First-principles characterization of the singlet excited state manifold in DNA/RNA nucleobases

ARTICLE

First-principles characterization of the singlet excited state manifold in DNA/RNA nucleobases

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Vishal K. Jaiswal, [a,b] Javier Segarra-Martí,*[b,c,d] Marco Marazzi,[e,f,g,h] Elena Zvereva,[†e,f,i] Xavier Assfeld,[e,f] Antonio Monari,[e,f] Marco Garavelli,*[a] and Ivan Rivalta*[a,b]

An extensive theoretical characterization of the singlet excited state manifold of the five canonical DNA/RNA nucleobases (thymine, cytosine, uracil, adenine and guanine) in gas-phase is carried out with time-dependent density functional theory (TD-DFT) and restricted active space second-order perturbation theory (RASPT2) approaches. Both ground state and excited state absorptions are analyzed and compared between these different theoretical approaches, assessing the performance of the hybrid B3LYP and CAM-B3LYP (long-range corrected) functionals with respect to the RASPT2 reference. By comparing the TD-DFT estimates with our reference for high-lying excited states, we are able to narrow down specific energetic windows where TD-DFT may be safely employed to qualitatively reproduce the excited state absorption (ESA) signals registered in non-linear and time-resolved spectroscopy for monitoring photoinduced phenomena. Our results show a qualitative agreement between the RASPT2 reference and the B3LYP computed ESAs of pyrimidines in the near-IR/Visible spectral probing window while for purines the agreement is limited to the near-IR ESAs, with generally larger discrepancies obtained with the CAM-B3LYP functional. This outcome paves the way for appropriate application of cost-effective TD-DFT approaches to simulate linear and non-linear spectroscopies of realistic multichromophoric DNA/RNA systems with biological and nanotechnological relevance.

Introduction

The study of the spectroscopic response of DNA/RNA remains a central topic towards understanding the photo-damaging mechanisms triggered in the genetic material upon ultraviolet (UV) light exposure as well as its photostability.[1-3] A deep understanding of the photophysics of the isolated choromophores, i.e. DNA/RNA nucleobases,[4-6] is essential as they represent the fundamental building blocks of nucleic acids and provides minimal models also allowing to disentangle the effects due to the chromophoric coupling in more complex structures (such as base pairs and oligomers). Over the years plenty of efforts have been devoted to characterize such spectral fingerprints and their capacity to monitor the complex excited state dynamics triggered upon UV light absorption.[7-14] Experimentally, significant advances have been made over the last two decades, where a pronounced increase in the spectral and temporal resolution has been witnessed. This has enhanced our understanding of the underlying photoinduced phenomena occurring in DNA at extremely short (and therefore experimentally challenging), sub-ps times.[15-19] These novel techniques currently range from standard pump-probe (PP) experiments in both the electronic (UV and visible, Vis)[20, 21] and infrared (IR)[15, 17, 22] regimes, to photoelectron[18] and fluorescence up-conversion[23] spectroscopies, more advanced multidimensional nonlinear electronic spectroscopic setups such as two-dimensional electronic spectroscopy (2DES),[19] and even time-resolved experiments featuring X-ray probing laser pulses.[24, 25] The sudden increase in spatial and temporal resolution of spectroscopic data is encompassed by the need for well-calibrated theoretical models to explain the plethora of spectroscopic signals recorded in these advanced experiments and mainly associated to high-lying electronic excited state manifold that remains relatively unexplored in the literature.

[a] Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy.
[b] Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France.
[c] Molecular Sciences Research Hub, White City Campus, Imperial College London, 80 Wood Lane, W12 0BZ London, UK.
[d] Instituto de Ciencia Molecular, Universitat de Valencia, P.O.Box 22085 Valencia, Spain.
[e] Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France.
[f] CNRS, Laboratoire de Physique et Chimie Théoriques. Vandoeuvre-lès-Nancy, 54506 France
[g] Department of Analytical Chemistry, Physical Chemistry and Chemical Engineering, Ctra, Madrid-Barcelona Km. 33,600, E-28805 Alcalá de Henares (Madrid), Spain.
[h] Chemical Research Institute "Andrés M. del Río" (IQAR), Universidad de Alcalá, E-28871 Alcalá de Henares (Madrid), Spain
[i] A.E. Arbuzov Institute of Organic and Physical Chemistry, Kazan Scientific Center, Russian Academy of Sciences

