

# Modeling the high-energy electronic state manifold of adenine: Calibration for nonlinear electronic spectroscopy

Artur Nenov,<sup>1,a)</sup> Angelo Giussani,<sup>1</sup> Javier Segarra-Martí,<sup>1</sup> Vishal K. Jaiswal,<sup>1</sup> Ivan Rivalta,<sup>2</sup> Giulio Cerullo,<sup>3</sup> Shaul Mukamel,<sup>4</sup> and Marco Garavelli<sup>1,2,b)</sup>

<sup>1</sup>Dipartimento di Chimica "G. Ciamician," Università di Bologna, Via Selmi 2, IT-40126 Bologna, Italy

<sup>2</sup>Université de Lyon, CNRS, Institut de Chimie de Lyon, École Normale Supérieure de Lyon, 46 Allée d'Italie, F-69364 Lyon Cedex 07, France

<sup>3</sup>Dipartimento di Fisica, Politecnico di Milano, IFN-CNR, Piazza Leonardo Da Vinci 32, IT-20133 Milano, Italy

<sup>4</sup>Department of Chemistry, University of California, Irvine, California 92697-2025, USA

(Received 6 February 2015; accepted 30 April 2015; published online 18 May 2015)

Pump-probe electronic spectroscopy using femtosecond laser pulses has evolved into a standard tool for tracking ultrafast excited state dynamics. Its two-dimensional (2D) counterpart is becoming an increasingly available and promising technique for resolving many of the limitations of pump-probe caused by spectral congestion. The ability to simulate pump-probe and 2D spectra from *ab initio* computations would allow one to link mechanistic observables like molecular motions and the making/breaking of chemical bonds to experimental observables like excited state lifetimes and quantum yields. From a theoretical standpoint, the characterization of the electronic transitions in the visible (Vis)/ultraviolet (UV), which are excited via the interaction of a molecular system with the incoming pump/probe pulses, translates into the determination of a computationally challenging number of excited states (going over 100) even for small/medium sized systems. A protocol is therefore required to evaluate the fluctuations of spectral properties like transition energies and dipole moments as a function of the computational parameters and to estimate the effect of these fluctuations on the transient spectral appearance. In the present contribution such a protocol is presented within the framework of complete and restricted active space self-consistent field theory and its second-order perturbation theory extensions. The electronic excited states of adenine have been carefully characterized through a previously presented computational recipe [Nenov *et al.*, *Comput. Theor. Chem.* **1040–1041**, 295-303 (2014)]. A wise reduction of the level of theory has then been performed in order to obtain a computationally less demanding approach that is still able to reproduce the characteristic features of the reference data. Foreseeing the potentiality of 2D electronic spectroscopy to track polynucleotide ground and excited state dynamics, and in particular its expected ability to provide conformational dependent fingerprints in dimeric systems, the performances of the selected reduced level of calculations have been tested in the construction of 2D electronic spectra for the *in vacuo* adenine monomer and the unstacked adenine homodimer, thereby exciting the L<sub>b</sub>/L<sub>a</sub> transitions with the pump pulse pair and probing in the Vis to near ultraviolet spectral window. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4921016>]

## I. INTRODUCTION

Electronic excited state dynamics and the underlying ramification of photo-processes that arise upon ultraviolet (UV)/visible (Vis) light irradiation in molecular systems is at the cornerstone of the chemical and physical sciences. A wide variety of systems, ranging from biomolecules to catalysts or light-harvesting complexes, make use of the excess energy obtained by light absorption, converting it to other forms of energy, such as heat by populating vibrational levels or potential energy by charge separation processes, promoting electronic movements from one part of the system to another. These unique features are often employed to reproduce specific effects to construct artificial optical nanotechnological devices

or help elucidate the photoinduced processes in biomolecules and their effects on life.

Photoinduced processes are therefore of paramount importance for a wide range of biological and technological studies. With this in mind, increasing efforts have been devoted over the last two decades to obtaining more accurate experimental and theoretical methods that enhance the intricate characterization of these photo-processes, obscure in nature due to their ultrafast character. Such ultrafast events require extremely precise experimental machinery that pushes the detection limits of the experiments towards the femtosecond and attosecond time scales, ideally resolving the signals registered both temporally and spatially. The advent of nonlinear optical spectroscopic methods, stressed out by the progressively increasing amount of publications related to this topic in recent years, offers diverse methods to accomplish this goal. Pump-probe spectroscopy, a 3rd order nonlinear technique, has been used

<sup>a)</sup>Artur.Nenov@unibo.it

<sup>b)</sup>marco.garavelli@unibo.it and marco.garavelli@ens-lyon.fr

extensively in the past to decipher the photophysics of single nucleotides in gas-phase and solution, so that nowadays we can claim to have gained an in-depth understanding of their excited state dynamics.<sup>1,2</sup> Two-dimensional optical spectroscopy, another 3rd order nonlinear technique, increasingly featured in for the IR regime,<sup>3-5</sup> and in constant development to be incorporated to the UV/Vis range (2DUV/Vis),<sup>6-10</sup> offers a solution to many of the problems emerging when the pump-probe method is applied to diribonucleosides and oligonucleotides. Two-dimensional electronic spectroscopy (2DES) allows disentangling the different signals in their diverse contributions referred to distinct electronic excited states and thus enables probing signals that remain hidden due to the overlap of individual decay pathways in pump-probe spectra. In a 2DES experiment, the system is excited by the sequence of three pulses, with controlled delays  $t_1$  (coherence time) and  $t_2$  (population or waiting time), and the third-order nonlinear signal  $S(t_1, t_2, t_3)$  is emitted, following pulse 3 at time  $t_3$ . By Fourier transforming with respect to  $t_1$  and  $t_3$ , one gets a sequence of 2D maps  $S(\Omega_1, t_2, \Omega_3)$  for different values of the waiting time  $t_2$ .

Nonlinear optical electronic spectroscopy, through a plethora of diverse experimental variations, provides an ultimate technique to track photoinduced events,<sup>11,12</sup> yet yielding a vast amount of signals hard to interpret and understand. It is here where theoretical chemistry comes at hand,<sup>13</sup> as it allows simulating the spectra and provides a route map to recognize the distinct fingerprints and assign them to specific molecular conformations and electronic excited state transitions. From a computational standpoint, the simulation of nonlinear spectra presents its challenges,<sup>14</sup> as it requires a vast amount of data whose cost is intimately associated to the method employed to obtain the energy levels and transition dipole moments (TDMs) characterizing the distinct transitions registered in the spectra. Parameterized model Hamiltonians, like the Frenkel excitonic Hamiltonian and related methods including electrostatic fluctuations,<sup>15-17</sup> make use of parameters obtained from *in vacuo ab initio* computations, which can then be assigned to model arbitrary size aggregates. This approach presents some shortcomings arising from the parameterized data as it is commonly based on single excitations from the lowest-lying states on separate monomers and therefore neglects charge transfer (CT) and doubly excited state contributions where two or more moieties are involved.<sup>18</sup> Furthermore, the neglect of van der Waals interactions and non-classical effects like quantum-mechanical exchange and mechanical coupling becomes important in closely stacked or conjugated aggregates and can no longer be described within the excitonic approximation.<sup>19-22</sup> Nevertheless, parameterized exciton models are commonly applied to study molecular aggregates as they allow for increasing the aggregate size at virtually no cost. In order to obtain more accurate estimates, we have recently introduced a computational protocol based on a sum-over-states approach<sup>23</sup> where *ab initio* calculations for dimeric and multimeric systems within a quantum mechanics/molecular mechanics (QM/MM) scheme<sup>18,24-27</sup> are coupled with nonlinear response theory,<sup>28</sup> so as to give a better account of the inter-monomer and monomer-solvent interaction effects simultaneously. Within this approach, the dimeric or multi-

meric system is treated at an *ab initio* level, thus including the interaction between monomers giving rise to mixed doubly excited and CT states, providing unique fingerprints arising from these interactions that allow disentangling the spectral signals to differentiate specific conformations along folding and unfolding protein processes,<sup>18,27</sup> or to those signals arising due to stacking interactions in a DNA/RNA double-stranded chain.<sup>29</sup>

A robust theoretical protocol that can be systematically improved in an *ab initio* manner is of paramount importance to predict accurate estimates of the fingerprints attained in non-linear spectroscopic experiments. Once defining a solid theoretical approach to compute the vertical excitation energies and transition dipole moments defining the distinct electronic promotions to be simulated within a well-defined degree of precision, further enhancements can be included through external perturbations, such as those arising from QM/MM approaches or due to the small energy fluctuations suffered in ultrafast excited-state dynamics.

We employ a multi-configurational self-consistent field approach due to the multi-configurational nature of the excited states and the need to describe doubly excited states, which is not possible with single determinant based approaches like time-dependent density functional theory (TD-DFT). We utilize the complete active space self-consistent field (CASSCF) and its perturbationally corrected counterpart (CASPT2) in the widely known CASSCF/CASPT2 scheme, because it is the cheapest multi-configurational methodology (cheaper than multi-reference configuration interaction (CI), for example) that gives reliable results.<sup>30-32</sup> The CASSCF/CASPT2 methodology treats electronic correlations on two levels: long range correlation effects (static correlation) are treated variationally (CASSCF) by constructing a multi-configurational wave function through permuting a set of electrons (active electrons) among a set of orbitals (active orbitals), thereby considering all the important electronic configurations, giving rise to excited states. Short range correlations are treated perturbatively (CASPT2), the multi-configurational counterpart of the Møller-Plesset second order perturbation (MP2) method. As the size of the active space must be kept relatively small (<16 orbitals/electrons), different schemes have been proposed in order to increase the accuracy and/or reduce the computational effort. As opposed to CASSCF, the restricted active space SCF (RASSCF) approach poses a limit on the maximal number of simultaneously permuted active electrons, which makes it similar in this sense to truncated CI schemes. Shift parameters (real and imaginary shift, ionization potential electron affinity (IPEA))<sup>33,34</sup> have been introduced in the zeroth order Hamiltonian during the years aiming at resolving intruder state problems, which had led in the past to an avoidable increase of the active space.

Thus, despite not being a black box method and bearing pitfalls when it comes to the construction of the active space or finding the balance between variationally and perturbatively correlation treatment, the proper handling of the CASSCF/CASPT2 protocol has the advantage of allowing a systematic improvement of the computation in an *ab initio* manner. This allows setting up a reference, as it is the aim of the present contribution, which is particularly necessary for

assessing the accuracy of the computation in the absence of experimental data regarding higher lying states.

In the present contribution, a thorough assessment of the computational protocol for the photoinduced dynamics is made on the nucleobase adenine and its homodimer in gas-phase to provide a clear process to improve the results obtained within an *ab initio* approach and to facilitate an example of a relevant biological application. We use 2DES, as the most sophisticated nonlinear optical spectroscopy available to date, to test and visualize the accuracy of the different levels of approximation and their ability to provide qualitatively and qualitatively correct 2DES maps. The method employed, based on CASSCF, its restricted active space counterpart (RASSCF), and second-order perturbation theory extensions (CASPT2/RASPT2), cannot be considered black box and therefore requires certain training and knowledge to be used efficiently.<sup>35</sup> Specific steps are given to improve the precision of the calculation to yield the most accurate electronic spectrum within the CASSCF/CASPT2 protocol to the present date. The fluctuations of spectral properties (transition energies and their associated dipole moments) as a function of the computational parameters (geometry optimization protocol, CAS/RAS schemes, active space size, and CASPT2 energy level shifts) and the effect of these fluctuations on the spectral appearance are demonstrated by simulating 2D electronic spectra in the near ultraviolet (NUV)-pump/Vis- and NUV-probe regimes. 2DES provides the highest possible combination of temporal and spectral resolutions within the framework of 3rd order nonlinear spectroscopy, which qualifies it as the most appropriate marker for the presented calibration protocol. Cost-efficient computational schemes for the monomer are derived to reproduce the essential signals appearing in the reference spectra. In the final calibration step, these cost-saving computational schemes are used to compute the spectra of the unstacked adenine homodimer *in vacuo*, to assess the influence of the vast number of additional excited states appearing in the dimer and considered in the CASSCF/RASSCF state-averaging procedure, and to provide estimates for the charge transfer events recorded in dimeric systems, which are intimately related to the photoinduced events occurring in the genomic material upon UV-light irradiation.<sup>2,31,36–39</sup> Based on the assessment protocol, we are able to provide an estimation about the computational accuracy and reliability of the computed electronic spectra.

## II. COMPUTATIONAL DETAILS

### A. Electronic structure theory calculations

All quantum-mechanical calculations were carried out with the software package Molcas 7.7.<sup>40,41</sup> The reference values for the adenine in gas-phase, used to calibrate different reduced active space schemes, were obtained with a procedure discussed in Sec. II D. The effect of the geometry optimization level on the electronic structure of adenine was studied by comparing the CASSCF results to MP2 theory and DFT making use of the B3LYP functional.<sup>42</sup> Geometry optimizations were done using  $C_s$ -symmetry. Different RASSCF schemes were assessed regarding their ability to reproduce the tran-

sition energies and dipole moments of the reference calculation: SA-25-RAS(4,6|0,0|4,4), SA-25-RAS(4,6|0,0|4,3), SA-20-RAS(4,4|0,0|4,3), and a SA-15-CAS(4,4), with RAS1 described by the maximum number of holes and orbitals, RAS2 by the number of electrons and orbitals, and RAS3 by the maximum number of excitations and orbitals. The exclusion of the  $n$  lone pair orbitals and their  $n\pi^*$  transitions is based on the relatively small dependence these states have on the dynamic correlation here benchmarked at the CASPT2/RASPT2 levels of theory. Their mediating role in the photoinduced events taking place in adenine is highly dependent on the environment surrounding the system,<sup>43</sup> as the  $n\pi^*$  state lays relatively low in energy and is therefore accessible *in vacuo* from the spectroscopic La state while being blue-shifted in polar solvents thus preventing its non-adiabatic population.<sup>44</sup> Additionally,  $n\pi^*$  states are usually dark in nature due to their exceedingly small oscillator strengths associated to their ground to excited state transitions and are therefore expected to yield negligible signals that are likely to be covered by the brighter  $\pi\pi^*$  contributions in the spectra occurring at similar energy ranges.

