization potential of benzene (at ~ 9.25 eV)¹⁷ and phenol (at ~ 8.5 eV)¹⁶ have been documented in numerous experiments,^{17,36–42} and their electronic nature has been described theoretically at different levels of theory, including the CASSCF//CASPT2 level.^{27,29,42–48} It has been demonstrated that a minimal active space (mAS) constructed out of the π -orbitals of benzene (i.e., CAS(6,6)) or phenol (i.e., CAS(8,7)) is insufficient to describe the valence excited states with covalent (e.g., ${}^1\text{B}_{2u}$ ${}^1\text{E}_{2g}$) and with ionic (e.g., ${}^1\text{B}_{1u}$ ${}^1\text{E}_{1u}$) character (in the valence bond description) at equal footing. The dynamical $\sigma-\pi$ polarization, which is completely neglected at the CASSCF level when a minimal active space is employed, is overestimated for ionic configurations at the CASPT2 level. Increasing the active spaces by including additional extravalence orbitals significantly

Table 1. PT2 Vertical GS → S_N Excitation Energies (eV) and Transition Dipole Moments Magnitudes (au) in the Benzene and Phenol Monomers Obtained at the RAS(0,0/6,6/2,12) and RAS(0,0/8,7/2,12) Level, Respectively, Using the ANO-L(432,21)-aug Basis Set^a

state	excitation energy	benzene				phenol				labels	
		TDM		transition	coeff.	TDM		transition	coeff.		
		GS	¹ B _{2u}			GS	¹ B _{2u}				
¹ B _{2u}	4.67	0.00		2 → 4	0.63	4.47	0.28	3 → 5	0.71	1	
	(4.89–4.90) ^b			3 → 5	0.63	(4.51–4.59) ^{d,e}		2 → 4	-0.53		
¹ B _{1u}	6.12	0.00	0.00	2 → 5	0.67	5.81	0.45	0.06	3 → 4	-0.73	
	(6.17–6.20) ^b			3 → 4	-0.67	(5.77–5.82) ^{d,e}		2 → 5	-0.55		
¹ E _{1u}	6.77	2.20	0.06	2 → 5	-0.54	6.51	2.21	0.17	2 → 5	0.73	
	(6.86–6.98) ^{b,c}			3 → 4	-0.54	(6.70) ^d		2 → 4	-0.50		
¹ E _{1u}	6.77	2.20	0.06	2 → 4	-0.54	7.02	1.83	0.11	2 → 4	0.60	
	(6.86–6.98) ^{b,c}			3 → 5	0.54	(6.70) ^d		3 → 5	0.46		
¹ ¹ E _{2g}	7.75	0.00	0.00	1 → 5	0.51	6.92	0.86	0.31	1 → 5	-0.54	
	(7.80–8.00) ^c			2,3 ⇒ 5	-0.37	(6.93) ^d		2,3 ⇒ 5	0.41		
				2 → 6	-0.31			2 → 4	-0.37		
				2 ⇒ 4,5	0.31						
¹ ¹ E _{2g}	7.75	0.00	0.00	1 → 4	-0.51	7.65	0.13	0.03	1 → 4	0.51	
	(7.80–8.00) ^c			2,3 ⇒ 4,5	-0.34	(7.50) ^f		3 ⇒ 4	0.44		
				3 ⇒ 4	0.32			3 → 6	-0.41		
¹ D*	9.70	0.00	0.00	2,3 ⇒ 4,5	-0.60	9.03	0.33	0.40	3 ⇒ 5	0.47	
	(9.30–9.50) ^c			2 ⇒ 5	0.41	(9.00) ^g		3 → 6	0.40		
				3 ⇒ 4	0.41			1 → 4	0.40		
								2,3 ⇒ 4,5	-0.37		
² ¹ E _{2g}	9.69	0.00	0.64	2 → 6	0.64	9.10	0.21	0.73	1 → 5	-0.49	
	(9.30–9.50) ^c			1 → 5	0.53	(9.00) ^g		3 → 4,5	-0.48		
² ¹ E _{2g}	9.69	0.00	0.64	3 → 6	-0.64	9.15	0.13	0.46	2,3 ⇒ 4,5	0.42	
	(9.30–9.50) ^c			1 → 4	0.53	(9.00) ^g		1 → 4	0.39		
								3 → 6	0.36		
								2 ⇒ 5	0.30		
								3 ⇒ 4	0.30		
² D*	10.43	0.00	0.00	2,3 ⇒ 4	0.46	9.76	0.28	0.64	3 ⇒ 4,5	0.50	
	(10.1–10.3) ^c			3 ⇒ 4,5	0.46			2 ⇒ 4,5	-0.40		
				2 ⇒ 4,5	-0.46			2,3 ⇒ 5	0.39		
³ D*	10.88	0.00	0.00	2,3 ⇒ 4,5	-0.45	10.30	0.02	0.54	2,3 ⇒ 4,5	-0.39	
				3 ⇒ 5	0.36			2 ⇒ 4	0.36		
				2 ⇒ 4	0.35						
¹ S*						10.10	0.66	0.50	(O _{lp}) → 5	0.73	
								2 → 6	-0.38		
² S*						10.47	0.30	0.42	(O _{lp}) → 4	-0.69	
								2 ⇒ 5	0.40		

^aDominant configurations (referred to orbitals in Figure S1 in the SI) and corresponding CI coefficients (≥ 0.30). States nomenclature refers to the ideal D_{6h} symmetry of benzene. Experimental transitions energies are reported in parentheses. ^bFrom refs 36–38. ^cFrom ref 17 (see also notes in the text). ^dFrom ref 36. ^eFrom refs 39 and 40. ^fFrom ref 15. ^gFrom ref 16.

improves the accuracy, reaching a quantitative agreement with experiments.

The experimental electronic spectrum of benzene below the first ionization potential shows four transitions appearing at ~4.90,^{36–38} ~6.20,^{36–38} ~6.90,^{17,36–38} and ~8.00 eV,¹⁷ corresponding to excitations from the ground state (GS) into the ¹B_{2u}, ¹B_{1u}, and the 2-fold degenerate ¹E_{1u} and ¹E_{2g} states (in the D_{6h} notation), respectively (see Table 1). Nakshima et al.¹⁷ reported 7.80 eV for photolyzed benzene solution, but this value was obtained by taking as reference value the 0–0 transition energy of 4.70 eV for benzene in cyclohexane, as reported by Eastman et al.⁴¹ Considering 4.9 eV as the

fundamental vertical transition energy,^{36–38} the upper limit of the GS → ¹E_{2g} absolute vertical excitation would be 8.00 eV. In Table 1, we report an experimental value of 7.80–8.00 eV. Our reference calculation at the RAS(0,0/6,6/2,12)/ANO-L(432,21)-aug level is in good agreement with these experimental data with a general minor underestimation of all vertical transition energies (deviations < 0.1 eV). However, for the fundamental GS → ¹B_{2u} transition, the absolute error with respect to the most widely accepted value (i.e., 4.90 eV)^{36–38} is 0.23 eV. Previously reported computations²¹ showed how this discrepancy could be reduced by increasing the IPEA shift, but our attempts to use shifts larger than 0.0 au, while conferring

better GS \rightarrow ${}^1\text{B}_{2u}$ transition energies, provided overestimations of higher transition energies and, overall, a worse picture of the singly and doubly excited state manifolds. The transitions from the GS to the ${}^1\text{B}_{2u}$ (covalent), ${}^1\text{B}_{1u}$ (ionic), and to the lowest pair of ${}^1\text{E}_{2g}$ (covalent) states (${}^1\text{E}_{2g}$) are dipole forbidden (see Table 1) but would become accessible due to vibronic coupling. Instead, the transition to the ionic ${}^1\text{E}_{1u}$ doublet is dipole allowed with high oscillator strength of ~ 0.80 ($|\text{TDM}| = 2.20$ au), which is in agreement with experimental results (~ 0.90)³⁷ and previous calculations (~ 0.82 ,²⁷ ~ 0.89 ,⁴³ ~ 0.79 ²⁹).

The phenol valence states have a similar nature to benzene (see Table 1), and for this reason, we keep using the D_{6h} notation (although phenol has only C_s symmetry). In the energy region below the phenol ionization limit (~ 8.5 eV), blue-shifted absorptions and splitting of both the ${}^1\text{E}_{1u}$ and ${}^1\text{E}_{2g}$ doublets are observed with respect to the benzene case, in agreement with experimental data of most substituted benzenes.³⁶ In particular, the first two excited states of phenol (i.e., ${}^1\text{B}_{1u}$ and ${}^1\text{B}_{2u}$) have been experimentally detected at ~ 4.51 – 4.69 eV^{36,39,40} and ~ 5.77 – 5.82 eV,^{36,39} respectively. A broad asymmetric UV band peaked at 6.70 eV with a shoulder at 6.93 eV has been observed for phenol by Kimura et al.,³⁶ indicating a significant splitting of the ${}^1\text{E}_{1u}$ doublet (estimated to be ~ 0.23 eV) with respect to benzene. The analysis of the three-photon ionization pathways of phenol has also provided experimental evidence of another valence state lying below the first ionization limit at energy around 7.50 eV.¹⁵ Our reference calculation at the RAS(0,0/8,7/2,12)/ANO-L(432,21)-aug level provides transition energies in good agreement with the experimental evidence with absolute errors for the first two excited states (i.e., ${}^1\text{B}_{1u}$ and ${}^1\text{B}_{2u}$) within 0.05 eV. At higher energies, we found a split ${}^1\text{E}_{1u}$ doublet with the two states lying at 6.51 and 7.02 eV, and with the lower-energy transition having larger oscillator strength ($f = 0.78$ and $|\text{TDM}| = 2.21$ au) with respect to the higher-energy transition (having $f = 0.58$ and $|\text{TDM}| = 1.83$ au). These results are in agreement with the broadened band observed experimentally and consistent with the functionalization effects generally observed for monosubstituted benzenes.³⁶ The ${}^1\text{E}_{2g}$ doublet is also predicted to split below the phenol ionization limit with transition energies at 6.92 and 7.65 eV and non-negligible oscillator strengths. This result suggests that the low-lying ${}^1\text{E}_{2g}$ state could contribute to the asymmetric broad band at around 7.00 eV,³⁶ whereas the valence state found at 7.50 eV and assigned to a single excitation from a low-lying occupied molecular orbital (MO) to the lowest unoccupied MO¹⁵ indeed corresponds to the high-lying ${}^1\text{E}_{2g}$ state calculated at 7.65 eV (see Table 1). Although for phenol we observe a nice agreement with the experimental data, it is necessary to mention that we believe that the ${}^1\text{E}_{1u}$ and ${}^1\text{E}_{2g}$ splittings are slightly overestimated²⁷ in our reference calculation (being 0.51 and 0.73 eV, respectively) due to wave function mixing at the RASSCF level. The reason is that the σ - π polarization is similar for the ground and the other covalent states, allowing a fairly good description of those states even without dynamic correlation, whereas ionic states, on the contrary, exhibit different polarization effects and are not described as properly at the RASSCF level. This can cause erroneous state ordering or, by mischance, artificial near degeneracy of energetically distant electronic states. In systems with low symmetry, such as phenol, this near degeneracy leads to artificial wave function mixing, as in the case of the high-lying (ionic) ${}^1\text{E}_{1u}$ state (whose energy is overestimated at the RASSCF, or CASSCF, level) and low-lying (covalent) ${}^1\text{E}_{2g}$