*E-mail: j.segarra-marti@imperial.ac.uk; marco.garavelli@unibo.it; i.rivalta@unibo.it
† Current address: IFP Energies Nouvelles, 1 - 4 Avenue de Bois-Préau, 92852 Rueil-Malmaison, France

From a computational standpoint, the models initially formulated to explain the photophysical properties of DNA/RNA nucleobases considered isolated nucleobases in the gas-phase and employed less correlated methods (e.g. limiting the size of the active space) than those exploitable nowadays, due to their high computational cost. These methods have been used for (static) characterization of critical points on the potential energy surfaces (PES) and the minimum energy path (MEP) associated to the excited states decays in single nucleobases.[26-31] Significant advances have been made over the last decade in three connected fronts: i) highly correlated approaches are currently used in systematic studies to rationalize DNA photoinduced phenomena, characterizing excited states PES and interstates crossings; [32-34] ii) non-adiabatic molecular dynamics can now be employed to obtain a time-dependent estimate of the ultrafast excited state evolution instead of relying on static studies;[35-37] and iii) hybrid quantum mechanics/molecular mechanics (QM/MM) approaches could be used at reasonable costs to provide description of the chromophore's molecular surrounding for a proper inclusion of environmental effects.[38-42] Despite all these advances, cost-effective approaches are still sought and must be used in order to tackle more realistic DNA systems, such as single/double-strands or oligomers. In this regard, time-dependent density functional theory (TD-DFT) has been shown to be a unique venue to pursue,[43-47] given its favorable computational cost, especially in light of the active development for its applicability in large-scale applications.[48]

We have recently performed a series of studies on the high-energy electronic states manifold of all canonical DNA/RNA nucleobases, including pyrimidines (uracil, thymine and cytosine) and purines (adenine and guanine), based on highly correlated multiconfigurational/multireference methods.[49-51] These studies aimed to characterize the high-lying excited states likely involved in advanced non-linear electronic spectroscopies. On the other hand, while TD-DFT methodologies have been largely employed to investigate low-lying excited states of nucleobases,[43-47] a comprehensive TD-DFT study of the high-lying singlet excited states of nucleobases is currently lacking. Thus, the main goal of this work is to evaluate the accuracy of cost-effective TD-DFT approaches in computing the high-lying electronic excited states in comparison to more computationally expensive multiconfigurational/multireference methods and assess whether TD-DFT is well-suited to simulate time-resolved non-linear electronic spectra. In particular, we characterize the singlet electronic states manifold of all canonical DNA/RNA nucleobases in gas-phase up to ca. 8.5 eV, i.e. close to their ionization energies. We employed the widely used B3LYP[52, 53] hybrid exchange–correlation functional and evaluated the effect of long-range corrections, by comparison with the CAM-B3LYP[54] functional, and of basis sets size, by assessing the performance of the Pople's polarized double-ζ 6-31G**[55, 56] basis set against the, computationally more expensive augmented triple-ζ 6-311++G**[55-57] basis sets. The proposed benchmarking of TD-DFT against strongly correlated restricted active space second-order perturbation theory (RASPT2) computations allows estimating the spectral windows where TD-DFT intrinsically retains an acceptable degree of accuracy. This knowledge will be crucial when studying the high-energy electronic states manifold of large and realistic molecular systems involving DNA/RNA nucleobases, which is notably responsible for the observed spectral signatures in time-resolved spectroscopies. Both RASPT2 and TD-DFT computations performed in this work are compared with available experimental data. The most comprehensive linear absorption spectra of isolated bases are provided by Yamada et al.[58] and Sinsheimer et al.[59], as recorded on sublimed films. The electronic spectra from sublimed films are always red-shifted compared to available vapor-phase spectra, as shown in the works of Clark et al. (adenine and uracil) [60], Li et al. (adenine) [61], Nowak et al.(uracil)[62] and Ito et al. (uracil)[63]. As a matter of fact, the absorption spectra on sublimed films are very similar to those recorded on solid thin films by Gomez et al.[64] Still, the absorption spectra from sublimed films reported by Yamada et al.[58] have been recorded up to high absorption energies, allowing a qualitative comparison with computed transition energies for high-lying single excited states (above 6.0 eV). As guanine has the lowest vapor pressure amongst canonical nucleobases,[60] its vapor-phase studies are difficult to undertake. However, matrix isolation spectra by Polewski et al.[65] in nitrogen-matrix at low temperature (15 K) provides UV absorption spectra for isolated guanine molecules. In addition to guanine, the matrix isolated spectra are also available for adenine in nitrogen-matrix (Polewski et al.[66]) and cytosine in argon-matrix by Bazsó et al.[67] For cytosine and thymine, electronic spectra obtained through electron impact energy loss (EEEL) methods have been also reported by Abouaf et al.[68, 69] The EEEL spectrum of cytosine features a low-energy absorption band in line with that recorded in argon-matrix and, as the EEEL spectrum of thymine, it provides absorption bands up to the vacuum-UV, allowing a comparison with computed excitations in gas-phase. Despite a quantitative comparison between experimental and theoretical data is beyond the scope of this work, as it would require accurate simulations of the absorption band shapes,[70, 71] the experimental linear absorption bands at high energies represent an important (although qualitative) reference for our computations in gas-phase.