The dynamic correlation was recovered with the multi-configurational second order perturbation theory approach in its single-state variant (i.e., SS-CASPT2/RASPT2).<sup>45,46</sup> All virtual orbitals were correlated in the perturbation procedure except for the 24 removed Rydberg orbitals with  $A''$ -symmetry (Sec. II D). An IPEA shift of 0.0 a.u. was used,<sup>34</sup> while the real and imaginary shift parameters were scanned in the range 0.2–0.7 a.u.<sup>33</sup> Transition dipole moments were calculated at SA-CASSCF/RASSCF level.<sup>47,48</sup> The Cholesky decomposition was used to speed up the calculation of two-electron integrals.<sup>49–51</sup>

The usage of a non-corrected IPEA Hamiltonian is based on previous experience reported in the literature for DNA/RNA systems,<sup>52–54</sup> where red-shifted estimates are obtained when using this correction. Zeroth order IPEA-shifted Hamiltonians are advised when dealing with open-shell cationic and anionic states, as they provide estimates closer to their experimental counterparts in terms of ionization potential,<sup>55</sup> electron affinities,<sup>56</sup> and dissociative electron attachments.<sup>57</sup> However, this behaviour is not general and appears to be rather system dependent,<sup>58</sup> thus justifying the use of non-IPEA Hamiltonians in the present study.

All computations for big active spaces, mainly those describing dimeric species and employing RASPT2 schemes, make use of large imaginary level shift values in order to correct the vertical excitation energies with respect to the reference calculation, which is to our knowledge the only reference available for the high-energy excited state manifold. It must be noted that the reference calculation uses an extended basis set and a large active space that is impractical for any actual application and the use of smaller active spaces introduces an error in the excitation energies that needs to be corrected for obtaining accurate spectra. To this end, the imaginary and standard level shift, the standard being reported in the supplementary material,<sup>61</sup> was employed in a semiempirical manner in order to obtain the closest estimates to the reference calculation. Large shift values have been previously reported in the literature for a range of multireference perturbation

theory techniques,<sup>59</sup> pointing towards an improvement in the rotational and dissociation constants when applied to diatomic molecules, as they do remove the effect of intruder states while getting closer estimates for the different magnitudes measured to the experiment, even if this effect is possibly due to error compensation.

The best performing reduced-cost computational schemes of the calibration procedure were applied to an unstacked adenine homodimer *in vacuo*, where the symmetry constraints were released. Thereby up to 150 excited states were included in the state averaging procedure.

## B. Nomenclature

Adenine has ten valence  $\pi$ -orbitals, six of which are occupied. The two highest occupied orbitals (HOMO - 1 and HOMO) as well as the two lowest unoccupied orbitals (i.e., LUMO and LUMO + 1) can be viewed as benzene-like orbitals (Fig. 1). Following Platt's notation,<sup>60</sup> the states associated with electronic transitions among these orbitals were named  $L_b$ ,  $L_a$ ,  $B_b$ , and  $B_a$ . Classification of higher lying excited states of polycyclic aromatic compounds beyond the  $L_{a/b}$  and  $B_{a/b}$  states is somewhat ambiguous due to the multi-configurational nature of the excited states. In the following, we adopt the nomenclature introduced in Ref. 26 (see Fig. 1). The categorization is done assuming that each bonding orbital

(excluding lone pairs) has an anti-bonding counterpart and that transitions from the bonding and into the anti-bonding orbitals are energetically degenerate and mix to give rise to the observed electronic transitions. Furthermore, relying on the electronic structure of benzene, it was assumed that HOMO - 1 and HOMO are degenerate, as are LUMO and LUMO + 1. Following these assumptions, the  $L_{a/b}$  and  $B_{a/b}$  states are united in the L band, while labeling of the remaining transitions is done in alphabetic order (e.g., A - 1, B + 2, etc.). The “ $\pm$ ” superscripts indicate that two sub-bands, each one consisting of two levels denoted by a subscript, emerge through different linear combinations of the same transitions.

## C. Parameters for the 2D spectroscopy and laser pulse setup

Using the computed SS-RASPT2 energies and the SA-RASSCF TDMs, quasi-absorptive 2DES maps were computed for each snapshot by the sum-over-states approach in the dipole approximation with Spectron 2.7.<sup>28</sup> Additional information regarding the specific theoretical framework employed to obtain the spectra can be found in the supplementary material.<sup>61</sup>

The 2DES experiments here simulated can be schematically described as shown in Figure 2. The molecular system under study is excited by three consecutive pulses, with two controllable relative delays  $t_1$  and  $t_2$ . This pulse sequence generates a nonlinear polarization in the medium that emits a field following pulse 3 with a delay time  $t_3$ . Pulse 4, known as local oscillator (LO), is then employed to fully measure this field registering it both in amplitude and phase. The results attained are then Fourier transformed with respect to  $t_1$  and  $t_3$  for a fixed value of *waiting time*  $t_2$ , resolving the 2D spectrum as a function of the *excitation frequency*  $\omega_1$  and *detection frequency*  $\omega_3$ . When the excitation pulses are sufficiently short, 2DES measures directly the third-order non-linear optical response of the system, providing the ultimate time-resolved non-linear optical experiment that surpasses standard pump-probe setups. Its implementation has been hampered due to the need to phase-lock between pulses 1 and 2 in order to be able to Fourier transform and between pulses 3 and 4 for spectral interferometry, which has prevented its wide use in applications. 2DES can be implemented in two configurations, the partially collinear pump-probe geometry being the one simulated in the present study. This technique uses two phase-locked collinear pump pulses and a non-collinear probe pulse, which is dispersed on a spectrometer (see Fig. 2). The probe pulse has the dual purpose of generating the non-linear polarization and heterodyning it.

The static picture is adopted for describing the dynamics occurring during the coherence time  $t_1$ , and the waiting time  $t_2$  between the pump pulse pair and the probe pulse was set to zero, thus, assuming that the relaxation dynamics in the system is slower than the time scale of the experiment. The static approximation allows generating the 2D spectra based solely on computations at the Frank-Condon (FC) point. This approximation, although being rather crude, is fully justified for our aim to study the effect of various computational parameters on the shape of the 2D maps. A constant line broadening of  $200\text{ cm}^{-1}$  was used throughout. All calculated signals use

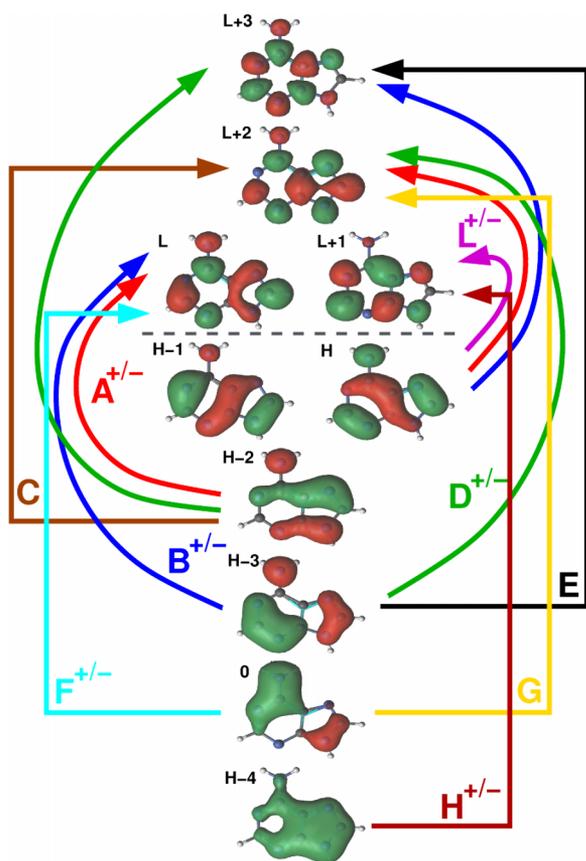


FIG. 1. List of the valence  $\pi$ -orbitals of adenine together with different single electron transitions which determine the absorption spectrum. Nearly degenerate transitions that are likely to mix are combined in groups denoted by a color code.

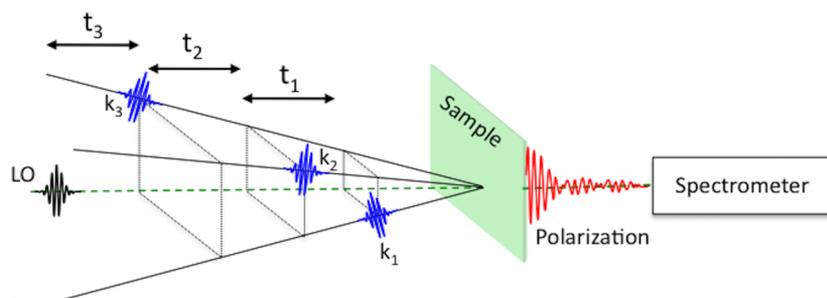


FIG. 2. Pulse sequence used in the simulated 2DES experiment based on a three-pulse photon echo setup.

the nonchiral *xxxx* pulse polarization configurations and are plotted on a logarithmic scale. Ground-state bleaching (GSB) and stimulated emission (SE) contributions appear as negative (blue) peaks, and excited-state absorptions (ESAs) appear as positive (red) peaks in the 2D spectra.

The  $L_b$  in adenine has a weaker oscillator strength compared to the  $L_a$  transition (Table I). Hence, its spectral resolution requires setting a narrowband pulse in resonance with the  $L_b$  transition. In the present study, pulses with a  $1000\text{ cm}^{-1}$  bandwidth were utilized to enhance the  $L_b$  trace, while still resolving the intense  $L_a$  traces. The sensitivity of the 2DES on the position of the pump pulse pair is also demonstrated by comparing pumping in the  $L_b$  region with pumping in the  $L_a$  region. Probing is performed in two spectral windows: NUV and Vis. Probing in the NUV resolves correlated transitions in both monomers and coupled aggregates. For adenine, we probe the spectral region between  $33\,000\text{ cm}^{-1}$  and  $44\,000\text{ cm}^{-1}$  to resolve the correlation between the  $L_b$  and  $L_a$  states of the monomer. Therefore, the probe pulse pair is centered in the same region as the pump pulse pair (i.e., around  $\sim 39\,000\text{--}40\,000\text{ cm}^{-1}$ , one-color regime, Fig. 3). Probing in the Vis covers the spectral region below the ionization potential and permits the collection of background-free signals. This is particularly interesting for the sake of characterizing charge transfer states in coupled chromophore aggregates, since interchromophore interactions are relatively weak and require regions free of background noise and overlapping intensive local absorptions. As we do not know *a priori* where the CT transitions are likely to appear, we probe a broad spectral window between  $12\,000\text{ cm}^{-1}$  and  $30\,000\text{ cm}^{-1}$  covering the near-IR to Vis region (Fig. 3). The probe pulse is centered at  $20\,000\text{ cm}^{-1}$  (i.e., two-color regime).

#### D. Reference methodology

Following the procedure described in the supplementary material,<sup>61</sup> 25 valence states are described at full valence CAS(12,10) level. Next, a gradual increase of the active space is performed until convergence of the CASPT2 excitation energies. Thereby, 4, 8, and 12 virtual extra-valence orbitals with the lowest diagonal elements in the effective state averaged Fock matrix are added. In order to reduce the computational cost, the increase of the active space was done in a RASSCF manner by allowing up to two electrons into the extra-valence orbitals in the RAS3 subspace. An increase of the active space promotes a better description of the dynamic correlation already at CASSCF level, effectively lowering the absolute energies of all states. Accordingly, the energy gap

between the reference states and the configurations used in the CASPT2 routine becomes larger, assuring that the contribution of each configuration to the dynamic correlation is small, as required for perturbation methods. The nature of the extra-valence orbitals is not crucial for the outcome of the procedure. Nevertheless, the lowest diagonal elements in the effective state averaged Fock matrix give the most stable results (in terms of active space orbital rotations) and it is these orbitals that are expected to contribute the most to the dynamic correlation, assuring that convergence will be achieved with only a small number of extra-valence orbitals. Upon convergence, the transition energies (and all other state properties as well) should become independent on whether further extra-valence orbitals are treated variationally or perturbationally. As long as this is not the case, the active space requires enlargement. The progressive active space increase is the optimal way to determine the minimal size required for obtaining the most stable CASSCF/CASPT2 results.

### III. REFERENCE SPECTRUM OF ADENINE IN GAS-PHASE

Table I summarizes the electronic spectrum of the ground state equilibrium geometry of adenine optimized at four computational levels, CAS(8,8)/6-31G\*, CAS(12,10)/ANO-L[432,21], MP2/ANO-L[432,21], and B3LYP/6-31G\*. In the case of the generally contracted basis set ANO-L the numbers in the parentheses denote the contraction schemes used for non-hydrogen (i.e., 4s3p2d) and hydrogen (i.e., 2s1p) atoms. Figures 4 and 5 show the NUV-pump Vis-probe (in the range between  $12\,000\text{ cm}^{-1}$  and  $30\,000\text{ cm}^{-1}$ ) and NUV-pump NUV-probe (in the range between  $33\,000\text{ cm}^{-1}$  and  $44\,000\text{ cm}^{-1}$ ) 2D spectra for the CAS(8,8)/6-31G\* ground state minimum obtained through the gradual active space increase outlined in Sec. II D.