state, finally decreasing the accuracy of the RASSCF/PT2 energies. The comparison between experimental and computed transition energies below the ionization potential of phenol, shown in Table 1, indicates that the wave function mixing increases the error for the low-lying ${}^1\text{E}_{1u}$ state and the high-lying ${}^1\text{E}_{2g}$ state energies for which absolute errors of 0.1–0.2 eV are observed. It is worth mentioning that this problem is expected to be even more important at higher energies, where the density of states increases. Although one possible solution to this problem could be the use of the multistate PT2 method,⁴⁹ we observed that in our cases this approach leads to too large of shifts (> 1 eV) because the off-diagonal elements in the multistate PT2 Hamiltonian are large (an order of magnitude larger than the 0.002 au upper limit), a condition in which the multistate PT2 is not intended to be more reliable than SS-PT2.

Regarding the general trend of TDMs, we observed that the hydroxyl functionalization of benzene induces an increase in the TDMs of almost all transitions, including those from the GS and the ${}^1\text{B}_{2u}$ excited state.

Time-resolved spectroscopy (using a 15 ns laser pulse) of a dilute benzene solution has shown the presence of discrete peaks above the ionization limit with a broad intensive absorption band at 4.60 eV and a sharp peak at 5.40 eV from the ${}^1\text{B}_{2u}$ state (i.e., involving superexcited valence states; as for the ${}^1\text{E}_{2g}$ states of benzene (see above)). Nakashima et al.¹⁷ reported absorptions of the two superexcited states at 9.30 and 10.10 eV using as reference the 0–0 transition energy at 4.70 eV. Therefore, in Table 1, we report energy ranges of 9.30–9.50 and 10.10–10.30 eV, respectively.¹⁷ Our reference calculation indicates the presence of the ${}^2\text{E}_{2g}$ doublet and the ${}^1\text{D}^*$ states at 9.69 and 9.70 eV from the GS, respectively. As expected, the benzene superexcited valence states are dark from the GS but can be excited by single electron excitations from ${}^1\text{B}_{2u}$. These high-lying singly excited states represent the ionic counterpart of the covalent ${}^1\text{E}_{2g}$ pair of states, with non-negligible contributions from double excitations, due to wave function mixing at the RASSCF level. Our results suggest that the broad band found experimentally at 4.60 eV¹⁷ from the ${}^1\text{B}_{2u}$ excited state could be assigned mainly to the ${}^1\text{B}_{2u} \rightarrow {}^2\text{E}_{2g}$ bright transitions (with $|\text{TDM}| = 0.64$ au) but also to the symmetry forbidden ${}^1\text{B}_{2u} \rightarrow {}^1\text{D}^*$ transition. The second doubly excited state (${}^2\text{D}^*$) of benzene is predicted to lie at 10.43 eV from the GS, indicating that the sharp peak observed experimentally at 5.40 eV from the ${}^1\text{B}_{2u}$ state could be assigned to the ${}^1\text{B}_{2u} \rightarrow {}^2\text{D}^*$ transition. In fact, although both ${}^1\text{D}^*$ and ${}^2\text{D}^*$ states are symmetry forbidden, the symmetry can break down due to vibronic coupling, and those transitions may be observed. Finally, we cannot exclude in this region the presence of $\sigma \rightarrow \pi^*$ transitions, which are neglected by our treatment.

Weber and co-workers investigated the role of superexcited states in femtosecond multiphoton photoelectron spectroscopy of phenol.^{15,16} They characterized a superexcited valence state around 9.0 eV, half an eV above the ionization limit of phenol,¹⁶ which is bright from both the ${}^1\text{B}_{2u}$ and ${}^1\text{B}_{1u}$ states and involves configurations with both double- and single-electron excitations. Our calculations confirm their analysis showing that the absorption around 9.0 eV can be associated with three closely lying bright states, the primary double ${}^1\text{D}^*$ excitation, and the two ${}^2\text{E}_{2g}$ single-electron excitations located at 9.03 and 9.10–9.15 eV, respectively. The low symmetry of phenol allows transitions from the ${}^1\text{B}_{2u}$ state to the ${}^1\text{D}^*$ doubly excited state. The ${}^2\text{E}_{2g}$ states are also bright, as in benzene, with

Table 2. PT2 Vertical GS → S_N Excitation Energies (in eV) of Benzene and Phenol Monomers Obtained at Different Levels of Theory, Including RAS Reference Calculations and Minimal Active Space Computations Using Different ANO-L Basis Sets^a

	¹ B _{2u}	¹ B _{1u}	¹ E _{1u}	¹ E _{lu}	¹ E _{2g}	¹ E _{2g}	1D*	² E _{2g}	² E _{2g}	2D*	3D*
Benzene											
RAS(0,0/6,6/2,12) ANO-L(432,21)-aug	4.67	6.12	6.77	6.77	7.75	7.75	9.70	9.69	9.69	10.43	10.88
CAS(6,6) ANO-L(432,21)-aug	4.64	5.74	6.27	6.27	7.65	7.65	9.34	9.05	9.06	9.96	10.54
CAS(6,6) ANO-L(432,21)	4.65	5.74	6.28	6.28	7.66	7.66	9.33	9.06	9.06	9.96	10.57
CAS(6,6) ANO-L(321,21)	4.69	5.86	6.45	6.45	7.72	7.72	9.44	9.27	9.27	10.12	10.67
CAS(6,6) ^δ ANO-L(321,21)	4.81	6.13	6.89	6.89	7.94	7.95	9.73	9.83	9.83	10.44	10.97
exp.	(4.89–4.90) ^b	(6.17–6.20) ^b	(6.86–6.98) ^{b,c}		(7.80–8.00) ^c			(9.30–9.50) ^c		(10.1–10.3) ^c	
labels	1	2	3	4	5	6	7	8	9	10	11
Phenol											
RAS(0,0/8,7/2,12) ANO-L(432,21)-aug	4.47	5.81	6.51	7.02	6.92	7.65	9.03	9.10	9.15	9.76	10.30
CAS(8,7) ANO-L(432,21)-aug	4.42	5.55	6.05	6.40	6.95	7.48	8.76	8.85	8.90	8.96	10.06
CAS(8,7) ANO-L(432,21)	4.45	5.55	6.06	6.36	7.05	7.52	8.75	8.76	8.84	9.14	9.78
CAS(8,7) ANO-L(321,21)	4.51	5.71	6.24	6.44	7.14	7.55	8.87	8.80	8.88	9.34	10.01
CAS(8,7) ^δ ANO-L(321,21)	4.64	5.96	6.63	6.86	7.43	7.76	9.29	9.35	9.33	9.72	10.34
exp.	(4.51–4.59) ^{d,e}	(5.77–5.82) ^{d,e}	(6.70) ^d	(6.93) ^d	(6.93) ^d	(7.50) ^f		(9.00) ^f			

^aStates nomenclature refers to the ideal D_{6h} symmetry of benzene. Experimental transitions energies are reported in parentheses. ^bFrom refs 36–38.

^cFrom ref 17 (see also notes in the text). ^dFrom ref 36. ^eFrom refs 39 and 40. ^fFrom ref 15.

characteristic large oscillator strengths from the ¹B_{2u} state. The ²E_{2g} doublet, in fact, represents an example of singly excited valence states dissolved in the ionization continuum that is expected to mainly contribute (as an excited state absorption signal) in multipulse UV electronic spectroscopy experiments. Finally, the electronic structures of benzene and phenol above their ionization limit, in the 10–11 eV energy range, are characterized by two doubly excited states (denoted as 2D* and 3D*), and in the case of phenol, by two other excited states arising from the single excitations involving the oxygen lone pair (O_{lp}) of the hydroxyl group, denoted as 1S* and 2S*. These latter states are characteristic signature bands of phenol in the far-UV. In summary, the above results provide good agreement with experiments and suggest new assignments for some discrete peaks observed above the ionization limits of benzene, confirming the presence of double excitations as superexcited states of both benzene and phenol. The gas-phase transition energies and dipole moments of high-lying excited states obtained here for the monomeric species could be used to construct Frenkel exciton Hamiltonians with a larger number of energy levels than those currently used to simulate 2DUV spectroscopy of aromatic side chains in proteins.⁵⁰

Minimal Active Spaces Calculations. In the previous section, we showed how large RAS active spaces, such as those used for the reference calculations of benzene and phenol monomers, can provide transition energies of monochromophoric systems with absolute errors lower than 0.2 eV compared to experiments. However, when dealing with dichromophoric aggregates, such as benzene/phenol homo- or heterodimers, the use of large active spaces for each monomer results in huge active spaces and unaffordable calculations. Therefore, one is limited to minimal active spaces (*mAS*) for each benzene and phenol monomer (i.e., CAS(6,6) and CAS(8,7), respectively), which will imply the following active spaces for the dimers: CAS(12,12) for the benzene dimer, CAS(16,14) for the phenol dimer, and CAS(14,13) for the benzene-phenol dimer. These active spaces already involve a large number of configuration state functions (CSFs) (e.g.,

736,164 CSFs in the case of CAS(14,13)) and such calculations require a significant computational cost. Moreover, to provide transition energies and dipole moments for simulations of 2DUV spectra, a large number of excited states (up to ~100) have to be included in the state-average procedure, making them computationally very demanding. At the same time, despite including the complete set of π-valence electrons and orbitals, *mAS* does not provide acceptable results for monomers (see Table 2) with the energies of the ionic states being strongly underestimated (up to 0.6 eV for the ¹E_{lu} state of both phenol and benzene and for the ¹B_{1u} and the ²E_{2g} states of benzene). In an effort to minimize the computational cost while preserving a reasonable accuracy of the transition energies, we now explore the possibility of reducing the basis set size and attuning *mAS* calculations to the reference computations. The optimal setups for monomer calculations have to provide 2DUV spectra of benzene and phenol in the gas phase that very closely resemble those obtained with the highest level of theory employed in this study, paving the way to accurate simulations of dimeric systems.