Results and discussion

1.Adenine

1.1 Ground state absorptions

A series of experimental UV absorption spectra in gas-phase are available for adenine,[58-61, 66] which are reported in Figure 1, along with a comparison of vertical $S_0 \rightarrow S_n$ excitation energies and corresponding oscillator strengths computed in gas-phase with different theoretical methods. As previously reported,[51] the computed RASPT2 vertical $S_0 \rightarrow S_n$ energies for the first two excited states of adenine, i.e. L_b (S_1) and L_a (S_2), fall in the 4.43-5.39 eV range (i.e. between 230-280 nm), where the first experimental band is recorded in both vapor by Clark et al.[60] and Li et al.[61] and in a nitrogen matrix by Polewski et al.[66]

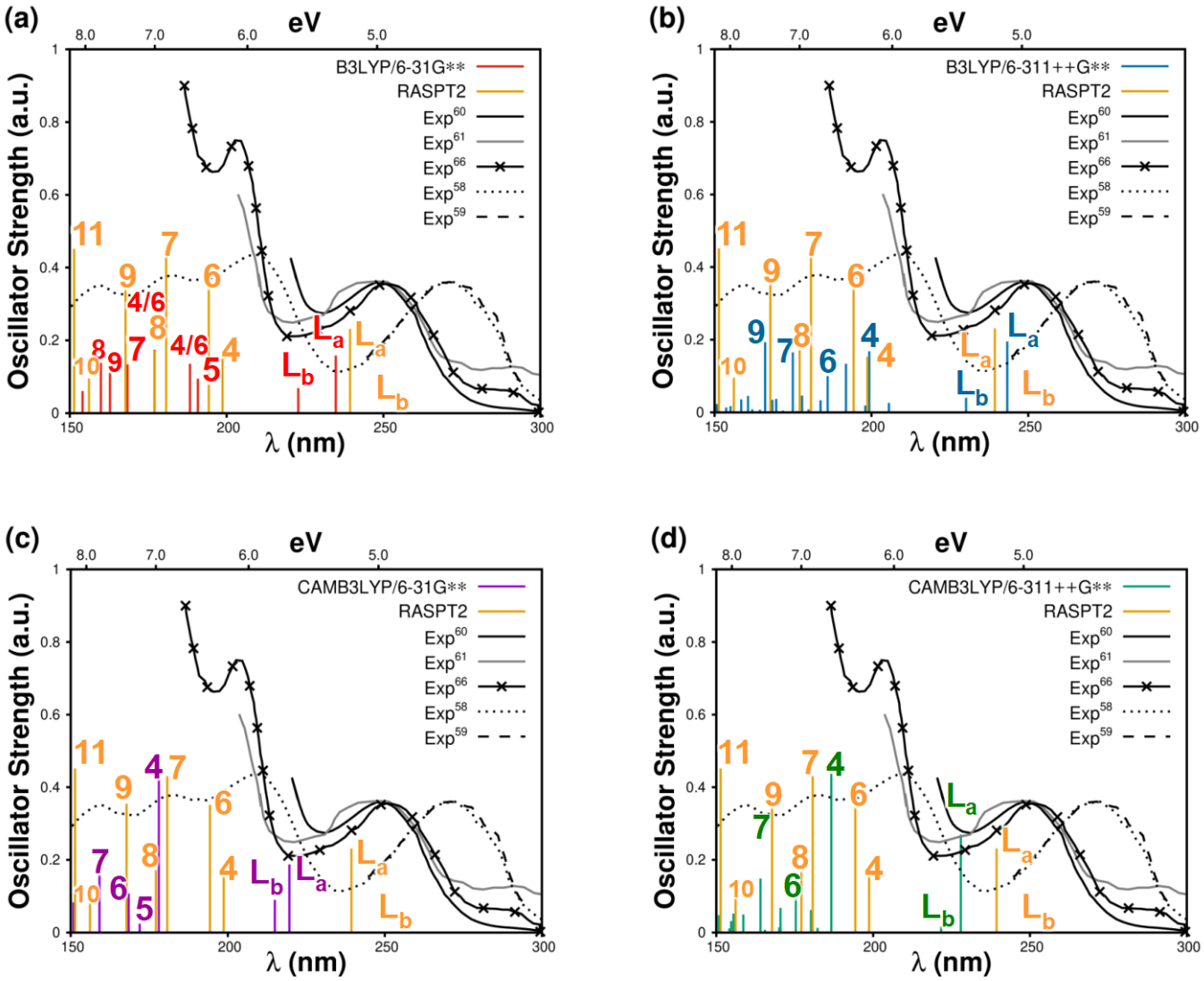

Figure 1. Experimental absorption spectrum in vapor phase (solid black line from Clark et al.[60], solid gray line from Li et al.[61]), matrix-isolated spectra (crossed black line from Polewski et al. [66]) and from sublimed films (dotted black line from Yamada et al[58], dashed black line from Sinsheimer et al.[59]) compared with computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with (a) B3LYP/6-31G**, (b) B3LYP/6-311++G**, (c) CAM-B3LYP/6-31G** and (d) CAM-B3LYP/6-311++G**. Only ππ* states are labeled, according to their root number (i.e. relative energy) in the reference RASPT2 computations (reported in Table S1).

As mentioned above, the experimental absorption bands reported by Sinsheimer et al.[59] and Yamada et al. [58] on adenine sublimed films are comparatively red-shifted to those in the vapor-phase and RASPT2 transition energies. This could be due to some residual intermolecular interactions in the samples, which is supported by the fact that the absorption maxima on sublimed films for all nucleobases are similar to those recorded on thin films by Gomez et al.