Before analyzing the spectra in detail, we draw several conclusions by visually inspecting Figs. 4 and 5. The gradual active space increase demonstrates the sensitivity of electronic spectra to dynamic correlation. First, and maybe most alerting, is the fact that full valence active space calculations *do not* provide converged energies and, hence, should not be used as a reference. Significant changes are observed when extra-valence orbitals are included in the active space. The spectra stabilize with four (NUV-pump Vis-probe, Figs. 4(a) and 4(b)) to eight (NUV-pump NUV-probe, Figs. 5(a) and 5(b)) additional orbitals in the active space. Adding more orbitals fine-tunes the positions of the peaks.

TABLE I. *In vacuo* vertical excitation energies (in eV, columns 2-5), transition dipole moments (in a.u.) associated to each transition arising from the GS (column 7), the first ( $L_b$ , column 8), and the second ( $L_a$ , column 9) excited states, leading wave function configurations (column 10), and corresponding coefficients (column 11) calculated at the SA-25-RAS(0,0|12,10|2,12)/SS-RASPT2 level of theory. Experimental values, taken from Ref. 62, are also provided. Labeling is given according to Fig. 3. Electronic levels are reported for four different ground-state optimization protocols: CAS(8,8)/6-31G\*, CAS(12,10)/ANO-L[432,21], MP2/ANO-L[432,21], and B3LYP/6-31G\*. Transition dipole moments are reported for the geometries optimized at CAS(8,8)/6-31G\* and MP2/ANO-L[432,21] (in brackets). Wave functions and coefficients are reported for the geometry optimized at the CAS(8,8)/6-31G\* level.

Label	Energy					TDM			Main configuration	Coefficient
	CAS88	CAS1210	MP2	DFT	Expt.	GS $\rightarrow$ $S_N$	$L_b \rightarrow S_N$	$L_a \rightarrow S_N$		
2A'(L <sub>1</sub> <sup>+</sup> )	4.94	4.90	4.74	4.75	4.6(I)	0.08(0.09)	...	...	H $\rightarrow$ L + 1 H - 1 $\rightarrow$ L	0.63 -0.54
3A'(L <sub>2</sub> <sup>+</sup> )	5.18	5.13	4.98	4.96	4.8(II)	1.35(1.41)	0.50(0.46)	...	H $\rightarrow$ L H - 1 $\rightarrow$ L + 1	0.79 0.30
4A'(L <sub>1</sub> <sup>-</sup> + A <sub>2</sub> <sup>+</sup> )	6.24	6.24	6.18	6.11	5.9(III)	1.00(1.51)	0.50(0.42)	0.70(0.83)	H - 1 $\rightarrow$ L H $\rightarrow$ L + 1 H - 2 $\rightarrow$ L	0.44 0.35 0.35
5A'(A <sub>1</sub> <sup>+</sup> )	6.37	6.37	6.15	6.19	6.3(IV)	0.15(0.13)	0.20(0.25)	1.20(1.09)	H $\rightarrow$ L + 2 H - 2 $\rightarrow$ L	0.53 0.36
6A'(A <sub>2</sub> <sup>+</sup> - L <sub>1</sub> <sup>-</sup> )	6.38	6.37	6.17	6.20		1.50(0.84)	0.33(0.47)	0.90(0.71)	H - 2 $\rightarrow$ L H $\rightarrow$ L + 2 H - 1 $\rightarrow$ L	0.48 -0.34 -0.32
7A'(L <sub>2</sub> <sup>-</sup> )	6.86	6.84	6.68	6.68		1.60(1.67)	0.44(0.50)	0.28(0.26)	H - 1 $\rightarrow$ L + 1 H $\rightarrow$ L	0.65 -0.27
8A'(A <sub>1</sub> <sup>-</sup> )	7.00	7.03	6.91	6.90	6.8(V)	1.00(0.50)	0.55(0.40)	0.56(0.55)	H - 1 $\rightarrow$ L + 2 H - 2 $\rightarrow$ L + 1	0.59 0.33
9A'(A <sub>2</sub> <sup>-</sup> )	7.39	7.40	7.24	7.29		1.40(1.25)	0.74(0.70)	0.07(0.13)	H - 2 $\rightarrow$ L + 1 H - 1 $\rightarrow$ L + 2	0.47 -0.28
10A'(B <sub>1</sub> <sup>+</sup> )	7.94	7.90	7.64	7.52	7.7(VI)	0.70(0.88)	0.18(0.24)	0.40(0.51)	H - 3 $\rightarrow$ L	0.46
11A'(C)	8.19	8.25	8.15	8.17		1.50(1.16)	0.33(0.27)	0.53(0.90)	H - 2 $\rightarrow$ L + 2	0.57
12A'(B <sub>2</sub> <sup>+</sup> )	8.53	8.53	8.35	8.32		0.80(0.82)	0.72(0.67)	0.13(0.13)	H - 3 $\rightarrow$ L + 1	0.53
13A'	8.83	8.76	8.25	8.33		0.30(0.69)	0.46(0.49)	1.15(0.99)	H $\Rightarrow$ L	0.47
14A'(F <sup>+</sup> + B <sub>1</sub> <sup>-</sup> )	9.14	9.12	8.92	8.91		0.50(0.50)	0.48(0.55)	0.26(0.17)	0 $\rightarrow$ L 0 $\rightarrow$ L + 1 H $\rightarrow$ L + 3	0.34 -0.34 0.27
15A'(F <sup>+</sup> )	9.19	9.19	9.09	8.79		0.45(0.76)	0.16(0.48)	0.31(0.26)	0 $\rightarrow$ L	0.48
16A'(F <sup>-</sup> )	9.38	9.29	8.88	9.05		0.70(0.55)	0.49(0.34)	0.00(0.15)	0 $\rightarrow$ L + 1	0.52
17A'	9.54	9.46	9.19	9.15		0.40(0.20)	0.82(0.83)	0.85(0.91)	H $\Rightarrow$ L, L + 1	0.58
18A'	10.04	10.06	9.77	9.82		0.45(0.48)	0.29(0.29)	0.32(0.25)	H - 3 $\rightarrow$ L + 2 H - 1, H $\Rightarrow$ L + 1, L + 2 0 $\rightarrow$ L + 2 H - 2 $\rightarrow$ L + 2	0.29 -0.25 0.23 0.20
19A'	10.22	10.25	10.00	10.00		0.45(0.49)	0.19(0.17)	1.10(0.95)	0 $\rightarrow$ L + 2 H - 1, H $\Rightarrow$ L + 1, L + 2 H $\Rightarrow$ L	0.28 -0.23 -0.22
20A'(D <sup>+</sup> )	10.31	10.28	9.96	9.99		0.20(0.26)	0.22(0.19)	0.35(0.37)	H $\Rightarrow$ L + 1, L + 2 H - 3 $\rightarrow$ L + 2	0.38 0.33
21A'(H <sup>+</sup> )	10.33	10.36	10.12	10.07		0.30(0.28)	0.19(0.20)	0.56(0.68)	H $\Rightarrow$ L + 1 H - 4 $\rightarrow$ L	0.35 -0.35

TABLE I. (Continued.)

Label	Energy					TDM			Main configuration	Coefficient
	CAS88	CAS1210	MP2	DFT	Expt.	GS $\rightarrow$ S <sub>N</sub>	L <sub>b</sub> $\rightarrow$ S <sub>N</sub>	L <sub>a</sub> $\rightarrow$ S <sub>N</sub>		
22A'(G)	10.51	10.49	10.14	10.16		0.15(0.16)	0.16(0.26)	0.39(0.89)	0 $\rightarrow$ L+2	0.45
23A'(G)	10.56	10.53	10.27	10.26		0.45(0.44)	0.52(0.57)	0.39(0.38)	0 $\rightarrow$ L+2	0.26
24A'(H <sup>+</sup> )	10.62	10.52	10.29	10.24		0.30(0.33)	0.24(0.09)	0.63(0.16)	H-4 $\rightarrow$ L H-4 $\rightarrow$ L+2	0.35 -0.29

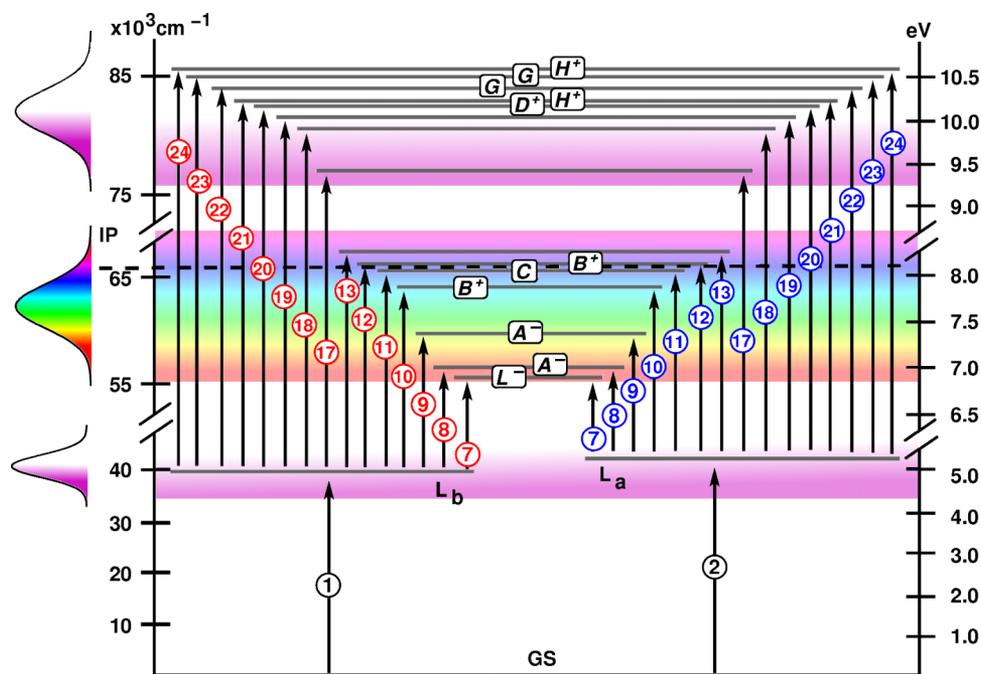


FIG. 3. Level scheme of adenine monomer including the higher excited states that can be probed out of the L<sub>b</sub> and L<sub>a</sub> states by using broadband pulses centered in the Vis and NUV spectral windows. Signals associated with transitions from the L<sub>b</sub> state are shown in red, whereas the associated transitions from L<sub>a</sub> are shown in blue. The numbering corresponds to the SA-25-RAS(0,0|12,10|2,12)/SS-RASPT2 state ordering *in vacuo* (Table I, 2nd column).

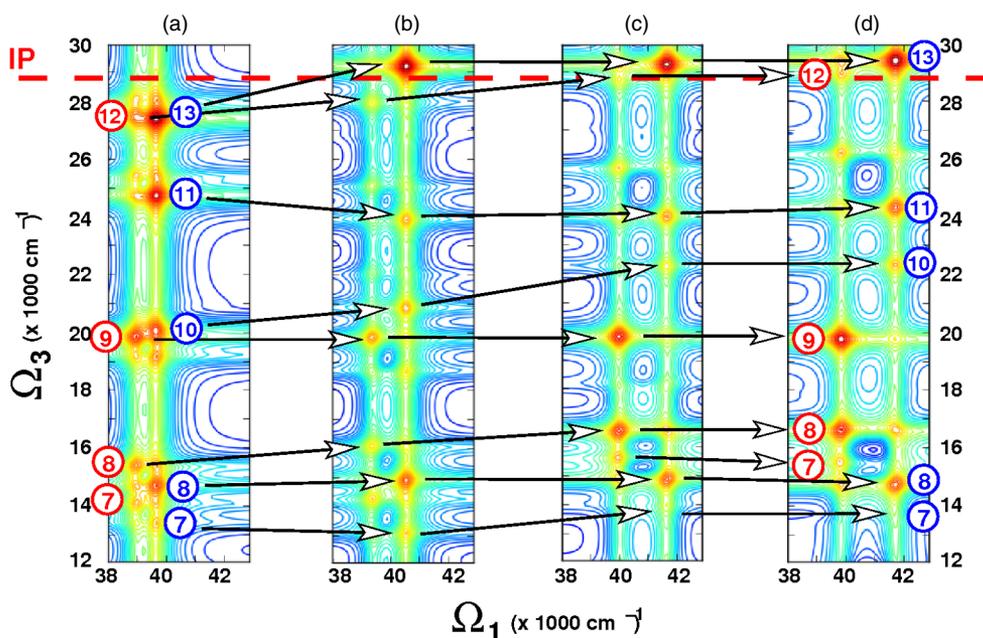


FIG. 4. 2D NUV-pump Vis-probe spectra of adenine monomer in gas-phase for different active space sizes: RAS(0,0|12,10|0,0), RAS(0,0|12,10|2,4), RAS(0,0|12,10|2,8), and RAS(0,0|12,10|2,12). Labeling according to Fig. 3.

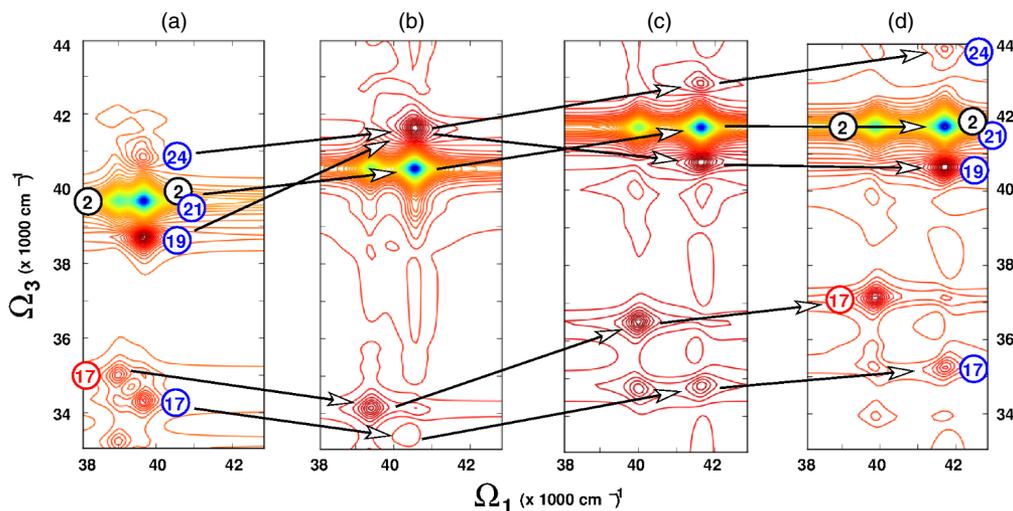


FIG. 5. 2D NUV-pump NUV-probe spectra of adenine monomer in gas-phase for different active space sizes: RAS(0,0|12,10|0,0), RAS(0,0|12,10|2,4), RAS(0,0|12,10|2,8), and RAS(0,0|12,10|2,12). Labeling according to Fig. 3.