As shown in Table 2, applying the same procedure as outlined for the reference computations but restricting the size of the active space to the minimal active spaces (with full-valence π-orbital set), the energies of the covalent ¹B_{2u} and ¹E_{2g} states of both benzene and phenol are underestimated only slightly and in good agreement with the RAS reference calculations. However, the ¹B_{1u}, ¹E_{lu}, and ²E_{2g} ionic states, as well as all of the doubly excited states, exhibit strong (nonrigid) red-shifts up to 0.60 eV with respect to reference energies. These results indicate that, with the *mAS*, the σ–π polarization effect is too large to be corrected at the perturbation level, resulting in an overstabilization of all ionic states. The removal of the Rydberg-type basis functions from the augmented ANO basis set has a minor effect on the transition energies when the *mAS* active space is employed (see ANO-L(432,21) results in Table 2). Instead, the polarization effect and, consequently, the red-shifts of ionic states with respect to the reference values are partially reduced when using a smaller ANO basis set (i.e.,

ANO-L(321,21)). Therefore, we observe a dramatic effect of the active space size on the excited state energies of each monomer, which is more relevant than the choice of the basis set. An alternative way to standardize *mAS* computations must be found to reduce the computational cost of the transition energy reference calculations.

Here, we propose an approach for obtaining accurate excited state energies of benzene and phenol monomers in the gas phase without increasing the size of the minimal active spaces. The basic idea is to delete a certain number of virtual orbitals in the perturbation treatment and reduce the number of dynamic correlations, aiming at minimizing the mean deviation from the reference values of the transition energies for all states. This procedure implies an increase in the absolute energies of all states that will be different for covalent and ionic states. The $\sigma-\pi$ polarization effects are similar for the covalent ground and valence excited states but are significantly different for ionic excited states. Therefore, a state-dependent blue-shift of the excitation energies is observed when discarding a set of virtual extravalence π^* -orbitals with higher angular momentum in the perturbation procedure (six π^* -orbitals for benzene, seven π^* -orbitals for phenol), following the localization procedure described in the Computational Details. By applying this procedure to the calculations with *mAS* and the ANO-L(321,21) basis set (hereafter named CAS(6,6) $^\delta$ and CAS(8,7) $^\delta$ for benzene and phenol, respectively), we obtained a blue-shift of 0.12–0.29 eV for the covalent ${}^1\text{B}_{2u}$ and ${}^1\text{E}_{2g}$ states and a more pronounced shift of 0.25–0.56 eV for the ${}^1\text{B}_{1u}$, ${}^1\text{E}_{1u}$, and ${}^2\text{E}_{2g}$ ionic states and the ${}^1\text{D}^*$, ${}^2\text{D}^*$, and ${}^3\text{D}^*$ double excitations in both benzene and phenol cases (see Table 2). For benzene, all excited states are between 0.01 and 0.20 eV from the reference energies and much closer to the experimental values than the standard *mAS* calculations. For phenol, all the energies are within 0.29 eV of the reference values except for the low-lying ${}^1\text{E}_{2g}$ state that is 0.51 eV blue-shifted with respect to the reference calculation. This large variation is due to the contemporary good description of the covalent ${}^1\text{E}_{2g}$ states already with the *mAS* (where they red-shift only slightly with respect to the reference values) and the small splitting of the ${}^1\text{E}_{2g}$ doublet (0.33 vs 0.73 eV found in the reference calculation) observed after removing the extravalence π^* -orbitals. Furthermore, the ${}^1\text{B}_{1u}$ doublet splitting is smaller with respect to the reference calculation (0.23 vs 0.51 eV) and in line with the experimental value (\sim 0.23 eV) estimated by Kimura et al.³⁶ As for the transition energies, we observed that the proposed procedure nicely reproduces almost all of the transition dipole moments calculated with the large active spaces, with average deviations of 0.1 au and the largest discrepancies for the high-energy ${}^1\text{B}_{2u} \rightarrow {}^2\text{D}^*$ and ${}^1\text{B}_{2u} \rightarrow {}^3\text{D}^*$ transitions in phenol, being 0.38 and 0.44 au smaller than the reference values (see Table S4 in the SI).

Here, we propose a methodology for calibrating the dynamic correlation by deleting virtual π^* -orbitals from the orbital set. A comparison of the proposed approach with large active spaces calculations, where the originally removed virtual extravalence π^* -orbitals are actually included in the AS, is reported in the SI (see Table S5). The outcome is that the two methodologies (removing virtual π^* -orbitals from the orbital set vs including them in the active spaces) give very similar results, demonstrating that the contribution of these orbitals to the dynamic correlation is overestimated in the perturbation treatment and that the deleting procedure simply reduce the overestimated CASPT2 dynamic correlation when small

(valence) active spaces are employed. Thus, the proposed computational approach can be seen as an alternative to the shift parameters used in CASPT2,^{51,52} as confirmed by testing *mAS* calculations with the real shift parameter set to 0.5 and 0.4 for benzene and phenol monomers, respectively (see Table S5, SI). However, the choice of shift parameter is chromophore-dependent and its application, while simpler with respect to orbital localization and removal procedure, is limited to systems where transition energies of different chromophores can be corrected with the same shift parameter, such as homodimeric systems. Removal of virtual π^* -orbitals from the perturbation treatment also affects the reference calculations of benzene (see Table S6 in the SI), determining a nonuniform shift of the transition energies with a better description of valence states and deterioration of superexcited valence states.

To evaluate the effectiveness of the proposed approach, we have tested the performances of the optimal setups in simulating the (k_1) rephasing signals of 2DUV electronic spectra compared to the standard *mAS* and the *lAS* reference calculations. Figure 2 shows the calculated spectra simulating

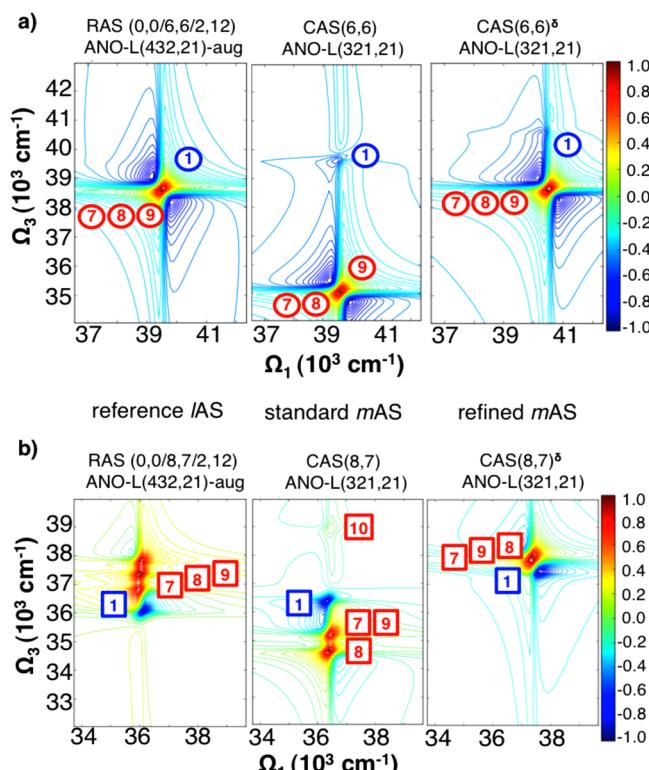


Figure 2. Simulated one-color 2DUV spectra of vapor benzene (a) and phenol (b) in the GS bleaching (negative peak 1) region. Calculations at different levels of theory are shown, highlighting the different positions of the ESAs (positive peaks 7–10) with respect to the GS bleaching. The complex part of the (k_1) rephasing signals with xxxx polarizations is plotted on a logarithmic scale. The signals are labeled according to Table 2.

the one-color 2DUV experiments of vapor benzene (Figure 2a) and phenol (Figure 2b) with identical pump and probe laser pulses centered at the fundamental GS \rightarrow ${}^1\text{B}_{2u}$ transition energies (39521 and 36700 cm^{-1} (i.e., 4.90 and 4.55 eV, respectively)) and with relatively large bandwidths (i.e., full width at half-maximum (fwhm) of 5864 cm^{-1} (corresponding to 5 fs pulses)). The compared spectra of benzene highlight

Table 3. PT2 Vertical GS → S_N Excitation Energies (eV) in the Noninteracting Benzene-Phenol Dimer Calculated at the CAS(14,13)^δ/ANO-L(321,21) Level^a

localized states			dimer states			
	VE	label			VE	
Mixed Double Excitations						
¹ B _{2u}	P	4.66 (4.64)	1P	¹ B _{2u} (B) + ¹ B _{2u} (P)	9.45 [-0.01]	
	B	4.80 (4.81)	1B			
¹ B _{1u}	P	6.17 (5.96)	2P	¹ B _{2u} (B) + ¹ B _{1u} (P)	10.96 [-0.01]	
	B	6.19 (6.13)	2B	¹ B _{1u} (B) + ¹ B _{2u} (P)	10.85 [0.00]	
¹ E _{1u}	P	6.62 (6.63)	3P	¹ B _{2u} (B) + ¹ E _{1u} (P)	11.44 [+0.02]	
	B	6.85 (6.89)	3B	¹ B _{2u} (P) + ¹ E _{1u} (B)	11.57 [+0.06]	
¹ E _{1u}	P	6.85 (6.86)	4P	¹ B _{2u} (B) + ¹ E _{1u} (P)	11.71 [+0.06]	
	B	6.88 (6.89)	4B	¹ B _{2u} (P) + ¹ E _{1u} (B)	11.57 [+0.03]	
¹ ¹ E _{2g}	P	7.45 (7.43)	5P	¹ B _{2u} (B) + ¹ ¹ E _{2g} (P)	12.25 [0.00]	
	B	7.96 (7.94)	5B	¹ B _{2u} (B) + ¹ ¹ E _{2g} (B)	12.49 [-0.01]	
¹ ¹ E _{2g}	P	7.70 (7.76)	6P	¹ B _{2u} (P) + ¹ ¹ E _{2g} (P)	12.61 [-0.01]	
	B	7.96 (7.95)	6B	¹ B _{2u} (P) + ¹ ¹ E _{2g} (B)	12.62 [0.00]	
Charge Transfers						
1D*	P	9.36 (9.29)	7P			
	B	9.77 (9.73)	7B	VE	transition	coeff.
² ¹ E _{2g}	P	9.40 (9.35)	8P	P → B	9.14	0.89
	B	9.82 (9.83)	8B	P → B	9.14	-0.89
² ¹ E _{2g}	P	9.32 (9.33)	9P	B → P	9.93	0.90
	B	9.82 (9.83)	9B	B → P	9.94	0.90
2D*	P	9.87 (9.72)	10P	P → B	10.00	0.89
	B	10.52 (10.44)	10B	P → B	10.01	-0.89
3D*	P	10.53 (10.34)	11P	B → P	10.37	-0.90
	B	11.00 (10.97)	11B	B → P	10.38	0.90