[64] As shown in Figure 1, all TD-DFT computations display a reversed energetic order for the lowest two excited states (L_b and L_a) with respect to RASPT2[72, 73], while preserving the oscillator strengths of the corresponding vertical excitations. Therefore, both TD-DFT and RASPT2 approaches suggest that the main contribution to the first experimental band (at 4.43-5.39 eV range) comes from the L_a state. The order of L_a and L_b excited states of adenine in gas-phase obtained in our reference RASPT2 calculations agrees with other highly correlated methods like CCSD(T)[74] and CCSD.[72]

At higher energies, the experimental spectra in vapor phase features a second more intense band below 230 nm,[60, 61] with the matrix-isolated spectrum displaying the next absorption band with maxima at ca. 203 nm[66] at the edge of deep-UV region and a rising third band below 196 nm in the vacuum-UV region. The adenine sublimed film[59] spectrum indicates the

presence of several transitions associated to this higher energy band (see Figure 1). In this energy window, RASPT2 computations locate several $S_0 \rightarrow S_n$ transitions associated to states 4 and 6-9 (according to root numbers reported in Table S1) with oscillator strengths of ca. 0.4, in line with experimental data. The vertical excitations associated to states 4 and 6-9 computed with B3LYP/6-311++G** are in qualitative agreement with RASPT2 computations, except for state 8, whose transition energy is significantly blue-shifted. Moreover, the corresponding oscillator strengths of these transitions at TD-DFT are generally underestimated when compared to RASPT2. In this spectral window, TD-DFT computations using the 6-311++G** basis-set also show the presence of several states associated to excitations from lone-pair on nitrogen/oxygen to □-type virtual orbitals (n→σ*), as reported for other similar organic systems.[41] These symmetry-allowed transitions are significantly blue-shifted when adopting the smaller (6-31G**) basis-set and are not computed in our reference RASPT2 calculations, as this would involve very large active spaces. However, these excited states generally feature very low oscillator strengths at TD-DFT and thus are not considered further in this work.