Second, the  $L_b/L_a$  gap shows a pronounced dependence on the active space. This may be understood in the framework of valence-bond theory,<sup>63</sup> which predicts a higher covalent character for the  $L_b$  state while the  $L_a$  state shows a more pronounced ionic character. As the ground state is itself of covalent nature, transitions to covalent excited states are fairly well described already at CASSCF level (the CASPT2 stabilization energy for the  $L_b$  state is  $\sim 0.5$  eV). Ionic states, on the contrary, exhibit different  $\pi \rightarrow \pi$  and  $\sigma \rightarrow \pi$  polarization effects and the quality of their description depends strongly on the treatment of the dynamic correlation (the CASPT2 stabilization energy for the  $L_a$  state is above 1.0 eV). Our calculations reveal that even a small modulation to the dynamic correlation introduced through extra-valence orbitals in the active space can alter the gap significantly. It is to be expected that truncation below the full valence active space will have an even more dramatic effect. Indeed, this is observed for the truncated RASSCF/RASPT2 schemes presented in Sec. IV. In the light of these findings, it should not be surprising that methods like CASSCF/CASPT2, MRCI, DFT, etc., which all describe the correlation on different theoretical foundations, give controversial results regarding the energetic order of the  $L_b$  and  $L_a$  states.<sup>64</sup> Our calculations are in favor of  $\sim 0.2$  eV gap between  $L_b$  and  $L_a$ , placing the  $L_a$  state higher in energy in agreement with results from previous studies using wave function-based methods.<sup>52,64–68</sup>

Experiments in aqueous solution assign a weak band at 4.6 eV to the  $L_b$  transition (denoted as band I in Table I), while a more intense peak around 4.8 eV is believed to arise from a transition to the  $L_a$  state (band II in Table I). Vapor experiments report only a peak at 4.9 eV, which we tentatively assign to the more intense  $L_a$  transition. Our results deviate by  $\sim 0.30$  eV and we do not expect the bathochromic effect of the solution to fully correct for this discrepancy. A possible explanation for the discrepancy is given later in this section.

As a third observation, we confirm that CASPT2, being a non-variational method, generally tends to overestimate the stabilization of the excited states. Describing a larger part of the dynamic correlation variationally (i.e., at CASSCF level)

leads to a blue shift of all states. As a result, a blue-shift of the spectral traces (along both  $\Omega_1$  and  $\Omega_3$ ) is observed, whereas the blue-shift along  $\Omega_3$  is more pronounced in the NUV-probe spectra (Fig. 5).

Next, we discuss the electronic spectra of adenine in detail considering the most accurate RAS(0,0|12,10|2,12) results (Figs. 4(d) and 5(d)). The VIS-probe spectrum (Fig. 4(d)) visualizes ESA (positive, red) from  $L_b/L_a$  to singly excited states below the ionization potential (8.3–8.5 eV corresponds to 26 000/28 000  $\text{cm}^{-1}$  from the  $L_b/L_a$ ), the only exception being peak 13, just above the ionization limit, which arises from a transition to the doubly excited state HOMO, HOMO  $\rightarrow$  LUMO, LUMO. Peak 13 is also the most intense signal in the Vis-spectrum of the  $L_a$  trace. Using IR-terminology it can be regarded as the overtone of the  $L_a$  transition, but with a large anharmonicity of 12 000  $\text{cm}^{-1}$ . The  $L_b$  trace is characterized by several intense peaks (peaks 7, 8, and 9) in the lower part of the spectrum, while the  $L_a$  trace shows more characteristic peaks around the ionization limit (peaks 10, 11, and 13). These peaks can be used as state recognition makers as it will be demonstrated later in Sec. IV. The existence of absorption-free windows between 20 000  $\text{cm}^{-1}$  and 26 000  $\text{cm}^{-1}$  along the  $L_b$  trace and between 16 000  $\text{cm}^{-1}$  and 22 000  $\text{cm}^{-1}$  along the  $L_a$  trace is promising for tracking CT states in aggregates which are expected to manifest as weak absorptions below the ionization limit.

The quality of the 2D spectra is now analyzed by comparison with experimental and theoretical data. The adenine excited states below the ionization limit have been observed experimentally by linear absorption, as well as by linear and circular dichroism spectroscopy. In a series of seminal papers, Roos and co-workers reported theoretical results on the energy spectra of purines and pyrimidines up to the ionization potential based on CASSCF/CASPT2 calculations.<sup>62,69,70</sup> An overview of experimental results is given in Refs. 1, 2, 38, and 71. Apart from the already discussed  $L_b$  (I) and  $L_a$  (II) bands, different experiments report four bands: 5.9 eV (III), 6.3 eV (IV), 6.8 eV (V), and 7.7 eV (VI), with bands III and V being the most intense ones ( $f_{\text{III}} = 0.25$  and  $f_{\text{V}} = 0.30$ ).

Our calculations show that these two bands are associated with transitions from the GS to the  $L^{-1}_1(B_b)$  and  $L^{-2}_2(B_a)$  states, which constitute the most intense transitions in the UV spectra of indole, benzene, and phenol as well.<sup>27,42,46,57,72</sup> Roos and co-workers also arrived at similar conclusions about the nature of the most intense transitions. As with the  $L_{b/a}$  states, our calculations tend to overestimate the transition energies for bands III–VI by up to 0.30 eV (Table I). We correctly predict band V (states 7A' and 8A') to be the most intense but fail to predict the intensity of band III (state 4A'). We believe that the reason is in the mixing of the  $L^{-1}_1$  state with the  $A^+_2$  state, which transfers oscillator strength to band IV (state 5A' and 6A'). As we will demonstrate now, these discrepancies are not caused by the CASSCF/CASPT2 computational procedure but are rather inherited by the GS optimization level. Table I also reports the excited energies of adenine obtained for a geometry optimized at MP2/ANO-L[432/21] level. It is evident that MP2 optimization tends to elongate the double bonds by  $\sim 0.03$  Å, similar to the DFT results. This leads to a red-shift of the transition energies by up to 0.20 eV so that we observe an agreement with the experimental values within 0.10 eV. The RAS(0,0|12,10|2,12)/RASPT2 calculation on top of the MP2 and DFT optimized geometries even correctly predicts the intensity of bands III and IV. We note, however, that bands III and IV are energetically not clearly separated at these computational levels. Figs. 6(b)–6(d) compare the NUV-pump Vis-probe spectra generated for the different GS geometries. Due to the overall red-shift of all transitions for the MP2 and DFT optimized geometries, the relative peak positions in the spectra remain nearly unchanged, with peaks 11 and 13 being the only outliers from the general trend. It should be pointed out that in the MP2 and DFT spectra, the  $L_b$  and  $L_a$  trace red-shift along  $\Omega_1$  by  $\sim 2000$   $\text{cm}^{-1}$ . States 11 and 13 exhibit a pronounced sensitivity to the computational level, which affects the level of state mixing and consequently the energies and the TDMS. Depending on the computational level, states 11 and 13 can be uncoupled, with peak 13 being very intense or can mix strongly, which causes peak 11 to blue-shift and gain in intensity at the expense of the red-shifted peak 13.

The NUV-probe spectrum (Fig. 5(d)) is dominated by a diagonal ( $\Omega_1 = \Omega_3 \sim 42\,000$   $\text{cm}^{-1}$ ) negative (blue) signal along the  $L_a$  trace and by an off-diagonal ( $\Omega_1 \sim 40\,000$   $\text{cm}^{-1}$ ;  $\Omega_1 \sim 42\,000$   $\text{cm}^{-1}$ ) negative signal along the  $L_b$  trace. The diagonal signal consists of two contributions,  $L_a$  GSB and SE from the  $L_a$  FC point. The off-diagonal signals appear for coupled transitions when pump and probe pulse excite different transitions. For non-correlated oscillators, these signals vanish due to exact cancellation of quantum pathways given by different Feynman diagrams. For correlated oscillators as in the case of transitions on the same chromophore, the off-diagonal bleach represents a characteristic fingerprint of the NUV 2D spectrum. The signals reflect intrinsic properties of the adenine chromophore and are present in the monomer as well as in aggregates. The corresponding GSB of the  $L_b$  state, which should appear at  $\Omega_1 = \Omega_3 \sim 40\,000$   $\text{cm}^{-1}$ , as well as the opposite off-diagonal bleach below the diagonal is not visible in the spectrum due to the lower oscillator strength of the  $L_b$  transition but could be enhanced by probing with a narrowband

pulse centered around  $40\,000$   $\text{cm}^{-1}$  (or  $38\,000$   $\text{cm}^{-1}$  if the MP2 or DFT optimized geometries are taken as reference). This has been recently demonstrated by 2DUV experiments conducted in the groups of Moran and Weinacht.<sup>15–17,58</sup>

The ESA in the NUV-window is characterized through transitions into states whose wave functions are a mixture of doubly excited configurations and single excitations out of the lone pairs “H-4” and “O” into the lowest unoccupied orbitals (Table I). Notably, the lack of leading configurations makes the assignment of the states ambiguous. A characteristic signature in the low-energy window appears through peak 17 and the state associated with this transition is a double excitation HOMO, HOMO  $\rightarrow$  LUMO, LUMO + 1. The peak constitutes the most intense signal along the  $L_b$  trace, which can be rationalized when viewing the doubly excited state as an  $L_a$  transition on top of the  $L_b$  transition. The high-energy window of the spectrum is characterized mainly through several transitions along the  $L_a$  trace (peaks 19, 21, and 24). Comparison between the CAS(8,8), MP2, and DFT optimization methods (Figs. 7(b)–7(d)) reveals that the region around the GSB is particularly sensitive to the optimization methods. While the GSB peaks red-shift, ESA peaks 19 blue-shifts, which causes the peaks to swap order in the spectrum and results in qualitatively different predictions. To the best of our knowledge, no experimental data are available for the higher excited states in order to give an estimation of the accuracy of the reported computations in the NUV.

For the sake of completeness, we report also the geometrical parameters (see the supplementary material),<sup>61</sup> energies (Table I), and spectra obtained from a GS geometry optimized at CAS(12,10)/ANO-L[432,21] level (Figs. 6(a) and 7(a)). The small geometrical differences result in spectra which are nearly identical to the spectra obtained at CAS(8,8)/6-31G\* level (Figs. 6(a), 6(b), 7(a), and 7(b)), thus confirming that CAS(8,8)/6-31G\* is sufficient for GS optimizations within the limits of the CASSCF framework.

Based on the above discussion, we conclude that the GS optimization methods introduce an uncertainty in the electronic spectrum of adenine, which generally leads to an uncertainty in the peak positions along  $\Omega_1$  and  $\Omega_3$  of around  $2000$   $\text{cm}^{-1}$ . The high density of excited states may cause state mixing and consequently pronounced peak shifts for even small geometry changes in the GS. The effects become more pronounced the higher the probing window. The Vis-window particularly below the ionization limit proves *stable with respect to the GS optimization level* and exhibits characteristic fingerprints that allow disentangling  $L_b$  and  $L_a$  traces, as well as broad signal-free windows for CT tracking.

#### IV. CALIBRATION OF ADENINE MONOMER AND HOMODIMER CALCULATIONS

The RAS(0,0|12,10|2,12)/ANO-L[432,21] results (referred to as reference from now on) presented in Sec. III probably constitute the most accurate RASSCF/RASPT2 data nowadays obtainable for the valence excited states of adenine below 10 eV at its equilibrium structure. They have been computed using what can be regarded as one of the most suitable strategies for the description of the photophysics and

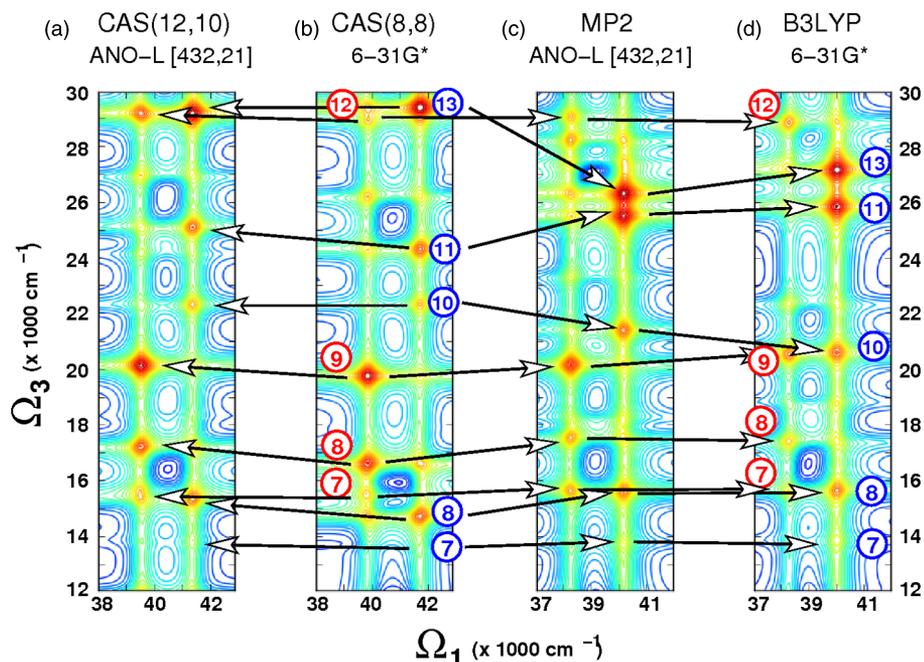


FIG. 6. Dependence of the 2D NUV-pump Vis-probe spectrum on the GS optimization method: CAS(12,10)/ANO-L[432/21] (a) vs. CAS(8,8)/6-31G\* (b) vs. MP2/ANO-L[432/21] (c) vs. B3LYP/6-31G\* (d). Labeling of the electronic transitions according to Fig. 3.

photochemistry of medium size systems (i.e., the RASSCF and RASPT2 methods),<sup>59,60</sup> testing the convergence of the outcomes wisely while increasing the selected active space and using diffused basis functions, thereby taking into account the related Rydberg-valence orbital and state mixing problems (see Sec. II D).