^aEnergies of excited states localized on benzene (B) and phenol (P) are compared with monomer values (reported in parentheses). Differences between mixed double excitations energies and the sums of the corresponding localized states are reported in brackets. Charge transfers from benzene to phenol (B → P) and vice versa (P → B) are reported with their dominant configurations and major CI coefficients, referring to active space orbitals depicted in Figure S2 in the SI. States nomenclature refers to the ideal D_{6h} symmetry of benzene. Labels refer to fundamental GS → ¹B_{2u} transitions of both benzene (transition 1B) and phenol (transition 1P) and ¹B_{2u} → S_N excited state absorptions (transitions 2B-11B and 2P-11P).

how the standard *mAS* calculation significantly red-shifts the 1D* double excitation and the ²¹E_{2g} ionic states that give rise to the ESAs positive signals (peaks 7 and 8–9, respectively), which appear well below the GS bleaching negative signal (peak 1). Upon removal of the extravalence π^* -orbitals, the CAS(6,6)^δ *mAS* calculation instead provides a 2D spectrum of gas-phase benzene that is very similar to that obtained with the reference *lAS* calculation, where the ESAs signals overlap with the weak bleaching signal. It is necessary to mention that in order to show more realistic representations of the gas-phase benzene spectra, the fundamental GS → ¹B_{2u} transition energy for the reference calculation has been blue-shifted to the experimental value (4.90 eV (i.e., 39521 cm⁻¹)) and by the same amount of energy (~0.23 eV) for the other levels of theory (i.e., to 4.92 and 5.04 eV for CAS(6,6) and CAS(6,6)^δ, respectively) for the sake of comparison. The GS → ¹B_{2u} |TDML was also increased (from zero) to 0.07 au (corresponding to $f = 6 \times 10^{-4}$) to reproduce the experimental estimation ($f = 6.4 \times 10^{-4}$).³⁸ As for benzene, the standard *mAS* calculation of phenol yields substantially different 2D spectrum with respect to the reference *lAS* calculation. The 7–9 ESAs are in fact significantly red-shifted, lying below the bleaching signal and with the 1D* double excitation (peak 8) located in between the two ²¹E_{2g} ionic states (peaks 7,9). Moreover, the red-shift also brings down the 2D* double excitation (peak 10), making it appear at 2580 cm⁻¹ above the GS bleaching. The refined minimal CAS(8,7)^δ active space calculation restores the

overlap between the ESAs and the GS bleaching observed in the reference calculation with a slight blue-shift of all signals. Overall, Figure 2 clearly shows that the proposed approach provides good reproduction of the expensive reference RAS calculations, allowing for extension to dimeric systems at a reasonable computational cost.

Noninteracting Benzene-Phenol Dimer. The protocol proposed for benzene and phenol monomers, consisting of removing the extravalence π^* -orbitals with a higher angular momentum from the set of molecular orbitals, can be extended to dimeric systems, such as the homodimers or the benzene-phenol dimer. In this section, we apply this approach to a noninteracting benzene-phenol dimer in the gas phase, where the two chromophores are oriented arbitrarily in space (i.e., without symmetry constraints) and separated by a distance of ~10 Å. This model allows for testing the method accuracy with respect to the monomer calculations and exploring other possibilities for further reduction of the computational cost.

The localization/delocalization procedure was applied to the noninteracting benzene-phenol dimer in the gas phase using the CAS(14,13) *mAS* and ANO-L(321,21) basis set (hereafter named CAS(14,13)^δ level). The resulting vertical excitations at PT2 level are reported in Table 3. As expected, the noninteracting dimer excited states localized on the single chromophores (localized excited states) show small deviations from the isolated monomers with average absolute deviation of 0.05 eV, except for the ¹B_{1u} state of phenol, which blue-shifts by

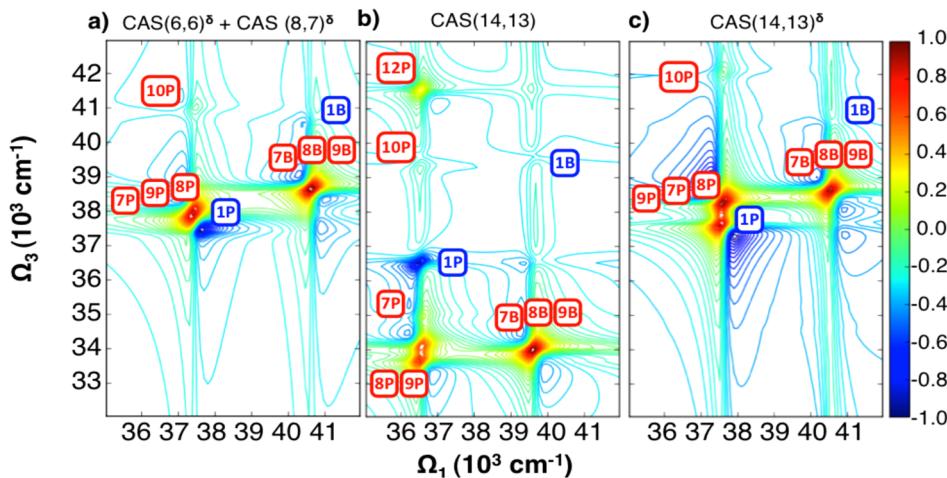


Figure 3. Simulated one-color 2DUV spectra of the noninteracting benzene-phenol dimer. The combined spectrum of monomers (a) at the $\text{CAS}(6,6)^{\delta}$ and $\text{CAS}(8,7)^{\delta}$ levels is compared with the dimer spectra at the (b) standard *mAS* $\text{CAS}(14,13)$ and (c) refined $\text{CAS}(14,13)^{\delta}$ levels. Pump and probe pulses have fwhm of 1466 and 10000 cm^{-1} , respectively. Pulse frequencies have been centered right below the benzene GS $\rightarrow {}^1\text{B}_{2u}$ transition energies at (b) 39000 cm^{-1} and (a,c) 40000 cm^{-1} . The complex part of the k_1 rephasing signals with xxxx polarizations is plotted on a logarithmic scale. The signals are labeled according to Table 3.

$\sim 0.20\text{ eV}$ becoming nearly degenerate with the analogous benzene state and the phenol doubly superexcited states (2D^* and 3D^*) that blue-shift by $0.15\text{--}0.19\text{ eV}$. Along with localized excited states, the electronic structure of the dimer involved several mixed doubly excited states, constructed by the collective one-electron excitations on both chromophores and appearing at energies given by the sum of the localized single excitations. If the two chromophores are electronically interacting, the energies of these mixed states are shifted with respect to the sum of the corresponding single excitations. This electronic coupling, denoted as “quartic” coupling,^{2,53} is expected to be absent or negligible in the noninteracting benzene-phenol dimer. To account for the numerous mixed doubly excited states arising from the different combinations of single excitations, several states have to be included in the state-average CASSCF calculations (e.g., the reported data for the dimer include 100 excited states). Table 3 shows that most of the mixed double excitation energies deviate only slightly from the expected values, demonstrating minor deterioration with respect to the monomer calculations and validating the proposed localization/delocalization procedure. The mixed doubly excited states involving the ${}^1\text{E}_{1u}$ states show the largest deviations (by at most $\sim 0.06\text{ eV}$) that, although small in absolute value, could give rise to small artificial quartic couplings. Regardless, these states are located at energies higher than 11.5 eV (i.e., well beyond the probing windows of 2DUV spectra). Charge-transfer (CT) states are also present in the electronic structure of the dimeric system, characterized by single excitations from molecular orbitals localized on one chromophore to those of the other chromophore. As expected, the CT states in the noninteracting benzene-phenol dimer are located above the ionization limits of the chromophores, and all the GS \rightarrow CT and ${}^1\text{B}_{2u}(\text{B},\text{P}) \rightarrow$ CT transitions are dark.

The $\text{CAS}(14,13)^{\delta}$ approach has been tested by simulating the 2DUV electronic spectra of the noninteracting dimer and comparing them with the standard $\text{CAS}(14,13)$ *mAS* and the combined $\text{CAS}(6,6)^{\delta}$ and $\text{CAS}(8,7)^{\delta}$ monomer calculations. Figure 3 shows the calculated (k_1) rephasing signals of the one-color 2DUV spectra of the noninteracting dimer in the gas phase. To distinguish the benzene signals (with low oscillator

strength) in the dimer spectra, the pump laser pulses have been tuned right below the benzene GS $\rightarrow {}^1\text{B}_{2u}$ transition frequencies with narrowband pump pulses (fwhm of 1466 cm^{-1}). Central frequencies of 40000 cm^{-1} have been used for the $\text{CAS}(14,13)^{\delta}$ and the combined $\text{CAS}(6,6)^{\delta}$ and $\text{CAS}(8,7)^{\delta}$ simulations for which benzene GS $\rightarrow {}^1\text{B}_{2u}$ transition frequencies were at 40570 cm^{-1} (5.03 eV) and 40650 cm^{-1} (5.04 eV), respectively. For the standard $\text{CAS}(14,13)$ simulations, central frequencies were set to 39000 cm^{-1} with fundamental benzene transition at 39618 cm^{-1} (4.91 eV). All benzene GS $\rightarrow {}^1\text{B}_{2u}$ transitions were blue-shifted by 0.23 eV , as in the gas-phase benzene reference spectra shown in Figure 2. The probe pulses, in contrast, have large bandwidths (fwhm of 10000 cm^{-1}) in order to properly resolve the ESA signals characteristic of each chromophore. The comparison clearly shows how the standard *mAS* calculations do not reproduce the spectrum of the ideal noninteracting dimer calculated as the sum of the monomer spectra obtained at the $\text{CAS}(6,6)^{\delta}$ and $\text{CAS}(8,7)^{\delta}$ levels with all of the ESAs strongly red-shifted. Instead, at the $\text{CAS}(14,13)^{\delta}$ level, the dimer spectrum is almost identical to the sum of the monomer spectra with the ESAs positive signals rising from the 1D^* double excitation and the 2E_{2g} ionic states of both benzene (peaks 7B and 8–9B, respectively) and phenol (peaks 7P and 8–9P, respectively) overlapping with the phenol GSB (peak 1P) and the benzene GSB (peak 1B), respectively. A slight blue-shift of the 2D^* superexcited state of phenol (peak 10P) is observed at $\text{CAS}(14,13)^{\delta}$ without significantly changing the overall spectrum. The absence of off-diagonal crossing peaks arising from a non-negligible quartic coupling between the ${}^1\text{B}_{2u}(\text{B})$ and ${}^1\text{B}_{2u}(\text{P})$ states is indicative of a good description of mixed doubly excited states.