When comparing the two basis-sets employed within TD-DFT, B3LYP/6-31G** computations showed significant mixing between $\pi\ \pi^*$ excitations, especially for states 4 and 6 (see Figure 1a), and an overall blue shift of the excited states with respect to B3LYP/6-311++G** computations. Moreover, state 5, which has a very low oscillator strength at both the reference RASPT2 and the B3LYP/6-311++G** level, gains larger oscillator strength when employing the double-ζ basis-set. By comparing CAM-B3LYP to B3LYP, an overall blue-shift of the high-lying excited state manifold is observed. The oscillator strengths at CAM-B3LYP/6-311++G** are in line with those of B3LYP/6-311++G** computations, except for state 4 that is particularly high. Unsurprisingly, similarly to B3LYP, the use of smaller basis-sets with CAM-B3LYP induces an overall blue-shift of the vertical excitation energies. To summarize, B3LYP/6-311++G** gives the closest agreement with RASPT2 calculations, showing that it can provide reliable transition energies up to ca. 7.5 eV from the GS (i.e. up to state 9) even if the oscillator strengths of higher-lying excited states are generally underestimated with respect to the reference RASPT2.

1.2 Excited state absorptions

Figure 2 reports the RASPT2 computed excited state absorption (ESA) energies for adenine, showing how the $S_n \rightarrow S_m$ spectra depends on the initially populated state discriminating between the first two quasi-degenerate singlet states L_a and L_b. In the near-IR probing window, between 1000-1200 nm, ESAs from L_a feature larger transition dipole moments (TDMs) involving states 4-6 in contrast to L_b-ESAs, which features only one transition lying at ca. 950 nm (involving state 4). In the Vis probing window, between 500-700 nm, weak ESA signals associated to states 7 and 8 are predicted to appear for both the L_a and L_b states, with one, slightly more intense ESA signal associated to state 9 appearing only in the L_b spectrum. Thus, in the near-IR probing window we observed striking differences between the computed L_a and L_b ESAs, suggesting specific spectroscopic fingerprints potentially describing the populations of the two states.

B3LYP/6-311++G** results and convoluted spectra are also reported in Figure 2, for a qualitative comparison with reference RASPT2 computations, showing that TD-DFT provides a similar picture to RASPT2 in the near-IR and Vis probing windows and may thus be used to discriminate between the low-lying spectroscopic fingerprints of adenine L_a and L_b states. When employing the CAM-B3LYP functional a large blue-shift of the excited state manifold is observed (as already showed for adenine linear absorption), which is detrimental for the characterizations of ESAs (see Figure S1/S2). The use of double-ζ basis-set leads to an overall blue-shift of the ESA transition energies (as for the $S_0 \rightarrow S_n$ transitions) and the more extended triple-ζ basis-set generally provides better agreement than smaller basis-set when compared with RASPT2 (see Figure S1/S2).

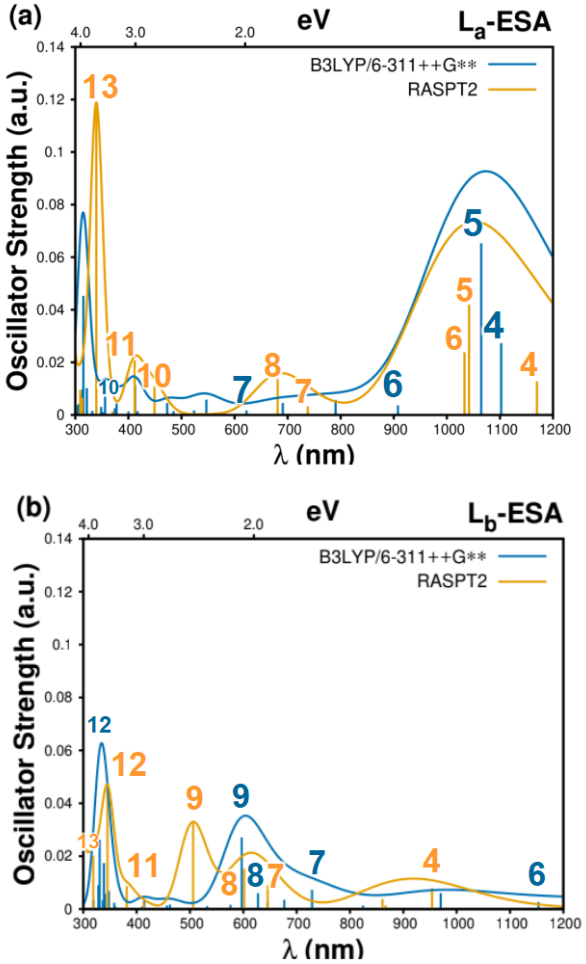

Figure 2. Computed ESAs associated to vertical $S_n \rightarrow S_m$ excitations (colored sticks) and corresponding convoluted spectra for adenine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) computations for (a) $S_n = L_a$ and (b) $S_n = L_b$. Only $\pi\ \pi^*$ states are labeled, according to root numbers in the reference RASPT2 computations (Tables S2-S3).