For the description of the NUV-pump/Vis-probe and NUV-pump/NUV-probe accessible excited states in the adenine dimer, the ideal solution would be to use the high level of theory characterized above. Since such a possibility would lead to an unaffordable computational effort (it would require to compute RASSCF and RASPT2 calculations with more

than  $10 \times 10^6$  configuration state functions), a reduced computational scheme demanding a reasonable computational cost but still able to provide the qualitative main features of the adenine excited states should be used. In order to determine the computational scheme that more closely fulfills such conditions, different excited state calculations on adenine reducing the employed level of theory have been performed, and the corresponding outcomes have been compared with respect to the reference computation.

We have tested the ability of the reduced computations to reproduce the correct energy differences and TDMs (which are the required quantities in order to build the spectra within the

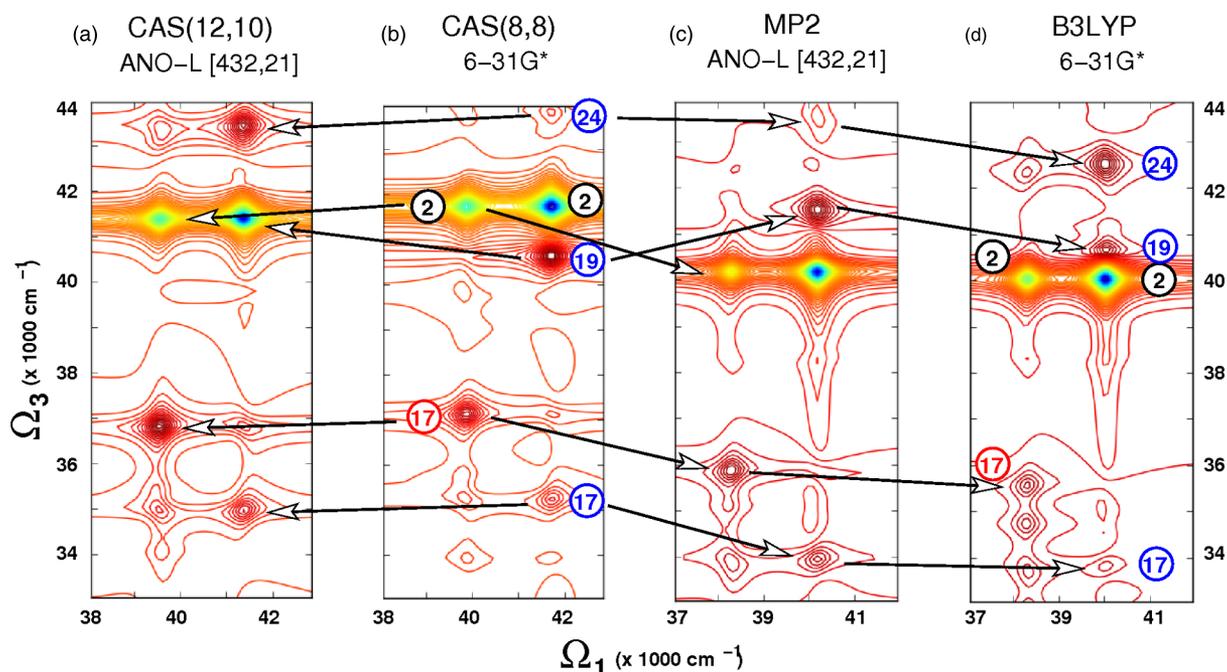


FIG. 7. Dependence of the 2D NUV-pump NUV-probe spectrum on the GS optimization method: CAS(12,10)/ANO-L[432/21] (a) vs. CAS(8,8)/6-31G\* (b) vs. MP2/ANO-L[432/21] (c) vs. B3LYP/6-31G\* (d). Labeling of the electronic transitions according to Fig. 3.

sum-over-states approach)<sup>26,30</sup> so to properly account for the relevant spectral ranges Vis-NUV. The reduction of the level of theory can be operated in different ways employing either (a) a smaller active space, (b) the less expensive RAS scheme (see computational method Sec. II), or (c) a less extended basis set. Such possible modifications should be dictated, apart from the above described computational issue, by the particular purpose for which the data will be used. In the present case, it must be kept in mind the need to account for excitations from the lowest valence occupied orbitals into the highest unoccupied valence orbitals, since the contribution of the corresponding configuration state functions becomes dominant in the higher excited states of the system that can be probed in the NUV-Vis and in particular in the NUV-NUV spectra.

In such a scenario, we have decided to use for all the reduced calculations the same basis set and RAS scheme and to mainly test the performance of the computations with respect to different active spaces. More specifically, the ANO-L basis set contracted to C,N [3s,2p,1d]/H[2s] has been employed, and up to four holes in the RAS1 subspace and up to 4 electrons in the RAS3 were allowed. Another possibility, already employed in other works,<sup>73</sup> would be to use a reduced RAS2 subspace and put the remaining orbitals in RAS1 and RAS3, thereby restricting the number of holes and excitations to two. Unfortunately, already this restricted scheme is beyond the computational capabilities. The adopted computational recipe has already been tested for the adenine system giving very similar results to the analogous CAS calculation.<sup>29</sup> Regarding the active space, three partitions have been used: RAS(4,6|0,0|4,4), in which all the valence  $\pi$  orbitals of the system have been distributed between the RAS1 and RAS3 subspaces, RAS(4,6|0,0|4,3), where the highest unoccupied valence  $\pi$  orbital (i.e., the L + 3) has been excluded from the RAS3, and RAS(4,4|0,0|4,3), in which the H - 4, the O, and the L + 3 orbitals have been deleted from the active space. Actually, also a RAS(4,4|0,0|4,4) space composed by the valence  $\pi$  orbitals excluding orbitals H - 4 and O has been tested, but since the corresponding results strongly deviate from the reference ones, they are reported in the supplementary material.<sup>61</sup>

The RAS(4,6|0,0|4,4) and RAS(4,6|0,0|4,3) active spaces were selected with the idea of reproducing the required data for both the NUV-Vis and the NUV-NUV spectra, while the RAS(4,4|0,0|4,3) scheme has been chosen in order to probe the accessible states in the visible. The operated active space reduction can also be seen as the result of the subsequent elimination of the lowest occupied and the highest unoccupied valence  $\pi$ -orbitals that were found to contribute on average less to the computed excited states in the reference computation.

For all calculations, it has been found that the excitation energies of the  $L_b$  and  $L_a$  states differ significantly from the values obtained at the reference RAS(0,0|12,10|2,12) level of theory (see the supplementary material).<sup>61</sup> As forewarned in Sec. III, the electronic gap between the  $L_b$  and the  $L_a$  states is very sensitive to the dynamic correlation treated perturbatively (see Table II). These problems are mainly due to the overestimated energy stabilization originating from the RASPT2 calculations. In fact, despite reducing the active space leads to the increase of the (both absolute and relative) RASSCF energies (as it is expected for RASSCF being a variational method),

the corresponding RASPT2 energies are decreasing, which means that the RASPT2 correction considerably increases with the reduction of the active space (see the supplementary material).<sup>61</sup> Although the increasing trend of the RASPT2 correction is actually well justified since the reduction of the active space translates into the description of a bigger part of the total correlation at the RASPT2 level, the decreasing of the absolute RASPT2 energies with the reduction of the active space clearly indicates the overestimation of the perturbation treatment. Tentatively, such a behavior can be attributed to the fact that configurations from the expansion set generated by single and double excitations from the reference state (i.e., the CASPT2 first-order wave function) with energies close to the reference CASSCF energy will lead to an overestimation or even singularities of the perturbation energy.<sup>74,75</sup> As perturbation theory holds only for small perturbations, such configurations should be included already in the preceding CASSCF calculation. Thus, the exclusion from the active space of valence  $\pi$ -orbitals which give rise to excited states within the energy window of interest is likely to lead to an overestimation of the CASPT2 energies of the corrected states.

Such an effect would not constitute a grave issue if the overestimation would equally affect all the excited states, since in order to compute the spectra, the quantities that must be correctly reproduced are energy differences and not their absolute values. As clearly shown in the supplementary material and discussed in Sec. III,<sup>61</sup> this is not the case even for the lowest  $L_b$  and  $L_a$  states.

An important diagnosis is that even a full-valence active space (i.e., RAS(4,6|0,0|4,4)) which can be seen as the equivalent of a full valence CAS(12,10) overestimates the energetic stabilization of ionic states. Thus, covalent states, e.g.,  $L_b$ , are reproduced with a red-shift of  $\sim 0.1$  eV from the reference, while ionic states like  $L_a$  can show over-stabilization of up to 0.4 eV. Large deviations can also be found for higher lying states, especially in the NUV range that red-shifts by up to 0.4 eV. Similar results have been observed for other conjugated systems with benzene the most prominent representative.<sup>42</sup> Another observation, which is related to the different accuracy of describing covalent and ionic states and is also in line with our observation for the progressive increase of the active space in the reference calculations, is that the gap between  $L_b$  and  $L_a$  depends largely on the active space, i.e., on the dynamic correlation treated perturbatively.

As discussed in Sec. III, the more natural solution in order to circumvent the described RASPT2 deficiency would be to increase the active space. When the complexity of the system does not allow such a strategy, a solution can be the manual shift of all the energies on the basis of some reference data or the increase of the real/imaginary shift parameter used in the RASPT2 calculations, which has the effect of increasing the electronic gap between the CASPT2 first-order wave function configurations and the reference CASSCF wave function and, hence, the effect of lowering the CASPT2 correlation contribution. In the supplementary material, we present a comparison of the energies of the first 25 states for the different active spaces varying the shift parameters by 0.1 in the range 0.2-0.7 a.u.<sup>61</sup> The mean deviation from the reference results is minimal in the range 0.5-0.6 a.u. and the mean absolute devi-

TABLE II. RASPT2 vertical excitation energies (in eV) at the CAS(8,8)/6-31G\* optimized ground state geometry of adenine in gas-phase, employing different levels of theory. From left to right: RAS(0,0|12,10|2,12)/ANO-L[432/21] with imaginary level shift parameter in the CASPT2 calculation of 0.2, RAS(4,6|0,0|4,4)/ANO-L[321/2] with an imaginary shift of 0.5, RAS(4,6|0,0|4,3)/ANO-L[321/2] with an imaginary shift of 0.6, RAS(4,4|0,0|4,3)/ANO-L[321/2] with an imaginary shift of 0.5, and CAS(4,4)/ANO-L[321/2] with an imaginary shift of 0.2. Corrected values used for the construction of the 2D electronic spectra displayed in Figs. 8 and 9 are given in brackets. Labeling according to Fig. 3.

Label	Reference	(6,4 0,0 4,4)	(6,4 0,0 4,3)	(4,4 0,0 4,3)	(4,4)	Main configuration
2A'(L <sub>1</sub> <sup>+</sup> )L <sub>b</sub>	4.94	4.93	4.67	4.72 (4.96)	4.33 (4.76)	H → L + 1 H - 1 → L
3A'(L <sub>2</sub> <sup>+</sup> )L <sub>a</sub>	5.18	4.99 (5.17)	5.05	5.24	4.73 (5.16)	H → L H - 1 → L + 1
7A'(L <sub>2</sub> <sup>-</sup> )	6.86	6.86	6.89	6.85	6.16 (6.61)	H - 1 → L + 1 H → L
8A'(A <sub>1</sub> <sup>-</sup> )	7.00	6.97	6.88	6.94	...	H - 1 → L + 2 H - 2 → L + 1
9A'(A <sub>2</sub> <sup>-</sup> )	7.39	7.55	7.25	7.63	...	H - 2 → L + 1 H - 1 → L + 2
10A'(B <sub>1</sub> <sup>+</sup> )	7.94	7.72	8.01	8.17	...	H - 3 → L
11A'C	8.19	8.35	8.28	8.86	...	H - 2 → L + 2
12A'(B <sub>2</sub> <sup>+</sup> )	8.53	8.64	8.54	8.87	...	H - 3 → L + 1
13A'	8.83	8.42	8.60	8.69	8.99 (9.42)	H => L
17A'	9.54	9.41	9.17	9.39	8.73 (9.16)	H => L, L + 1
18A'	10.04	10.07	9.79	9.83	...	H - 3 → L + 2 H - 1, H => L + 1, L + 2 0 → L + 2 H - 2 → L + 2
19A'	10.22	10.16	10.43	10.37	9.46 (9.89)	0 → L + 2 H - 1, H => L + 1, L + 2 H => L
20A'(D <sup>+</sup> )	10.31	10.16	...	10.97	...	H => L + 1, L + 2 H - 3 → L + 2
21A'(H <sup>+</sup> )	10.33	10.44	10.10	...	9.68 (10.11)	H => L + 1 H - 4 → L
22A'(G)	10.51	10.49	10.93	...	...	0 → L + 2
23A'(G)	10.56	...	11.14	...	...	0 → L + 2
24A'(H <sup>+</sup> )	10.62	10.55	10.41	...	...	H - 4 → L H - 4 → L + 2

ations from the reference energies are within 0.2 eV. Effectively, covalent states become somewhat overestimated, but ionic states improve substantially, thus globally giving the best compromise for the description of both covalent and ionic states. Hence, the overall accuracy of the energetic position of the peaks cannot be expected to be better than 0.2 eV ( $\sim 1500 \text{ cm}^{-1}$ ).