To further decrease the computational cost of the proposed methodology for dimeric systems, we have examined the possibility of reducing the size of the multiconfigurational problem. The $\text{CAS}(14,13)^{\delta}$ treatment of the benzene-phenol dimer, in fact, still involves a huge number of CFSs (i.e., 736,164 CFSs), which can be significantly decreased by reduction (or restriction) of the active spaces. Because the wave function analysis of monomers (see Table 1) shows that

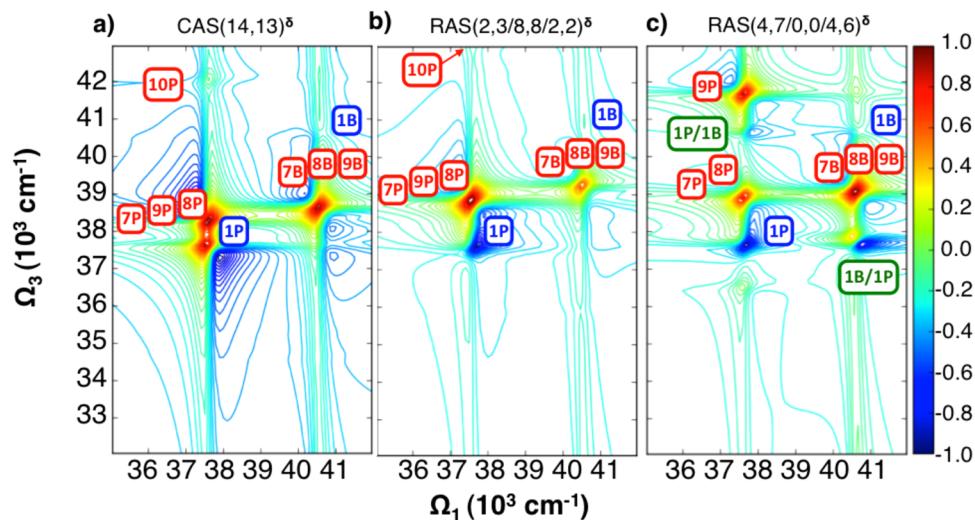


Figure 4. Comparison of simulated one-color 2DUV spectra of the noninteracting benzene-phenol dimer with different active spaces. Pump and probe pulses have fwhm of 1466 and 10000 cm^{-1} , respectively, with central frequencies set right below the benzene GS \rightarrow ${}^1\text{B}_{2u}$ transition energy at 40000 cm^{-1} . The complex part of the (k_l) rephasing signals with xxxx polarizations is plotted on a logarithmic scale. The signals are labeled according to Table 3, and crosspeaks arising from “artificial” quartic coupling are indicated as 1B/1P and 1P/1B.

the frontier π -orbitals (i.e., the three highest occupied π -orbitals and the three lowest unoccupied π^* -orbitals) are essential for the description of singly and doubly excited state manifolds of the chromophores, an active space comprising all these orbitals (i.e., at least a (12,12) active space for a dimer) is mandatory for resolving the complete manifold of high-lying excited states of dimeric systems. However, a CAS(12,12) active space would not include the phenol oxygen lone pair, which contributes with non-negligible CI coefficients in the ${}^2\text{E}_{2g}$ states wave functions, leading to a large blue-shift of these states, near degeneracy and erroneous mixing with the ${}^2\text{E}_{2g}$ states of benzene. The high-lying ${}^2\text{E}_{2g}$ doublet of phenol that appears in the one-color 2DUV spectra of the dimer (peaks 8–9P in Figure 3), and it represents the main spectroscopic fingerprint of phenol in the UV probing window. Therefore, complete active spaces smaller than (14,13) would not provide accurate results. The RAS technique can be used to significantly reduce the computational cost by restricting the total number of simultaneously excited electrons within the active orbitals, maintaining the total number of active electrons while reducing the number of CSFs. Here, we have employed RAS schemes that maintain the 14 active electrons and 13 active orbitals of the (14,13) mAS while significantly reducing the number of CSFs, such as RAS(2,3/8,8/2,2) and RAS(4,7/0,0/4,6). In these RAS calculations, the localization/delocalization procedure and the removal of virtual extravalence π^* -orbitals has been applied, as in the CAS(14,13) $^\delta$ treatment. In the RAS(2,3/8,8/2,2) scheme, a complete (8,8) active subspace is employed with single and double excitations allowed from the three occupied orbitals in RAS1 to the eight orbitals in RAS2 and the two RAS3 unoccupied orbitals. This RAS scheme provides very interesting results with PT2 energies of localized, mixed doubly excited state energies and CT states that deviate in average from the CAS(14,13) $^\delta$ values by \sim 0.06 eV. The major differences were found for phenol localized states with ${}^1\text{B}_{1u}$ (P) red-shifted by 0.17 eV (lying at 6.02 eV, very close to the experimental value of 5.77–5.82 eV) and the high-lying ${}^1\text{E}_{2g}$ and superexcited 2D* states blue-shifted by 0.19 and 0.17 eV, respectively. Common mixed states, indicative of interchromophore interactions via electronic coupling, are calculated with the

same accuracy as in CAS(14,13) $^\delta$ with deviations from the expected values within 0.06 eV. As shown in Figure 4, the differences between RAS(2,3/8,8/2,2) and CAS(14,13) $^\delta$ transition energies do not significantly affect the one-color 2DUV spectrum of the noninteracting dimer, which appears very similar in the two cases, except for the slight blue-shift of the phenol superexcited 2D* state (peak 10P). The use of the RAS(2,3/8,8/2,2) scheme reduces the number of CSFs by a factor of 3.3 with respect to CAS(14,13) $^\delta$ (i.e., from 736,164 to 219,048 CSFs) with a gain of computer time preserving a reliable one-color 2DUV spectrum.

The RAS(4,7/0,0/4,6) scheme does not include any complete active subspace (i.e., unoccupied RAS2), with all seven occupied orbitals located in RAS1, where up to four holes are allowed, and all six unoccupied orbitals present in RAS3, where up to four electrons can be allocated. This RAS scheme reduces the number of CSFs by an order of magnitude with respect to the CAS(14,13) $^\delta$ mAS calculations, decreasing up to 52,641. Interestingly, most of the transition energies at the RAS(4,7/0,0/4,6) level show nice agreement with the RAS(2,3/8,8/2,2) and the CAS(14,13) $^\delta$ ones, with average absolute energy deviations of 0.01 and 0.08 eV, including CT states, that show largest deviations of 0.03 and 0.04 eV, respectively. However, although this RAS scheme is very appealing because it significantly reduces the computational cost of the CAS(14,13) $^\delta$ calculations, some relevant differences worsen the one-color 2DUV spectrum. In particular, at the RAS(4,7/0,0/4,6) level, the high-lying ${}^2\text{E}_{2g}$ state of phenol is significantly blue-shifted with respect to the RAS(2,3/8,8/2,2) and CAS(14,13) $^\delta$ calculations (by 0.35 and 0.43 eV, respectively), and the corresponding ESA signal (peak 9P) is no longer overlapping with the GS bleaching but rather lies well above it, appearing as a distinct peak (see Figure 4c). More importantly, the mixed doubly excited states energies at the RAS(4,7/0,0/4,6) level are not as accurate as in RAS(2,3/8,8/2,2) and CAS(14,13) $^\delta$, causing the appearance of off-diagonal crossing peaks associated with an “artificial” quartic coupling between the ${}^1\text{B}_{2u}$ (B) and ${}^1\text{B}_{2u}$ (P) states (see peaks 1B/1P and 1P/1B in Figure 4c). In summary, the RAS(2,3/8,8/2,2) scheme is best suited for generating the excited state spectrum of the benzene-phenol

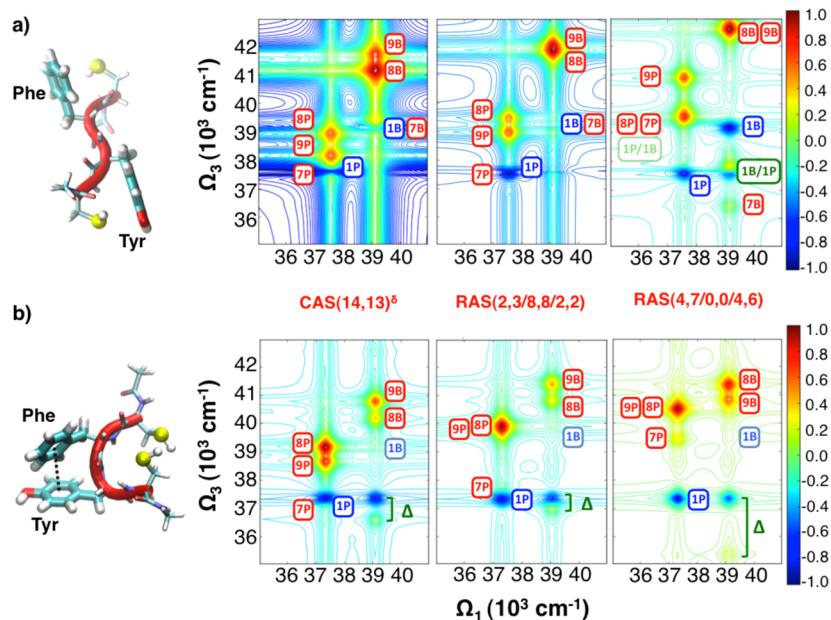


Figure 5. Comparison of simulated one-color 2DUV spectra of the CFYC tetrapeptide in solution. Pump and probe pulses with fwhm of 733 and 2932 cm⁻¹ have been used, respectively, with central frequencies set right below the benzene GS → ¹B_{2u} transition energies at 39000 cm⁻¹. The complex part of the ($k_I + k_{II}$) quasi-absorptive signals with xxxx polarizations is plotted on a logarithmic scale. The signals are labeled according to Table 3 with quartic couplings indicated with Δ and crosspeaks arising from “artificial” coupling indicated as 1B/1P and 1P/1B.

heterodimer with a significant gain of computational time, whereas RAS(4,7/0,0/4,6) should be avoided if the energy window of interest involves mixed doubly excited states or superexcited valence states (i.e., with probe pulses in the UV region).