In the UV probing window, the computed RASPT2 ESAs also suggest differences between the L_a and L_b states. The most intense ESA signal in this region arises from L_a and involves a double excitation (state 13), consistently with previously reported experimental and theoretical pump-probe spectra of solvated adenine.[70, 75] Unfortunately, TD-DFT computations in the adiabatic approximation do not account for double excitations thus missing the most important ESA signal in the UV probing window. The L_b-ESAs are significantly less intense than the L_a-ESAs, and in addition to doubly-excited state 13 also involve single excitations (states 11,12) that are adequately described at the TD-DFT levels. In particular, B3LYP computations feature slightly blue-shifted transition energies and smaller TDMs for single excitations with respect to RASPT2 calculations (see Figure 2). This can be observed for state 10 in the L_a-ESAs or for state 12 in the L_b-ESAs. Coherently with what experienced for the other spectroscopic windows, CAM-B3LYP blue-shifts the UV ESAs of both L_a and L_b states with respect to B3LYP, thus moving them outside the spectral window considered here (see Figure S2). In the UV probing window, TD-DFT computations employing larger (i.e. more diffuse) basis-set show the appearance of some high-lying π *-Rydberg mixed states, which are not considered in RASPT2 calculations (see Computational details). It is worth mentioning that Rydberg-like states are significantly blue-shifted in solution with respect to gas-phase and thus are expected to have a negligible role when extending this type of computations to realistic models.
In summary, we observed that ESAs computed at the B3LYP/6-311++G** level provide convoluted spectra that are in qualitative agreement with RASPT2 calculations for excited states lying below 7.5 eV (as for linear absorption), and therefore for spectral windows probing the ESAs of L_a and L_b states in both near-IR and Vis regions. At higher-energies, in the UV probing window, TD-DFT computations are generally facing the limitation of missing double excitations, particularly important for the L_a-ESAs. For the sake of brevity, we will compare only the B3LYP/6-311++G** level of theory with the RASPT2 calculations for the remainder of this work, documenting the results of the other TD-DFT levels employed here in the Supporting Information.

2.Guanine

2.1 Ground state absorptions

Clark et al. recorded absorption maxima of guanine in vapor phase[60] at 293 and 284 nm, without reporting the corresponding absorption spectrum while noting evidence of significant decomposition. However, matrix-isolated spectra at low temperature (15K)[65] provide the UV absorption spectra in gas phase up to ca. 160 nm. The absorption spectra of sublimed films[58-59] are red-shifted compared to matrix-isolated spectra similar to adenine while displaying a similar trend across UV absorption window. Figure 3(a) reports the vertical $S_0 \rightarrow S_n$ excitations of guanine in the gas-phase, computed at RASPT2 and B3LYP/6-311++G** levels of theory together with the experimental spectra of guanine in nitrogen-matrix[65] and sublimed films[58-59]. In contrast to adenine, the lowest-lying $\pi\pi^*$ excited states (labeled L_a and L_b for adenine) located in the 4.5-5.3 eV range (ca. 230-280 nm) are found in the same energetic order by TD-DFT and RASPT2 computations, with L_a featuring a lower oscillator strength than L_b.