Consequently, in the present contribution, an imaginary shift parameter equal to 0.5 has been used in the RAS(4,6|0,0|4,4) and RAS(4,4|0,0|4,3) computations, while the same parameter has been set to 0.6 in the RAS(4,6|0,0|4,3) case. For the RAS(4,6|0,0|4,4) and RAS(4,4|0,0|4,3) computations, a further manual shift has been also applied in order to recover

the right splitting between the L<sub>b</sub> and L<sub>a</sub> states: in the RAS(4,6|0,0|4,4), the L<sub>a</sub> has been shifted by  $+1500 \text{ cm}^{-1}$ , while in the RAS(4,4|0,0|4,3) results, the L<sub>b</sub> state position has been moved up by  $2000 \text{ cm}^{-1}$ . The corresponding data are reported in the supplementary material.<sup>61</sup> The NUV-pump/Vis-probe spectra obtained with the chosen active space are reported in Fig. 8, while the corresponding NUV-pump/NUV-probe spectra are reported in Fig. 9. The described procedure, although making use of imaginary shift values higher than the ones suggested in the literature, has its foundation in the comparison with the reference data and allows for the construction of electronic 2D spectra, which are comparable with the ones obtained using the reference data.

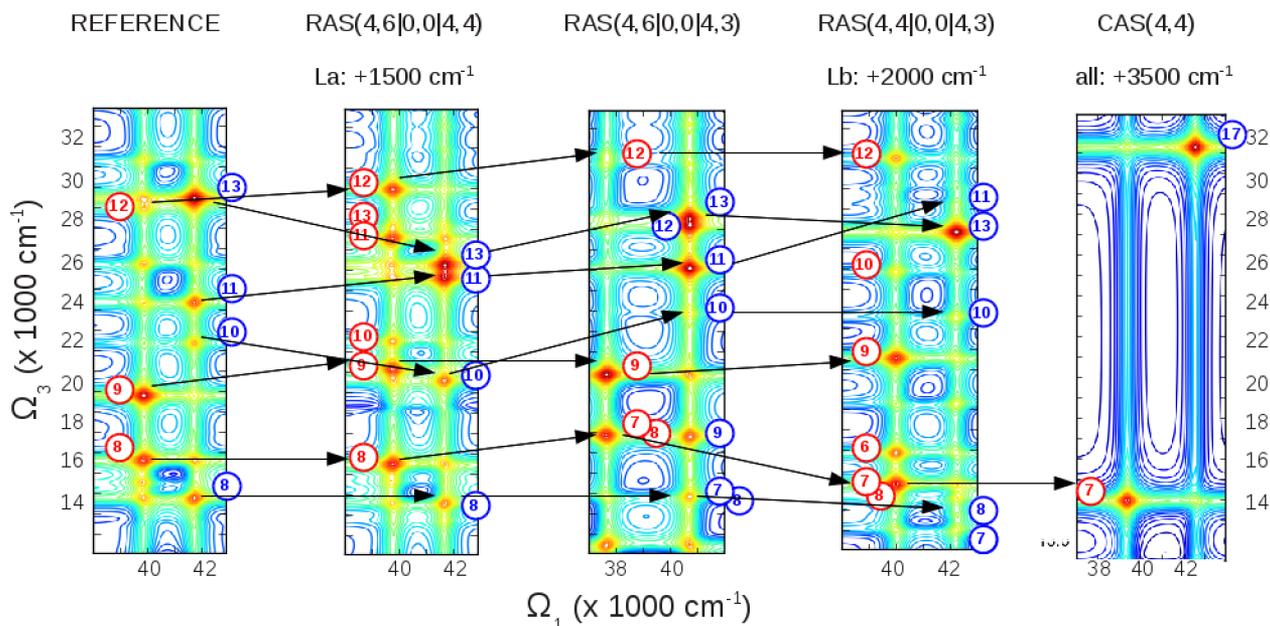


FIG. 8. Comparison of the NUV-pump/Vis-probe 2D spectra of adenine computed at different levels of theory. From left to right: RAS(0,0|12,10|2,12)/ANO-L[432/21], RAS(4,6|0,0|4,4)/ANO-L[321/2], RAS(4,6|0,0|4,3)/ANO-L[312/2], RAS(4,4|0,0|4,3)/ANO-L[321/2], CAS(4,4)/ANO-L[321/2]. The most significant signals appearing in the reference spectrum are clearly labeled, and their fluctuation in the different spectra is highlighted by arrows. Labeling according to Fig. 3.

The reference NUV-pump/Vis-probe 2D spectrum is characterized by two traces at  $\Omega_1$  around  $\sim 39\,900\text{ cm}^{-1}$  and  $\sim 41\,700\text{ cm}^{-1}$ , originated from the electronic excitations involving the  $L_b$  and the  $L_a$  states, respectively. In the following, we will focus only on the positions of the most intense transitions along these traces. Along the  $L_b$  trace, three main signals can be distinguished: an ESA to  $8A'$  at  $\Omega_3 = 16\,600\text{ cm}^{-1}$ , to  $9A'$  at  $\Omega_3 = 19\,700\text{ cm}^{-1}$ , and to  $12A'$  at  $\Omega_3 = 29\,000\text{ cm}^{-1}$ .

The three signals can be characterized in the spectra computed with the reduced active spaces. In particular, the spectrum obtained at the RAS(4,6|0,0|4,4) level nicely reproduces the positions of all main peaks, displaying the largest de-

viation for peak 9, which is blue-shifted by around  $1000\text{ cm}^{-1}$ ; for the spectrum created using the RAS(4,6|0,0|4,3) scheme, a qualitatively correct agreement can still be appreciated with signals 8 and 12 being blue-shifted by around  $1000\text{ cm}^{-1}$  and  $2000\text{ cm}^{-1}$ , respectively, so that peak 8 overlaps with peak 7; the RAS(4,4|0,0|4,3) calculation reproduces closely the features reported for the RAS(4,6|0,0|4,3) spectrum, with the only exception of signal 8, which is now red-shifted with respect to the reference spectrum by  $1000\text{ cm}^{-1}$ . In Figure 8, we also present the spectrum computed with the CAS(4,4) scheme, including in the active space the HOMO - 1, HOMO, LUMO, and LUMO + 1 orbitals, manually shifting all the

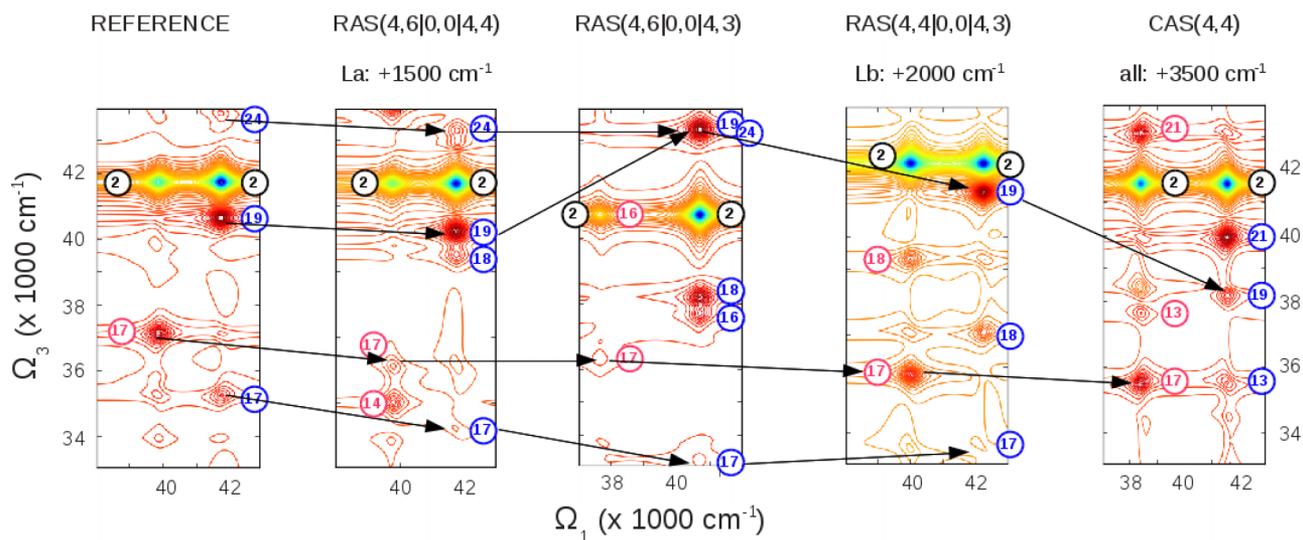


FIG. 9. Comparison of the NUV-pump/NUV-probe 2D spectra of adenine computed at different levels of theory. From left to right: RAS(0,0|12,10|2,12)/ANO-L[432/21], RAS(4,6|0,0|4,4)/ANO-L[321/2], RAS(4,6|0,0|4,3)/ANO-L[321/2], RAS(4,4|0,0|4,3)/ANO-L[321/2], CAS(4,4)/ANO-L[321/2]. The most significant signals appearing in the reference spectrum are clearly labeled, and their evolution in the different spectra is highlighted by arrows. Labeling according to Fig. 3.

values by  $3500\text{ cm}^{-1}$  (see the supplementary material).<sup>61</sup> As clearly shown, such a level of theory is not the most appropriate for a quantitative construction of the NUV-pump/Vis-probe 2D spectra.

Along the  $L_a$  trace of the reference NUV-NUV spectrum, four main signals can be found at  $\Omega_3$  equal to  $14\,700$ ,  $22\,300$ ,  $24\,300$ , and  $29\,400\text{ cm}^{-1}$ , due, respectively, to the electronic promotion into the  $8A'$ ,  $10A'$ ,  $11A'$ , and  $13A'$  states. For all RASSCF schemes, peaks 11 and 13 are blue- and red-shifted, respectively. The  $L_a \rightarrow 11A'$  transition borrows oscillator strength from the  $L_a \rightarrow 13A'$  transition, whereas the level of mixing depends strongly on the computational level as discussed in Sec. III. As shown in Figure 6, a similar shifting is observed comparing the NUV-pump/Vis-probe 2D spectra obtained at the geometries optimized at the CAS level with respect to the one using the MP2 ground state minimum. For the spectrum computed using the RAS(4,6|0,0|4,4) data, both peaks appear degenerate. Peak 10 is red-shifted by around  $1000\text{ cm}^{-1}$ ; peak 8 preserves the right position but it has low intensity. It is not consequently clear if the shifting of the  $13A'$  peak is either the result of the reduced level of theory or is due to the intrinsic difficulty to describe correctly the electronic gap between states  $11A'$  and  $13A'$ . The RAS(4,6|0,0|4,3) scheme manages to recover partially the splitting between peaks 11 and 13; the signal associated with the 10 state is blue-shifted by  $1000\text{ cm}^{-1}$ ; peak 8 overlaps with peak 7, as observed already for the signals along the  $L_b$  trace. The RAS(4,4|0,0|4,3) active space inverts the energetic order of peaks 11 and 13, the state-mixing decreases, and the  $L_a \rightarrow 11A'$  transition recovers its low intensity. Peak 10 is blue-shifted by  $1000\text{ cm}^{-1}$  and signal 8 appears at the correct energy, but both are quite weak.

On the basis of the presented comparisons, we conclude that both the RAS(4,6|0,0|4,4) and the RAS(4,6|0,0|4,3) level of theories are able to qualitatively reproduce the adenine NUV-pump/Vis-probe 2D spectra. The RAS(4,6|0,0|4,3) can however be considered a better choice since it gives a correct  $L_b$ - $L_a$  energy gap (no manual shift is needed); peaks 11 and 13 displayed along the  $L_a$  trace are not degenerate in energy unlike the RAS(4,6|0,0|4,4) scheme; last but not least, it requires a lower computational effort. Consequently, the RAS(4,6|0,0|4,3) level should be used for the computation of the adenine dimer, implying a RAS(4,12|0,0|4,6).

A similar analysis has also been performed in order to find the most appropriate computational level for the construction of NUV-pump/NUV-probe 2D spectra. Naturally, the reference spectrum is described by two traces, one at lower energy associated with GSB, SE, and ESA from the  $L_b$  state and the other related to the corresponding electronic transitions from the  $L_a$  state. Besides the off-diagonal bleach signal at  $\Omega_3 = 41\,900\text{ cm}^{-1}$ , a characteristic ESA peak is present at  $\Omega_3 = 37\,100\text{ cm}^{-1}$  along the  $L_b$  trace, related to the absorption to the  $17A'$  state. The  $L_a$  trace is characterized by three main peaks: at  $\Omega_3$  equal to  $35\,200$ ,  $40\,700$ , and  $43\,900\text{ cm}^{-1}$ , due, respectively, to the electronic promotion into the  $17A'$ ,  $19A'$ , and  $24A'$  states.

The performances of the reduced calculations are displayed in Figure 9. Overall the deviations from the reference computation are more pronounced in this region, with erroneous state ordering and fluctuating TDMs. The  $L_b$ -trace

characteristic signal associated to the  $L_b \rightarrow 17A'$  electronic promotion, although becoming more and more red-shifted as the active space is reduced, is in all computed spectra in a qualitatively correct position. However, the intensity of the transition exhibits pronounced fluctuation with the computational level, being weak at RAS(4,6|0,0|4,4) and RAS(4,6|0,0|4,3) levels and strong at RAS(4,4|0,0|4,3) and CAS(4,4) levels.