Modeling Chromophore Stacking in a Tetrapeptide.

The calibration results given in the previous section demonstrate how the computational cost of simulating 2D spectra of a benzene-phenol dimeric species in vacuo can be considerably reduced by employing RAS schemes without significant loss of accuracy. Here, we apply this approach to the CFYC tetrapeptide in solution, a realistic system, and a protein model containing two aromatic residues (i.e., the F and Y residues) with a phenyl and a phenolic group as chromophores, respectively. As we have previously shown,^{4,5} the use of the CAS(14,13)^δ level within an SOS//QM/MM approach allows us to simulate 2D electronic spectra of the CFYC tetrapeptide that contains clear spectroscopic fingerprints characterizing different folding states, an unfolded state with noninteracting chromophores, and a folded one with T-stacked chromophores. We aim to find the most appropriate RAS method that allows accurate and fast simulations of one-color (UV–UV) and two-color (UV–vis) 2D spectra, which should provide clear distinction between unstacked and T-stacked conformations of aromatic residues in the CFYC tetrapeptide. Therefore, we compare the performances of RAS(2,3/8,8/2,2) and RAS(4,7/0,0/4,6) methods with respect to the CAS(14,13)^δ level, analyzing two structures selected from the unbiased unfolding dynamics previously reported (see also the Computational Details section) and representing tetrapeptide configurations with interacting and noninteracting aromatic chromophores (Figure 5).

Figure 5 shows the one-color 2DUV spectra (in the probe window between 35000 and 43000 cm⁻¹) in the unfolded (Figure 5a) and folded (Figure 5b) configurations calculated at different levels of theory. The pump pulses were set at 39000

cm^{-1} (4.83 eV), right below the calculated benzene GS \rightarrow ${}^1\text{B}_{2u}$ transition frequencies (e.g., 39115 cm^{-1} (4.85 eV) at the CAS(14,13) $^\delta$ level), which were found to be very close to the experimental values of benzene molecules in water solution (reported at $\sim 39390 \text{ cm}^{-1}$ (4.88 eV)).⁵⁴ Narrowband pump and probe pulses with fwhm of 733 and 2932 cm^{-1} , respectively, were necessary to properly resolve all the signals in the calculated one-color 2DUV spectra. Analogously to the noninteracting dimer in vacuo (Figure 4), the one-color 2DUV spectrum of the unfolded configuration in solution calculated at the CAS(14,13) $^\delta$ level (Figure 5a) contains ESAs signals rising from 1D* double excitations (peaks 7B and 7P) and 2^1E_{2g} ionic states (peaks 8–9B and 8–9P) of both chromophores. While the double excitation signals still overlap with the GSBs (peaks 1P and 1B) as in vacuo, the ionic states lose their degeneracy and slightly blue-shift with respect to the gas phase. An analogous blue-shift due to solvent effects is observed for the phenol superexcited 2D* state (peak 10P), which disappears from the selected probing window. Notably, the comparison between the CAS(14,13) $^\delta$ level and the cheaper RAS(2,3/8,8/2,2) and RAS(4,7/0,0/4,6) methods shows analogies to what was observed in vacuo (and reported in Figure 4). In fact, although the RAS(2,3/8,8/2,2) simulations provide a one-color 2DUV spectrum very similar to that obtained at the CAS(14,13) $^\delta$ level, the RAS(4,7/0,0/4,6) results are unsatisfactory. Similarly to the gas phase, the 9P signal is significantly blue-shifted and an “artificial” quartic coupling is observed in the solution spectrum at the RAS(4,7/0,0/4,6) level. Moreover, the separation between the ESAs of phenylalanine (peaks 7B and 8–9B) is overestimated at the RAS(4,7/0,0/4,6) level, further deteriorating the 2DUV spectrum. Therefore, the RAS(2,3/8,8/2,2) scheme is confirmed to be the minimal RAS scheme that can provide qualitatively good one-color 2DUV spectrum for a dimeric system with noninteracting chromophores, as in the gas phase (Figure 4) as in solution (Figure 5a).

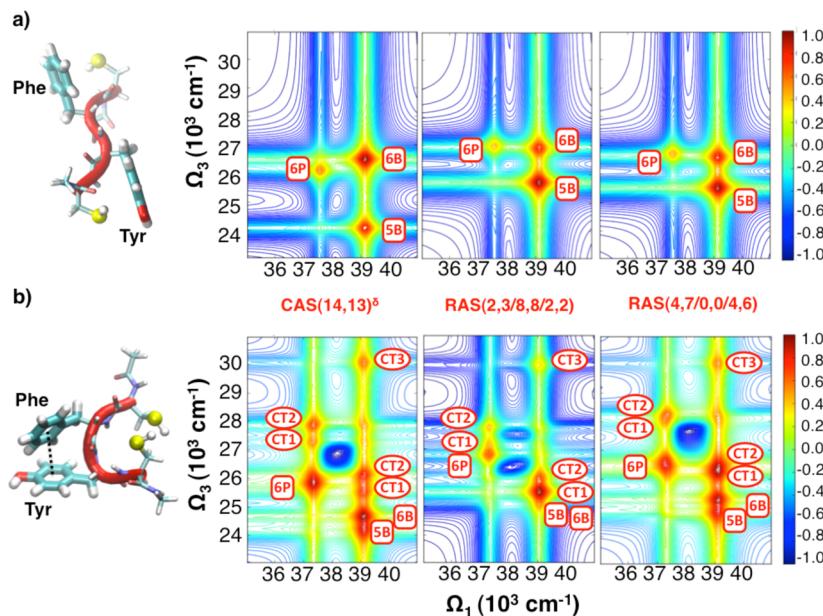


Figure 6. Comparison of simulated two-color 2DUV–vis spectra of the CFYC tetrapeptide in solution. Pump and probe pulses with fwhm of 733 and 5864 cm^{-1} have been used, respectively, with central frequencies set at 39000 and 28000 cm^{-1} , respectively. The complex part of the ($k_1 + k_{\text{II}}$) quasi-absorptive signals with xxxx polarizations is plotted on a logarithmic scale. The signals are labeled according to Table 3, including numbered charge transfer (CT) states.

To test the performances of the RAS schemes for dimeric systems with interacting chromophores, we extended the analysis to the case of the folded CFYC tetrapeptide with T-stacked aromatic side chains. Figure 5b shows the comparison of the one-color 2DUV spectra of the folded tetrapeptides calculated at different levels of theory. It is worth noting that, as previously reported,^{4,5} the interchromophore interactions introduce, in the one-color 2DUV spectrum of the tetrapeptide, an off-diagonal crossing peak arising from the quartic coupling between the fundamental transitions of the chromophores. This characteristic feature of the T-stacked conformation is properly captured at the CAS(14,13)^δ level, where the quartic splitting (Δ) is calculated to be $\sim 759 \text{ cm}^{-1}$. At the RAS(2,3/8,8/2,2) level, the quartic splitting is only slightly underestimated ($\Delta = 348 \text{ cm}^{-1}$), whereas at the RAS(4,7/0,0/4,6) level, the presence of the “artificial” coupling, already found in the noninteracting gas-phase dimer and in the unfolded tetrapeptide, causes an overestimation of the splitting with $\Delta = \sim 2200 \text{ cm}^{-1}$ (see Figure 5b). Therefore, as for the case of the unfolded tetrapeptide, the RAS(4,7/0,0/4,6) results appear to be inappropriate for reproducing the results of higher levels of theory, and the RAS(2,3/8,8/2,2) scheme can be used instead of the CAS(14,13)^δ level to simulate the one-color 2DUV spectra of the CFYC tetrapeptide in solution with a significant reduction of computational cost.

The most compelling signature of coupled chromophore aggregates in the CFYC model is the presence of charge-transfer (CT) state signals in the visible probing region of the 2D electronic spectra.^{4,5} In fact, CT states of the CFYC tetrapeptide are dark in the unfolded conformation but significantly red-shift, and transitions ${}^1\text{B}_{2u}(\text{B},\text{P}) \rightarrow \text{CT}$ gain brightness in folded configurations upon interchromophore interactions. One-electron excitations involving CT states can be found by probing in the visible region, a spectral region below the ionization potential of the chromophores, where background-free signals can be more clearly detected. Therefore, here we focus our attention on the use of RAS schemes to

simulate two-color 2DUV–vis spectra of the CFYC tetrapeptide.

Figure 6 shows the two-color 2DUV–vis spectra of the two selected conformations of the CFYC tetrapeptide at different levels of theory. Pump and probe pulses with fwhm of 733 and 5864 cm^{-1} , respectively, were used to resolve the 2D signals in the calculated two-color spectra. Probe pulses have their central frequency set to 28000 cm^{-1} (3.47 eV). As expected, for the unfolded tetrapeptide with noninteracting chromophores, CT signals are not found in the two-color 2D spectrum (Figure 6a) and only low-lying ESAs can be detected. In particular, in the selected probing window between 23000 and 31000 cm^{-1} , only ESAs involving the ${}^1\text{E}_{2g}$ states of both chromophores can be found (i.e., peaks 5P and 5–6B for Y and F residues, respectively). Interestingly, the 2DUV–vis spectrum at the CAS(14,13)^δ level seems to be nicely reproduced by both RAS(2,3/8,8/2,2) and RAS(4,7/0,0/4,6) schemes, except for a slight blue-shift of the 5B signal at the RAS levels.