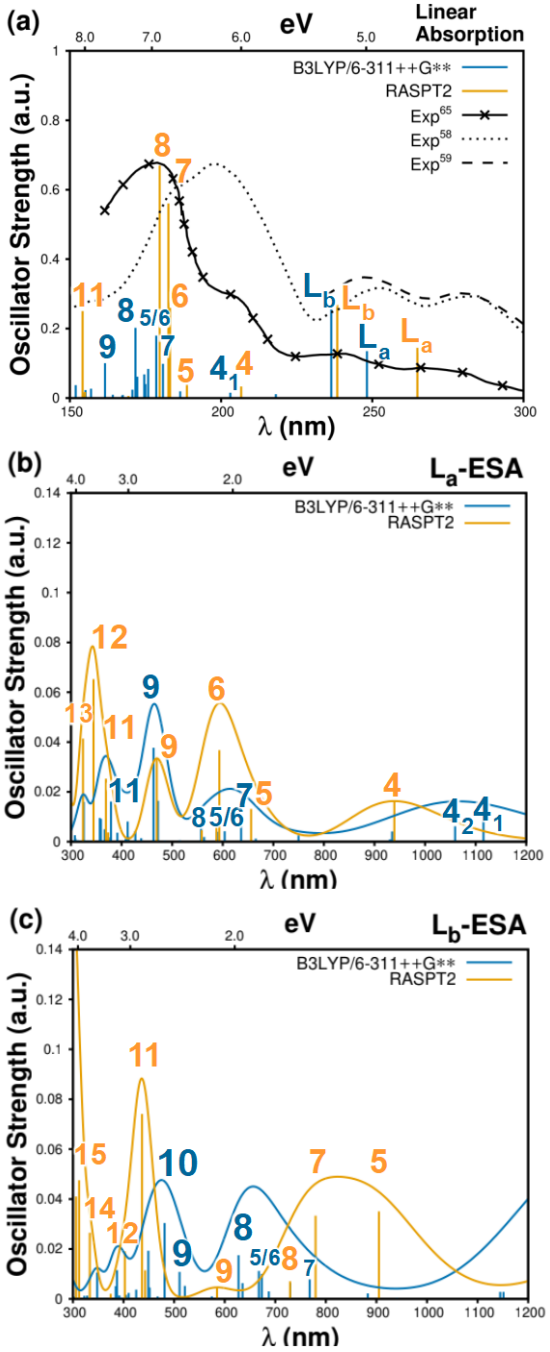

Figure 3. (a) Experimental absorption spectra of guanine in nitrogen-matrix (crossed black line from Polewski et al.[65]), from sublimed films (dotted black line from Yamada et al[58], dashed black line from Sinsheimer et al.[59]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of guanine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) for (a) $S_0 \rightarrow S_m$, (b) $L_a \rightarrow S_m$ and (c) $L_b \rightarrow S_m$ transitions and corresponding convoluted spectra (colored lines). Only $\pi\pi^*$ states are labeled, according to RASPT2 reference (Table S4).

The transition energies at RASPT2 for these excitations (264 nm and 238 nm for L_a and L_b respectively) are in line with the two absorption maxima in matrix-isolated spectra occurring at 277 nm and 241 nm reported by Polewski et al. allowing the assignments of these bands. At B3LYP/6-311++G**, the transition energy of L_b is in good agreement with RASPT2, while L_a displays a more pronounced blue-shift. Both excitations are significantly blue-shifted at other TD-DFT levels employed here (see Figure S3). It is worth noting that the L_a/L_b Platt's notation of benzene-like systems used here for the lowest-lying ππ* states of adenine[51] and guanine[50] and is employed for consistency with previous literature, but those states actually involve different electronic excitations in these two nucleobases. This remark also explains why the discrepancy in energetic order of L_b /L_a between RASPT2 and TD-DFT, as mentioned above, is not observed in guanine.

At a higher energy window, Polewski et al. report another absorption band at 200 nm[65]. In this region RASPT2 computations predict an excitation to state 4 at 206 nm with low oscillator strength of ca. 0.03. At B3LYP/6-311++G** level, mixing of this ππ* state with n-σ* states result in two states of similar character in the same region (labelled 4_1 and 4_2), with the one having more oscillator strength located at 203 nm, in line with RASPT2. In the vacuum-UV region, the experimental absorption spectra in nitrogen-matrix features the most intense band at 176 nm[65] with spectra on sublimed film[58] displaying a similar trend with red-shifted transition energies. RASPT2 computations in the vacuum-UV spectral region (i.e. below 190 nm) also predict multiple $S_0 \rightarrow S_n$ states in a narrow spectral window between 180-190 nm. The primary contributions arise from states 7 and 8 along with states 5 and 6 having lower oscillator strength (Figure 3a). While these states are also found by all TD-DFT computations (Figure S3), their transition energies are blue-shifted, and their oscillator strengths are generally lower than the reference. At the B3LYP/6-311++G** level, the $\pi \pi^*$ states 5 and 6 feature a significant mixing (see Figure 3a) that is not found with CAM-B3LYP functional or with small basis-set. In addition, as remarked earlier in the section for adenine, the use of bigger triple-ζ basis-set at TD-DFT also litters the higher energy spectral window with excitations involving mixed π*-Rydberg-type orbitals having appreciable oscillator strength.