Regarding the  $L_a$  trace, the RAS(4,6|0,0|4,4) scheme performs well, with peaks 19 and 24 both being well reproduced, while peak 17 appears red-shifted by  $\sim 1000\text{ cm}^{-1}$  and is weak in intensity. However, we note that the wave function of state 19 has a considerably different composition, and we assign the  $L_a \rightarrow 19A'$  transition mainly based on the large oscillator strength along the  $L_a$  trace which we believe is due to the pronounced contribution from the HOMO, HOMO  $\rightarrow$  LUMO, LUMO CSF to the wave function of state  $19A'$  both in the reference and in the reduced schemes. In the spectrum obtained using the RAS(4,6|0,0|4,3), data pronounced differences are present with respect to the reference one: first, intense peak 19 blue-shifts significantly and appears above the GSB peak, thereby overlapping with peak 24; second, peak 17 is almost red-shifted out of the probing window; third, state 16, which is outside of the probing window in the reference spectrum, blue-shifts and mixes with peak 18 to give rise to intense signal. The RAS(4,4|0,0|4,3) scheme reproduces the position of main peaks 17 and 19 qualitatively well, whereas peak 17 is red-shifted by  $\sim 2000\text{ cm}^{-1}$  and almost dark. At this level of active space reduction, some electronic levels (i.e., states 22-24) are no longer describable within the CASSCF configuration space (see Table II). The CAS(4,4) scheme performs better in the NUV region than in the visible although the agreement is still not completely satisfactory. The energies and intensities of main signals 17 and 19 associated with two-electron excitations among the HOMO - 1, HOMO, LUMO, and LUMO + 1 are qualitatively well reproduced, even if peak 17 strongly red-shifts along the  $L_a$  trace and appears in the Vis window (see Fig. 8). However, the strong blue-shift of nearly  $5000\text{ cm}^{-1}$  associated with peak 13, which appears in the reference spectrum around  $30\,000\text{ cm}^{-1}$  (i.e., in the Vis spectrum, see Fig. 8), together with the missing signals from transitions like 24 proves that CAS(4,4) can be used only to reproduce in a crude approximation doubly excited states in the NUV<sup>29</sup> and is otherwise insufficient to reproduce the correct energy positions of the accessible adenine excited states in the Vis to NUV.

On the basis of the performed comparison, we conclude that the RAS(4,6|0,0|4,4) computation is able to qualitatively reproduce the NUV-pump/NUV-probe 2D spectrum of adenine and could be applied to compute poly-adenine aggregates.

In order to move towards the application of the above-calibrated protocol to dimeric systems, several extra criteria have to be satisfied besides a good agreement of energies and TDMs: the absence of symmetry in a realistic aggregate, the averaging of the CASSCF wave function over a larger number of states (ranging over 100 states), and the influence of electronic states with mixed doubly excited and CT character in the dimer computations, whose presence could potentially perturb the monomer-like states. In particular, it is questionable

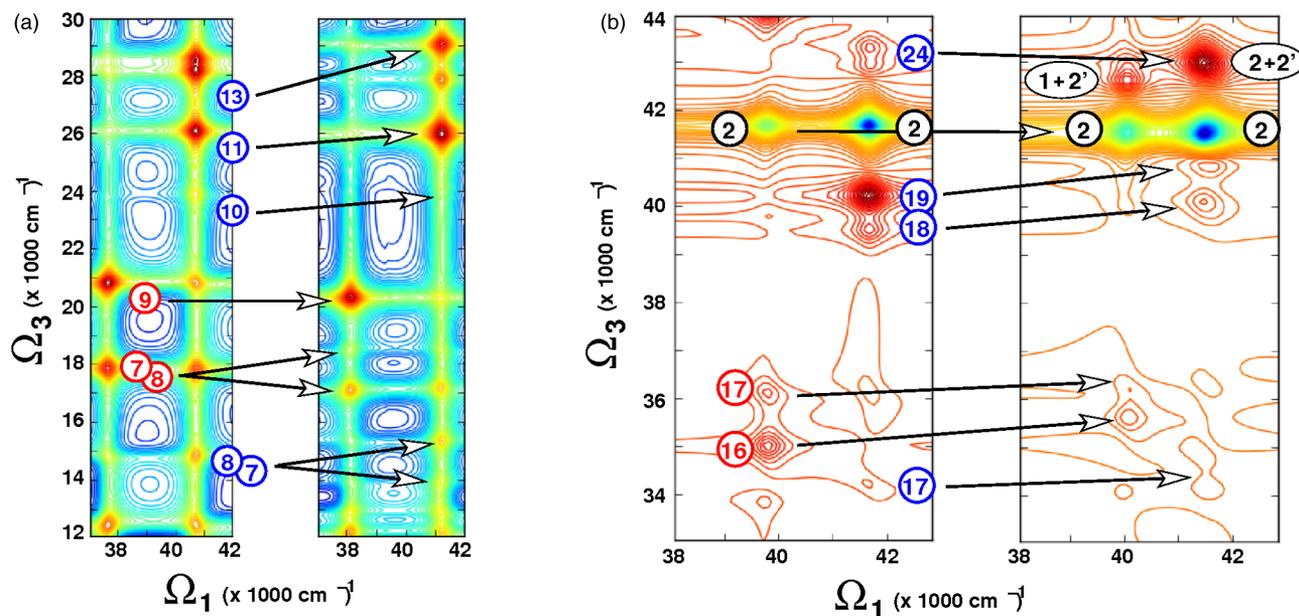


FIG. 10. Comparison between the gas-phase spectra of the adenine monomer and unstacked homodimer. In the left-hand side, the Vis SS-PT2(shift = 0.6)//SA-25-RAS(4,6|0,0|4,3) is compared against SS-PT2(shift = 0.6)//SA-150-RAS(4,12|0,0|4,6). In the right-hand side, the NUV SS-PT2(shift = 0.5)//SA-25-RAS(4,6|0,0|4,4) is compared against SS-PT2(shift = 0.5)//SA-150-RAS(4,12|0,0|4,8). Detailed tables with leading wave function contributions, energies, TDM magnitudes are provided in the supplementary material.<sup>61</sup>

if the different computational schemes provide a converging description of the energies of the CT level for which we do not have reference values. In the following, we evaluate the effect of these parameters on the energetic positions and TDMs of the states.

For this purpose, an *in vacuo* adenine dimer was generated by replicating the adenine molecule. The distance between the two chromophores was set to approximately 10 Å to minimize any inter-chromophore interactions. In addition, the molecules were rotated arbitrary in space and the calculation was performed without any symmetry constraints. From this starting geometry, we have computed the NUV/Vis and NUV/NUV spectra at the SA-100-RAS(4,12|0,0|4,6) and SA-150-RAS(4,12|0,0|4,8) level of theories, respectively, and compared them to the corresponding RAS(4,6|0,0|4,3) and RAS(4,6|0,0|4,4) spectra of the monomer (see Fig. 10).

As shown in Fig. 10(a), comparing the NUV/Vis spectra, the main signals along the  $L_b$  and  $L_a$  traces characterizing the monomeric species are nicely reproduced in the gas-phase dimer, with the main difference being the splitting of the 7 and 8 peaks. We note that the observed splitting between peaks 7 and 8, as well as the more pronounced splitting of peaks 11 and 13, gives a much better agreement with the reference spectrum of the gas-phase monomer (Fig. 8). Similar  $L_b$  and  $L_a$  vertical excitation energies have been obtained out of both the monomer and dimer, which consequently leads to a comparable  $L_b$  and  $L_b$  splitting in the two corresponding spectra. The presence of only two traces in the homodimer spectrum reflects the missing interaction between the monomers, absorbing the two  $L_a$  and two  $L_b$  states within the same trace (see the supplementary material).<sup>61</sup> We can conclude that the lack of symmetry and the computation of a large number of roots in the RAS averaging procedure do not affect significantly the NUV/Vis accessible monomer-like excited states.

A similar conclusion regarding the NUV/NUV accessible monomer-like excited states can be drawn by inspection of Fig. 10(b). In fact, the same signals appearing in the monomer at the NUV/NUV range can be appreciated at around the same positions in the corresponding dimeric NUV/NUV spectrum. However, there is one alerting observation. Apparent small deviations of  $\sim 0.05$  eV at CASPT2 level for mixed doubly excited states lead to the appearance of ESA which should not be present in the non-interacting system (overlapping peaks  $1 + 2'$  and  $1' + 2$  along the  $L_b$  trace and peak  $2 + 2'$  along the  $L_a$  trace). They correspond to transitions to states where both monomers are excited simultaneously, whereas the probe pulse excites the bright  $L_a$  transition. These signals represent a characteristic signature of an interacting system. The electronic separation between the ESA and the corresponding off-diagonal bleach signals (quartic splitting) is an indicator of the coupling strength, as both peaks should coincide and cancel in a non-interacting system. We note that this condition is fulfilled at RASSCF level, yet the CASPT2 correction introduces a splitting. The inaccuracy of the CASPT2 energies is within the energy range usually observed for quartic couplings yet they are due to the limitations of the method by not being size-consistent.<sup>76</sup>

As mentioned above, we do not have reference values for the energies of the CT states in the adenine dimer. The latter are known to show large fluctuations correlated to the level of geometry optimization, the treatment of the solute-solvent interactions, and the excited state computations.<sup>77</sup> As the RAS(4,6|0,0|4,4) scheme was found to reproduce overall the electronic levels in the monomer up to 10 eV (Table II) and applying the scheme to the dimer (implying RAS(4,12|0,0|4,8)) does not affect the properties of the monomer-like states (Fig. 10), we assume that CT states in the dimer are described with similar accuracy. We encounter a trend of blue-shifting

TABLE III. Charge transfer energetics (in eV) present in the unstacked adenine homodimer at three different computational levels.

	RAS(4,12 0,0 4,8) imaginary shift = 0.5 a.u.	RAS(4,12 0,0 4,6) imaginary shift = 0.6 a.u.	RAS(4,8 0,0 4,6) imaginary shift = 0.5 a.u.
CT <sub>1</sub>	8.35	8.50	8.64
CT <sub>2</sub>	8.43	8.57	8.73
CT <sub>3</sub>	9.08	9.15	9.26
CT <sub>4</sub>	9.17	9.24	9.37
CT <sub>5</sub>	9.49	9.60	9.69
CT <sub>6</sub>	9.59	9.70	9.78
CT <sub>7</sub>	9.90	10.03	10.12
CT <sub>8</sub>	9.95	10.08	10.18
CT <sub>9</sub>	10.22	10.26	10.31
CT <sub>10</sub>	10.28	10.37	10.42
CT <sub>11</sub>	10.33	10.39	
CT <sub>12</sub>	10.40	10.50	

the energies of the CT states with reducing the active space but an overall good agreement is observed (see Table III). In particular, the CT states are uniformly blue-shifted by  $\sim 0.1$  eV for the RAS(4,12|0,0|4,6) and by  $\sim 0.3$  eV at the RAS(4,8|0,0|4,6) level of theory. The agreement between the results of the RAS(4,12|0,0|4,8) and the RAS(4,12|0,0|4,6) schemes motivates the final conclusion that *the RAS(4,12|0,0|4,8) scheme is most suited for describing the electronic structure of adenine monomer and multimers up to 10 eV, while the reduced RAS(4,12|0,0|4,6) scheme can be used for describing states up to the ionization potential.*

## V. CONCLUSIONS

The present study is an indispensable step towards translating the computational mechanistic analysis (i.e., energetic profiles and geometrical deformations) into experimental observables through the simulation of pump-probe and two-dimensional electronic spectra from first principles. Different levels of theory based on the RASSCF/RASPT2 framework have been tested in order to get the best compromise between accuracy and computational efforts for the simulation of nonlinear electronic spectroscopy experiments of the adenine monomer and homodimer. Following a previously reported recipe,<sup>26</sup> which takes into account valence-Rydberg mixing problems, the active space of adenine has been progressively enlarged yielding the most accurate estimate within the RASSCF/RASPT2 framework corresponding to a RAS(0,0|12,10|2,12)/SS-RASPT2, which is here used as a reference. The computed transition energies and dipole moments here provided can be used for building parameterized exciton models for DNA/RNA nucleobases, similarly to what has been done in the past for aminoacids.<sup>15</sup>

Compared with the reference values, we have obtained that the following reduced active spaces, RAS(4,6|0,0|4,4) and RAS(4,6|0,0|4,3), provide reasonable estimates for the NUV-pump/NUV-probe and NUV-pump/Vis-probe windows, respectively. Their performance has been then evaluated for an unstacked adenine homodimer in gas-phase, which is built by duplicating the active space over the two adenine moieties thus yielding the RAS(4,12|0,0|4,8) and RAS(4,12|0,0|4,6)

schemes. The assessment renders these computational schemes eligible for simulating pump-probe and 2D electronic spectra in a more general environment and conformation with an accuracy of  $\sim 2000$  cm<sup>-1</sup> ( $\sim 0.25$  eV). The latter is here demonstrated through the generation of the 2D NUV-pump/NUV-probe and NUV-pump/Vis-probe spectra of adenine. The results suggest that bidimensional spectra provide a way to tackle dimeric systems within an *ab initio* approach. Upcoming calculations are being performed to elucidate the role of the conformation and environment and its effect on the essential photoinduced events occurring in the adenine homodimer related to its photostability.

## ACKNOWLEDGMENTS

M.G. and G.C. acknowledge support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). A.G. acknowledges financial support from Italian Ministry of Education and Research (MIUR)-Grant No. RBFR1248UI. We also acknowledge the use of HPC resources of the “Pôle Scientifique de Modélisation Numérique” at the ENS-Lyon, France.

<sup>1</sup>C. E. Crespo-Hernandez, B. Cohen, P. M. Hare, and B. Kohler, *Chem. Rev.* **104**(4), 1977–2019 (2004).