The 2DUV–vis spectra of the folded CFYC tetrapeptide are depicted in Figure 6b. Here, the ${}^1\text{E}_{2g}$ ESAs signals are almost unchanged with respect to the unfolded structure (except for peak 6B, which is significantly red-shifted close to 5B) and is accompanied by the appearance of three signals associated with bright ${}^1\text{B}_{2u}(\text{B},\text{P}) \rightarrow \text{CT}$ transitions (CT1–3). It is necessary to note that transition energies of CT states are not experimentally available for benzene-phenol or analogous dimeric/oligomeric systems and that their energies are strongly dependent on the relative geometric orientation and distance of the chromophores. Thus, because no experimental reference can be used to benchmark CT transition energies, here we assume that our methodology is able to describe CT states with the same accuracy as the local excitations. As local excitations with ionic character, in fact, CT states show strong dependence on the amount of dynamic correlation and exhibit a significant blue-shift upon discarding virtual orbitals in the perturbation treatment. For their nature, CT states are accessible from ${}^1\text{B}_{2u}$ states of both aromatic side chains. Thus, the spectroscopic

signatures of these CT states would appear twice in the 2D spectra of CFYC at two different pump frequencies corresponding to the ${}^1\text{B}_{2u}$ states of Y and F (i.e., $\Omega_1(\text{Y}) = 37331 \text{ cm}^{-1}$ and $\Omega_1(\text{F}) = 39115 \text{ cm}^{-1}$, respectively) and at “constantly” different probe frequencies (i.e., rigidly shifted along Ω_3 by the energy difference between the two ${}^1\text{B}_{2u}$ states). The CAS(14,13) $^\delta$ 2D spectrum of the folded structure, indeed, is characterized by the presence of CT1 and CT2 states at probing frequencies of 27276 and 27885 cm^{-1} , respectively, for $\Omega_1(\text{Y})$ and 25492 and 26101 cm^{-1} , respectively, for $\Omega_1(\text{F})$. In the selected probing window, a third CT state (CT3) is predicted to appear at $\Omega_3 = 30093 \text{ cm}^{-1}$ and $\Omega_1(\text{F})$. Notably, all of these features are found in both RAS(2,3/8,8/2,2) and RAS(4,7/0,0/4,6) 2DUV-vis spectra (Figure 6b), providing similar spectra with respect to the higher CAS(14,13) $^\delta$ level of theory.

In summary, the proposed RAS schemes are both suitable for reproducing 2DUV-vis spectra of unfolded and folded CFYC tetrapeptides with a decrease in CSFs by an order of magnitude with respect to the CAS(14,13) $^\delta$ calculations in the RAS(4,7/0,0/4,6) case and a tremendous reduction of the computational cost.

CONCLUSIONS

Benzene, phenol, and their dimer are minimal models for the study of electronic coupling of UV-chromophores in proteins. They have been targeted in this work for developing a protocol that allows accurate and, at the same time, computationally affordable simulations of 2DUV spectra based on extensive multiconfigurational/multireference calculations. By employing large active spaces within a restricted active space (RAS) scheme, we have initially set the reference methodology for reproducing the experimental transition energies of high-lying excited states of the monomeric UV-chromophores in the gas phase, including superexcited states above their ionization limits. Good agreement with previously reported calculations and experimental data on valence states below the ionization limits has been observed, allowing for characterization of previously unassigned superexcited states. The outcome of this study also provided transition energies and dipole moments that could be used in the future for developing more accurate Frenkel exciton Hamiltonians than those currently available. Restricted large active state calculations of the aromatic monomers have then been used to set the reference 2DUV spectra for benchmarking the use of less expensive minimal active spaces to treat a gas-phase noninteracting benzene-phenol dimer. We have shown that a calibration of the dynamic correlation (by deleting virtual π^* -orbitals from the orbital set of the multiconfigurational treatment) introduced with the multireference perturbation treatment can provide the required accuracy to properly simulate the 2DUV spectrum of the noninteracting dimer while using minimal complete active spaces. In an effort to further reduce the computational cost of the minimal complete active space calculations (involving 736,164 configuration state functions, CSF) we have employed different RAS schemes with a reduced number of CSFs. We found that the RAS(2,3/8,8/2,2) scheme (with 219,048 CSF) provides a reliable one-color spectrum (i.e., both pump and probe pulses in the UV region) of the noninteracting dimer, whereas the RAS(4,7/0,0/4,6) scheme, with only 52,641 CSFs, could not properly describe mixed doubly excited states and superexcited valence states located in the UV probing region. Finally, the restricted minimal active spaces have been tested

within a hybrid QM/MM scheme on a realistic proteic model containing the benzene-phenol dimer, the CFYC tetrapeptide in solution. Interestingly, we found that the RAS(2,3/8,8/2,2) scheme can be successfully used to simulate not only the one-color 2DUV spectrum of the unfolded tetrapeptide (with noninteracting UV-chromophores) but also the spectrum of the folded configuration with T-stacked and strongly interacting aromatic side-chains, suggesting that this approach can be used to speed up simulations of one-color 2DUV spectroscopy, tracking the folding/unfolding dynamics of CFYC in solution. However, the most compelling spectroscopic fingerprints of electronic couplings between aromatic side chains are held in the charge transfer states (CT), which red-shift and become bright upon interchromophore interactions and appear in two-color 2DUV spectra (with probe pulses in the visible region). Notably, we observed that the RAS(4,7/0,0/4,6) scheme provides reliable two-color 2DUV spectra of both unstacked and T-stacked configurations of the CFYC tetrapeptide, allowing reduction of more than 1 order of magnitude in the number of CSFs and a significant gain in computational time. In summary, our gas-phase results provide basic ingredients for the development of new Frenkel exciton Hamiltonians of aromatic side chains that include several high-lying excited states and allow scale-up of 2DUV spectroscopy simulations to large proteic systems. The proposed recipe for reducing the computational cost of multiconfigurational calculations of the excited state manifolds of dimeric aromatic systems assures accurate and computationally affordable simulations of their nonlinear 2DUV spectroscopy. The opportunity for performing extensive multiconfigurational computations within the hybrid SOS//QM/MM scheme is open to theoretical characterization solvents and thermally induced line broadenings in 2DUV spectra and for direct comparisons with future experiments.

ASSOCIATED CONTENT

S Supporting Information

Active space orbitals, Cartesian coordinates, transition dipole moments, and excitation energy benchmarks. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00443.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: artur.nenov@unibo.it.

*E-mail: marco.garavelli@unibo.it or marco.garavelli@ens-lyon.fr.

*E-mail: ivan.rivalta@ens-lyon.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

S.M. gratefully acknowledges the support of the National Science Foundation (grant CHE-1361516), and of the Chemical Sciences, Geosciences, and Biosciences division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy. M.G. acknowledges support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). I.R. gratefully acknowledges the support of the École Normale Supérieure de Lyon (Fonds Recherche 900/S81/BS81-FR14). We acknowledge the use of HPC resources of the “Pôle Scientifique de Modélisation Numérique” at the ENS-Lyon, France.

■ ABBREVIATIONS

phenylalanine, Phe; F; tyrosine, Tyr, Y; ultraviolet, UV; two-dimensional, 2D; two-dimensional electronic spectroscopy, 2DES; two-dimensional ultraviolet spectroscopy, 2DUV; quantum mechanics, QM; quantum mechanics/molecular mechanics, QM/MM; sum-over-states, SOS; charge transfer, CT; double excitations, D; restricted active space self-consistent field, RASSCF or RAS; multireference perturbation theory, PT2; complete active space self-consistent field, CASSCF or CAS; cysteine-phenylalanine-tyrosine-cysteine, CFYC; transition dipole moments, TDM; single state, SS; state-average, SA; ionization-potential-electron-affinity, IPEA; minimal active spaces, mAS; large active spaces, lAS; local oscillator, LO; ground state bleach, GSB; excited state absorption, ESA; stimulated emission, SE; molecular orbital, MO; full width at half-maximum, fwhm

■ REFERENCES

- (1) Trinquier, G.; Sanejouand, Y. H. Which Effective Property of Amino Acids Is Best Preserved by the Genetic Code? *Protein Eng., Des. Sel.* **1998**, *11*, 153–169.
- (2) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford University Press: New York, 1995.
- (3) Mukamel, S. Multidimensional Femtosecond Correlation Spectroscopies of Electronic and Vibrational Excitations. *Annu. Rev. Phys. Chem.* **2000**, *51*, 691–729.
- (4) Nenov, A.; Rivalta, I.; Cerullo, G.; Mukamel, S.; Garavelli, M. Disentangling Peptide Configurations Via Two-Dimensional Electronic Spectroscopy: Ab Initio Simulations Beyond the Frenkel Exciton Hamiltonian. *J. Phys. Chem. Lett.* **2014**, *5*, 767–771.
- (5) Nenov, A.; Beccara, S. A.; Rivalta, I.; Cerullo, G.; Mukamel, S.; Garavelli, M. Tracking Conformational Dynamics of Polypeptides by Nonlinear Electronic Spectroscopy of Aromatic Residues: A First-Principles Simulation Study. *ChemPhysChem* **2014**, *15*, 3282–3290.
- (6) Ginsberg, N. S.; Cheng, Y.-C.; Fleming, G. R. Two-Dimensional Electronic Spectroscopy of Molecular Aggregates. *Acc. Chem. Res.* **2009**, *42*, 1352–1363.
- (7) West, B. A.; Moran, A. M. Two-Dimensional Electronic Spectroscopy in the Ultraviolet Wavelength Range. *J. Phys. Chem. Lett.* **2012**, *3*, 2575–2581.
- (8) Krebs, N.; Pugliesi, I.; Hauer, J.; Riedle, E. Two-Dimensional Fourier Transform Spectroscopy in the Ultraviolet with Sub-20fs Pump Pulses and 250,720nm Supercontinuum Probe. *New J. Phys.* **2013**, *15*, 085016–085032.
- (9) Consani, C.; Auböck, G.; van Mourik, F.; Chergui, M. Ultrafast Tryptophan-to-Heme Electron Transfer in Myoglobins Revealed by UV 2D Spectroscopy. *Science* **2013**, *339*, 1586–1589.
- (10) Mukamel, S.; Abramavicius, D. Many-Body Approaches for Simulating Coherent Nonlinear Spectroscopies of Electronic and Vibrational Excitons. *Chem. Rev.* **2004**, *104*, 2073–2098.
- (11) Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. Ultraviolet Spectroscopy of Protein Backbone Transitions in Aqueous Solution: Combined Qm and Mm Simulations. *J. Phys. Chem. B* **2010**, *114*, 8270–8277.
- (12) Rivalta, I.; Nenov, A.; Cerullo, G.; Mukamel, S.; Garavelli, M. Ab Initio Simulations of Two-Dimensional Electronic Spectra: The SOS//Qm/Mm Approach. *Int. J. Quantum Chem.* **2014**, *114*, 85–93.
- (13) Nenov, A.; Rivalta, I.; Mukamel, S.; Garavelli, M. Bidimensional Electronic Spectroscopy on Indole in Gas Phase and in Water from First Principles. *Comput. Theor. Chem.* **2014**, *1040*, 295–303.
- (14) Stolow, A.; Bragg, A. E.; Neumark, D. M. Femtosecond Time-Resolved Photoelectron Spectroscopy. *Chem. Rev.* **2004**, *104*, 1719–1757.
- (15) Schick, C. P.; Weber, P. M. Ultrafast Dynamics in the Three-Photon, Double-Resonance Ionization of Phenol Via the S-2 Electronic State. *J. Phys. Chem. A* **2001**, *105*, 3735–3740.
- (16) Schick, C. P.; Weber, P. M. Ultrafast Dynamics in Superexcited States of Phenol. *J. Phys. Chem. A* **2001**, *105*, 3725–3734.
- (17) Nakashima, N.; Sumitani, M.; Ohmine, I.; Yoshihara, K. Nanosecond Laser Photolysis of the Benzene Monomer and Eximer. *J. Chem. Phys.* **1980**, *72*, 2226–2230.
- (18) Platt, J. R. Classification of Spectra of Cata-Condensed Hydrocarbons. *J. Chem. Phys.* **1949**, *17*, 484–495.
- (19) Karlstrom, G.; Lindh, R.; Malmqvist, P. A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P. O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. Molcas: A Program Package for Computational Chemistry. *Comput. Mater. Sci.* **2003**, *28*, 222–239.
- (20) Aquilante, F.; De Vico, L.; Ferre, N.; Ghigo, G.; Malmqvist, P. A.; Neogrady, P.; Pedersen, T. B.; Pitonak, M.; Reiher, M.; Roos, B. O.; Serrano-Andres, L.; Urban, M.; Veryazov, V.; Lindh, R. Software News and Update Molcas 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224–247.
- (21) Ghigo, G.; Roos, B. O.; Malmqvist, P. A. A Modified Definition of the Zeroth-Order Hamiltonian in Multiconfigurational Perturbation Theory (Caspt2). *Chem. Phys. Lett.* **2004**, *396*, 142–149.
- (22) Forsberg, N.; Malmqvist, P. A. Multiconfiguration Perturbation Theory with Imaginary Level Shift. *Chem. Phys. Lett.* **1997**, *274*, 196–204.
- (23) Aquilante, F.; Lindh, R.; Pedersen, T. B. Unbiased Auxiliary Basis Sets for Accurate Two-Electron Integral Approximations. *J. Chem. Phys.* **2007**, *127*, 114107–114113.
- (24) Nenov, A.; Giussani, A.; Segarra-Martí, J.; Jaiswal, V. K.; Rivalta, I.; Cerullo, G.; Mukamel, S.; Garavelli, M. Modelling the High Energy Electronic State Manifold of Adenine: Calibration for Nonlinear Electronic Spectroscopy. *J. Chem. Phys.* **2015**, *142*, 212443–212459.
- (25) Lorentzon, J.; Fulscher, M. P.; Roos, B. O. Theoretical-Study of the Electronic-Spectra of Uracil and Thymine. *J. Am. Chem. Soc.* **1995**, *117*, 9265–9273.
- (26) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. Towards an Accurate Molecular Orbital Theory for Excited States: Ethene, Butadiene, and Hexatriene. *J. Chem. Phys.* **1993**, *98*, 3151–3162.
- (27) Lorentzon, J.; Malmqvist, P. A.; Fulscher, M.; Roos, B. O. A Caspt2 Study of the Valence and Lowest Rydberg Electronic States of Benzene and Phenol. *Theor. Chim. Acta* **1995**, *91*, 91–108.
- (28) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab initio and Semiempirical Linear Combination of Atomic Orbital Wave-Functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- (29) Rogers, D. M.; Hirst, J. D. Ab Initio Study of Aromatic Side Chains of Amino Acids in Gas Phase and Solution. *J. Phys. Chem. A* **2003**, *107*, 11191–11200.
- (30) Brixner, T.; Stiopkin, I. V.; Fleming, G. R. Tunable Two-Dimensional Femtosecond Spectroscopy. *Opt. Lett.* **2004**, *29*, 884–886.
- (31) Cowan, M. L.; Ogilvie, J. P.; Miller, R. J. D. Two-Dimensional Spectroscopy Using Diffractive Optics Based Phased-Locked Photon Echoes. *Chem. Phys. Lett.* **2004**, *386*, 184–189.
- (32) Tian, P.; Keusters, D.; Suzuki, Y.; Warren, W. S. Femtosecond Phase-Coherent Two-Dimensional Spectroscopy. *Science* **2003**, *300*, 1553–1555.
- (33) Abramavicius, D.; Palmieri, B.; Voronine, D. V.; Sanda, F.; Mukamel, S. Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle Versus Supermolecule Perspectives. *Chem. Rev.* **2009**, *109*, 2350–2408.
- (34) Grumstrup, E. M.; Shim, S.-H.; Montgomery, M. A.; Damrauer, N. H.; Zanni, M. T. Facile Collection of Two-Dimensional Electronic Spectra Using Femtosecond Pulse-Shaping Technology. *Opt. Express* **2007**, *15*, 16681–16689.
- (35) Voronine, D. V.; Abramavicius, D.; Mukamel, S. Manipulating Multidimensional Electronic Spectra of Excitons by Polarization Pulse Shaping. *J. Chem. Phys.* **2007**, *126*, 044508–044515.
- (36) Kimura, K.; Nagakura, S. Vacuum Ultra-Violet Absorption Spectra of Various Mono-Substituted Benzenes. *Mol. Phys.* **1965**, *9*, 117–135.

- (37) Pantos, E.; Philis, J.; Bolovinos, A. Extinction Coefficient of Benzene Vapor in Region 4.6 to 36 Ev. *J. Mol. Spectrosc.* **1978**, *72*, 36–43.
- (38) Hiraya, A.; Shobatake, K. Direct Absorption-Spectra of Jet-Cooled Benzene in 130–260-Nm. *J. Chem. Phys.* **1991**, *94*, 7700–7706.
- (39) Petruska, J. Changes in Electronic Transitions of Aromatic Hydrocarbons on Chemical Substitution 0.2. Application of Perturbation Theory to Substituted-Benzene Spectra. *J. Chem. Phys.* **1961**, *34*, 1120–1136.
- (40) Martinez, S. J.; Alfano, J. C.; Levy, D. H. Rotationally Resolved Fluorescence Excitation-Spectra of Phenol and 4-Ethylphenol in a Supersonic Jet. *J. Mol. Spectrosc.* **1992**, *152*, 80–88.
- (41) Eastman, J. W.; Rehfeld, S. J. Interaction of the Benzene Molecule with Liquid Solvents. Fluorescence Quenching Parallels (0–0) Ultraviolet Absorption Intensity. *J. Phys. Chem.* **1970**, *74*, 1438–1443.
- (42) Jones, D. B.; da Silva, G. B.; Neves, R. F. C.; Duque, H. V.; Chiari, L.; de Oliveira, E. M.; Lopes, M. C. A.; da Costa, R. F.; Varella, M. T. d. N.; Bettega, M. H. F.; Lima, M. A. P.; Brunger, M. J. An Experimental and Theoretical Investigation into the Excited Electronic States of Phenol. *J. Chem. Phys.* **2014**, *141*, 074314–074321.
- (43) Hashimoto, T.; Nakano, H.; Hirao, K. Theoretical Study of the Valence Pi->Pi* Excited States of Polyacenes: Benzene and Naphthalene. *J. Chem. Phys.* **1996**, *104*, 6244–6258.
- (44) Hashimoto, T.; Nakano, H.; Hirao, K. Theoretical Study of Valence and Rydberg Excited States of Benzene Revisited. *J. Mol. Struct.: THEOCHEM* **1998**, *451*, 25–33.
- (45) Hirao, K.; Nakano, H.; Hashimoto, T. Multireference M?ller-Plesset Perturbation Treatment for Valence and Rydberg Excited-States of Benzene. *Chem. Phys. Lett.* **1995**, *235*, 430–435.
- (46) Hirao, K.; Nakano, H.; Nakayama, K.; Dupuis, M. A Complete Active Space Valence Bond (Casvb) Method. *J. Chem. Phys.* **1996**, *105*, 9227–9239.
- (47) Matos, J. M. O.; Roos, B. O.; Malmqvist, P.-Å. A Casscf-Cci Study of the Valence and Lower Excited States of the Benzene Molecule. *J. Chem. Phys.* **1987**, *86*, 1458–1466.
- (48) Bernhardsson, A.; Forsberg, N.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. A Theoretical Study of the 1b2u and 1b1u Vibronic Bands in Benzene. *J. Chem. Phys.* **2000**, *112*, 2798–2809.
- (49) Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. The Multi-State Caspt2Method. *Chem. Phys. Lett.* **1998**, *288*, 299–306.
- (50) Jiang, J.; Mukamel, S. Two-Dimensional Ultraviolet (2duv) Spectroscopic Tools for Identifying Fibrillation Propensity of Protein Residue Sequences. *Angew. Chem., Int. Ed.* **2010**, *49*, 9666–9669.
- (51) Forsberg, N.; Malmqvist, P.-Å. Multiconfiguration Perturbation Theory with Imaginary Level Shift. *Chem. Phys. Lett.* **1997**, *274*, 196–204.
- (52) Roos, B. O.; Andersson, K.; Fülscher, M. P.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M.; Molina, V. Applications of Level Shift Corrected Perturbation Theory in Electronic Spectroscopy. *J. Mol. Struct.: THEOCHEM* **1996**, *388*, 257–276.
- (53) Hamm, P.; Zanni, M. *Concepts and Methods of 2d Infrared Spectroscopy*; Cambridge University Press: Cambridge, U.K., 2011.
- (54) Bayliss, N. S.; Hulme, L. Solvent Effects in the Spectra of Benzene, Toluene, and Chlorobenzene at 2600-a and 2000-A. *Aust. J. Chem.* **1953**, *6*, 257–277.