<sup>2</sup>J. Chen, Y. Zhang, and B. Kohler, in *Photoinduced Phenomena in Nucleic Acids II*, edited by M. Barbatti, A. C. Borin, and S. Ulbrich (Springer International Publishing, 2015), Vol. 356, pp. 39–87.

<sup>3</sup>L. E. Buchanan, J. K. Carr, A. M. Fluit, A. J. Hoganson, S. D. Moran, J. de Pablo, J. L. Skinner, and M. T. Zanni, *Proc. Natl. Acad. Sci.* **111**(16), 5796–5801 (2014).

<sup>4</sup>A. Ghosh, J.-J. Ho, A. L. Serrano, D. R. Skoff, T. Zhang, and M. T. Zanni, *Faraday Discuss.* **177**, 493 (2015).

<sup>5</sup>M. R. Panman, C. N. van Dijk, H. Meuzelaar, and S. Woutersen, *J. Chem. Phys.* **142**(4), 041103 (2015).

<sup>6</sup>B. A. West, J. M. Womick, and A. M. Moran, *J. Phys. Chem. A* **117**(29), 5865–5874 (2012).

<sup>7</sup>B. A. West, J. M. Womick, and A. M. Moran, *J. Phys. Chem. A* **115**(31), 8630–8637 (2011).

<sup>8</sup>B. A. West and A. M. Moran, *J. Phys. Chem. Lett.* **3**(18), 2575–2581 (2012).

<sup>9</sup>P. Krause, S. Matsika, M. Kotur, and T. Weinacht, *J. Chem. Phys.* **137**(22), 22A537 (2012).

<sup>10</sup>R. B. Varillas, A. Candeo, D. Viola, M. Garavelli, S. De Silvestri, G. Cerullo, and C. Manzoni, *Opt. Lett.* **39**(13), 3849–3852 (2014).

<sup>11</sup>G. D. Scholes, *Proc. Natl. Acad. Sci.* **111**(28), 10031–10032 (2014).

- <sup>12</sup>V. M. Huxter, T. A. A. Oliver, D. Budker, and G. R. Fleming, *Nat. Phys.* **9**(11), 744–749 (2013).
- <sup>13</sup>T. A. A. Oliver, N. H. C. Lewis, and G. R. Fleming, *Proc. Natl. Acad. Sci.* **111**(28), 10061–10066 (2014).
- <sup>14</sup>T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, *Chem. Rev.* **112**(1), 543–631 (2012).
- <sup>15</sup>J. Jiang and S. Mukamel, *Phys. Chem. Chem. Phys.* **13**(6), 2394–2400 (2011).
- <sup>16</sup>J. Jiang, D. Abramavicius, B. M. Bulheller, J. D. Hirst, and S. Mukamel, *J. Phys. Chem. B* **114**(24), 8270–8277 (2010).
- <sup>17</sup>W. Zhuang, T. Hayashi, and S. Mukamel, *Angew. Chem., Int. Ed.* **48**(21), 3750–3781 (2009).
- <sup>18</sup>A. Nenov, I. Rivalta, G. Cerullo, S. Mukamel, and M. Garavelli, *J. Phys. Chem. Lett.* **5**, 767–771 (2014).
- <sup>19</sup>D. Nachtigallova, P. Hobza, and H.-H. Ritze, *Phys. Chem. Chem. Phys.* **10**(37), 5689–5697 (2008).
- <sup>20</sup>D. Nachtigallova, T. Zelený, M. Ruckebauer, T. Müller, M. Barbatti, P. Hobza, and H. Lischka, *J. Am. Chem. Soc.* **132**(24), 8261–8263 (2010).
- <sup>21</sup>M. T. Oakley and J. D. Hirst, *J. Am. Chem. Soc.* **128**(38), 12414–12415 (2006).
- <sup>22</sup>G. D. Scholes and K. P. Ghiggino, *J. Phys. Chem.* **98**(17), 4580–4590 (1994).
- <sup>23</sup>S. Mukamel, R. Oszwaldowski, and D. Abramavicius, *Phys. Rev. B* **75**(24), 245305 (2007).
- <sup>24</sup>I. Rivalta, A. Nenov, O. Weingart, G. Cerullo, M. Garavelli, and S. Mukamel, *J. Phys. Chem. B* **118**(28), 8396–8405 (2014).
- <sup>25</sup>I. Rivalta, A. Nenov, G. Cerullo, S. Mukamel, and M. Garavelli, *Int. J. Quantum Chem.* **114**(2), 85–93 (2014).
- <sup>26</sup>A. Nenov, I. Rivalta, S. Mukamel, and M. Garavelli, *Comput. Theor. Chem.* **1040–1041**, 295–303 (2014).
- <sup>27</sup>A. Nenov, S. a Beccara, I. Rivalta, G. Cerullo, S. Mukamel, and M. Garavelli, *ChemPhysChem* **15**(15), 3282–3290 (2014).
- <sup>28</sup>D. Abramavicius, B. Palmieri, D. V. Voronine, F. Sanda, and S. Mukamel, *Chem. Rev.* **109**, 2350 (2009).
- <sup>29</sup>A. Nenov, J. Segarra-Martí, A. Giussani, I. Conti, I. Rivalta, E. Dumont, V. K. Jaiswal, S. F. Altavilla, S. Mukamel, and M. Garavelli, *Faraday Discuss.* **177**, 345–362 (2015).
- <sup>30</sup>A. Giussani, *J. Chem. Theory Comput.* **10**(9), 3987–3995 (2014).
- <sup>31</sup>A. Giussani, J. Segarra-Martí, D. Roca-Sanjuán, and M. Merchán, in *Photo-induced Phenomena in Nucleic Acids I*, edited by M. Barbatti, A. C. Borin, and S. Ullrich (Springer International Publishing, 2015), Vol. 355, pp. 57–97.
- <sup>32</sup>L. Serrano-Andrés and M. Merchán, *J. Mol. Struct.: THEOCHEM* **729**, 99–108 (2005).
- <sup>33</sup>N. Forsberg and P. A. Malmqvist, *Chem. Phys. Lett.* **274**(1-3), 196–204 (1997).
- <sup>34</sup>G. Ghigo, B. O. Roos, and P.-Å. Malmqvist, *Chem. Phys. Lett.* **396**(1–3), 142–149 (2004).
- <sup>35</sup>V. Varyazov, P. Å. Malmqvist, and B. O. Roos, *Int. J. Quantum Chem.* **111**(13), 3329–3338 (2011).
- <sup>36</sup>J. Chen and B. Kohler, *J. Am. Chem. Soc.* **136**(17), 6362–6372 (2014).
- <sup>37</sup>C. Su, C. T. Middleton, and B. Kohler, *J. Phys. Chem. B* **116**, 10266–10274 (2012).
- <sup>38</sup>C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernandez, and B. Kohler, *Annu. Rev. Phys. Chem.* **60**, 217–239 (2009).
- <sup>39</sup>D. Markovitsi, T. Gustavsson, and I. Vayá, *J. Phys. Chem. Lett.* **1**(22), 3271–3276 (2010).
- <sup>40</sup>F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P. A. Malmqvist, P. Neogrady, T. B. Pedersen, M. Pitonak, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Varyazov, and R. Lindh, *J. Comput. Chem.* **31**(1), 224–247 (2010).
- <sup>41</sup>F. Aquilante, T. B. Pedersen, V. Varyazov, and R. Lindh, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **3**(2), 143–149 (2013).
- <sup>42</sup>A. D. Becke, *J. Chem. Phys.* **98**(7), 5648–5652 (1993).
- <sup>43</sup>A. S. Chatterley, C. W. West, G. M. Roberts, V. G. Stavros, and J. R. R. Verlet, *J. Phys. Chem. Lett.* **5**(5), 843–848 (2014).
- <sup>44</sup>S. F. Altavilla, J. Segarra-Martí, A. Nenov, I. Conti, I. Rivalta, and M. Garavelli, *Front. Chem.* **3**, 29 (2015).
- <sup>45</sup>K. Andersson, P. A. Malmqvist, and B. O. Roos, *J. Chem. Phys.* **96**(2), 1218–1226 (1992).
- <sup>46</sup>P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, *J. Chem. Phys.* **128**(20), 204109 (2008).
- <sup>47</sup>P.-Å. Malmqvist and B. O. Roos, *Chem. Phys. Lett.* **155**(2), 189–194 (1989).
- <sup>48</sup>P. Å. Malmqvist, B. O. Roos, and B. Schimmelpfennig, *Chem. Phys. Lett.* **357**(3–4), 230–240 (2002).
- <sup>49</sup>F. Aquilante, P.-Å. Malmqvist, T. B. Pedersen, A. Ghosh, and B. O. Roos, *J. Chem. Theory Comput.* **4**(5), 694–702 (2008).
- <sup>50</sup>T. Pedersen, F. Aquilante, and R. Lindh, *Theor. Chem. Acc.* **124**(1–2), 1–10 (2009).
- <sup>51</sup>J. Bostrom, M. G. Delcey, F. Aquilante, L. Serrano-Andrés, T. B. Pedersen, and R. Lindh, *J. Chem. Theory Comput.* **6**(3), 747–754 (2010).
- <sup>52</sup>L. Serrano-Andrés, M. Merchán, and A. C. Borin, *Chem. - Eur. J.* **12**(25), 6559–6571 (2006).
- <sup>53</sup>L. Serrano-Andrés, M. Merchán, and A. C. Borin, *J. Am. Chem. Soc.* **130**, 2473–2484 (2008).
- <sup>54</sup>M. Merchán, R. González-Luque, T. Climent, L. Serrano-Andrés, E. Rodríguez, M. Reguero, and D. Pelaez, *J. Phys. Chem. B* **110**(51), 26471–26476 (2006).
- <sup>55</sup>D. Roca-Sanjuán, M. Rubio, M. Merchán, and L. Serrano-Andrés, *J. Chem. Phys.* **125**(8), 084302 (2006).
- <sup>56</sup>D. Roca-Sanjuán, M. Merchán, L. Serrano-Andrés, and M. Rubio, *J. Chem. Phys.* **129**(9), 095104 (2008).
- <sup>57</sup>I. González-Ramírez, J. Segarra-Martí, L. Serrano-Andrés, M. Merchán, M. Rubio, and D. Roca-Sanjuán, *J. Chem. Theory Comput.* **8**(8), 2769–2776 (2012).
- <sup>58</sup>J. Segarra-Martí, M. Merchán, and D. Roca-Sanjuán, *J. Chem. Phys.* **136**, 244306 (2012).
- <sup>59</sup>S.-W. Chang and H. A. Witek, *J. Chem. Theory Comput.* **8**(11), 4053–4061 (2012).
- <sup>60</sup>J. R. Platt, *J. Chem. Phys.* **17**(5), 484–495 (1949).
- <sup>61</sup>See supplementary material at <http://dx.doi.org/10.1063/1.4921016> for further details on the theoretical framework employed for generating the spectra, as well as for the tables with the energy values and transition dipole moments obtained with the different levels of theory reported.
- <sup>62</sup>M. P. Fülcher, L. Serrano-Andrés, and B. O. Roos, *J. Am. Chem. Soc.* **119**(26), 6168–6176 (1997).
- <sup>63</sup>M. Klessinger and J. Michl, *Excited States and Photochemistry of Organic Molecules* (VCH Publishers Inc., New York, 1995).
- <sup>64</sup>F. Santoro, R. Impropa, T. Fahleson, J. Kauczor, P. Norman, and S. Coriani, *J. Phys. Chem. Lett.* **5**(11), 1806–1811 (2014).
- <sup>65</sup>L. Serrano-Andrés, M. Merchán, and A. C. Borin, *Proc. Natl. Acad. Sci. U.S.A.* **103**(23), 8691–8696 (2006).
- <sup>66</sup>M. Barbatti and S. Ullrich, *Phys. Chem. Chem. Phys.* **13**(34), 15492–15500 (2011).
- <sup>67</sup>S. Matsika, *J. Phys. Chem. A* **109**(33), 7538–7545 (2005).
- <sup>68</sup>W. Credo Chung, Z. Lan, Y. Ohtsuki, N. Shimakura, W. Domcke, and Y. Fujimura, *Phys. Chem. Chem. Phys.* **9**(17), 2075–2084 (2007).
- <sup>69</sup>J. Lorentzon, M. P. Fülcher, and B. O. Roos, *J. Am. Chem. Soc.* **117**(36), 9265–9273 (1995).
- <sup>70</sup>M. P. Fülcher and B. O. Roos, *J. Am. Chem. Soc.* **117**(7), 2089–2095 (1995).
- <sup>71</sup>K. Kleiner, D. Nachtigallova, and M. S. de Vries, *Int. Rev. Phys. Chem.* **32**(2), 308–342 (2013).
- <sup>72</sup>A. Giussani, M. Merchán, D. Roca-Sanjuán, and R. Lindh, *J. Chem. Theory Comput.* **7**(12), 4088–4096 (2011).
- <sup>73</sup>T. G. Schmalz, L. Serrano-Andrés, V. Sauri, M. Merchán, and J. M. Oliva, *J. Chem. Phys.* **135**(19), 194103 (2011).
- <sup>74</sup>B. O. Roos, K. Andersson, M. P. Fülcher, L. Serrano-Andrés, K. Pierloot, M. Merchán, and V. Molina, *J. Mol. Struct.: THEOCHEM* **388**, 257–276 (1996).
- <sup>75</sup>B. O. Roos and K. Andersson, *Chem. Phys. Lett.* **245**(2-3), 215–223 (1995).
- <sup>76</sup>B. O. Roos, K. Andersson, M. P. Fülcher, P. A. Malmqvist, L. Serrano-Andrés, K. Pierloot, and M. Merchán, *Adv. Chem. Phys.* **93**, 219–331 (1996).
- <sup>77</sup>J. Segarra-Martí and P. B. Coto, *Phys. Chem. Chem. Phys.* **16**(46), 25642–25648 (2014).