2.2 Excited state absorptions

Figure 3(b, c) reports RASPT2 computed vertical $S_n \rightarrow S_m$ transition energies for guanine and the corresponding convoluted spectra, according to the initially populated S_n state, i.e. either L_a or L_b. RASPT2 computations indicate that in the near-IR probing window the excited state absorption from L_a involves state 4 while two ESAs (involving states 5 and 7) of higher oscillator strength are found for L_b (Figure 3). In contrast, two major transitions from L_a to states 6 and 9 are found in the Vis probing window (at 593 and 469 nm, respectively) while the L_b-ESAs are dominated by an intense transition to state 11 at 436 nm. In the near-UV probing region (i.e. below 380 nm), the ESAs computed at RASPT2 from both L_a and L_b states do not show any appreciable differences, featuring several intense $S_n \rightarrow S_m$ transitions, including contributions from double excitations (e.g. state 13).

Comparing B3LYP with RASPT2 computations we found red-shifted L_a-ESAs in the near-IR probing window associated to state 4 and blue-shifted L_b-ESAs involving states 5 and 7 (see Fig. 3 and Fig. S4/S5). As mentioned above for adenine, n→σ* states get involved in the ESAs signals when the 6-311++G** basis-set is employed, which leads to the splitting of state 4 into two almost degenerate states (i.e. 4_1 and 4_2) in the case of guanine. The oscillator strengths of both L_a- and L_b-ESAs are generally underestimated at B3LYP with respect to CASPT2. In the near-UV probing region, while reference RASPT2 computations of L_a-ESAs feature multiple $\pi \pi^*$ transitions with high TDMs (states 11,12,13), all TD-DFT methods predict only the transition involving state 11 (Figure 3b/Figure S4). In the case of L_b-ESAs, the manifold of bright states predicted by RASPT2 are not found at the TD-DFT levels in this probing window (Figure 3c/Figure S5), as they are most likely shifted to higher energies or they have some contribution from double excitations. As remarked earlier for adenine, employing CAM-B3LYP functional blue-shifts all ESAs of both L_a and L_b states with respect to B3LYP (Figure S5).

3. Uracil

3.1 Ground state absorptions

Uracil is the only canonical pyrimidine for which gas-phase spectrum in vapor phase[60, 62-63] is available in addition to sublimed film.[58-59] These experimental spectra are shown in Figure 4(a) along with computed transitions at RASPT2 and B3LYP/6-311++G** levels. In the deep-UV region RASPT2 computations predict a single transition to state 2 at ca. 238 nm in line with the first absorption band recorded in vapor spectra.[60, 63] Similar to purines, the gas-phase experimental spectra on sublimed films are red-shifted with respect to vapor-phase spectra and RASPT2 transition energies. Moving higher up in energy, the vapor phase spectra feature a rising second band in the vacuum-UV (below 230 nm) spectral window, also predicted by RASPT2 computations and associated to states 3, 4 and 5. The TDMs of the RASPT2 transitions increase with the excitation energies, in line with the intensities trend of the first two experimental bands below 250 nm observed in sublimed films.[58, 62] All $S_0 \rightarrow S_n$ RASPT2 transitions of uracil are well reproduced at B3LYP/6-311++G** level (see Fig. 4a). In analogy to purines, the use of the 6-311++G** basis-set identifies electronic transitions of weak intensity to states of mixed nσ*/ππ* nature at energies above 6.5 eV, i.e. around state 5 in the case of uracil. The employment of smaller basis-set induces a slight blue-shift of all transitions, while a more pronounced shift is observed when employing the CAM-B3LYP functional (see Figure S6).

3.2 Excited state absorptions

In contrast to purines, the first two singlet excited states of uracil (and other pyrimidines in the gas phase) are well separated in energy and involve different electronic excitations (see Tables S1, S4, S7, S9 and S11). Thus, we will consider here only the $S_1 \rightarrow S_m$ excitations as potential ESAs detectable in non-linear spectroscopies set-up with standard pump UV-B (at 267nm).

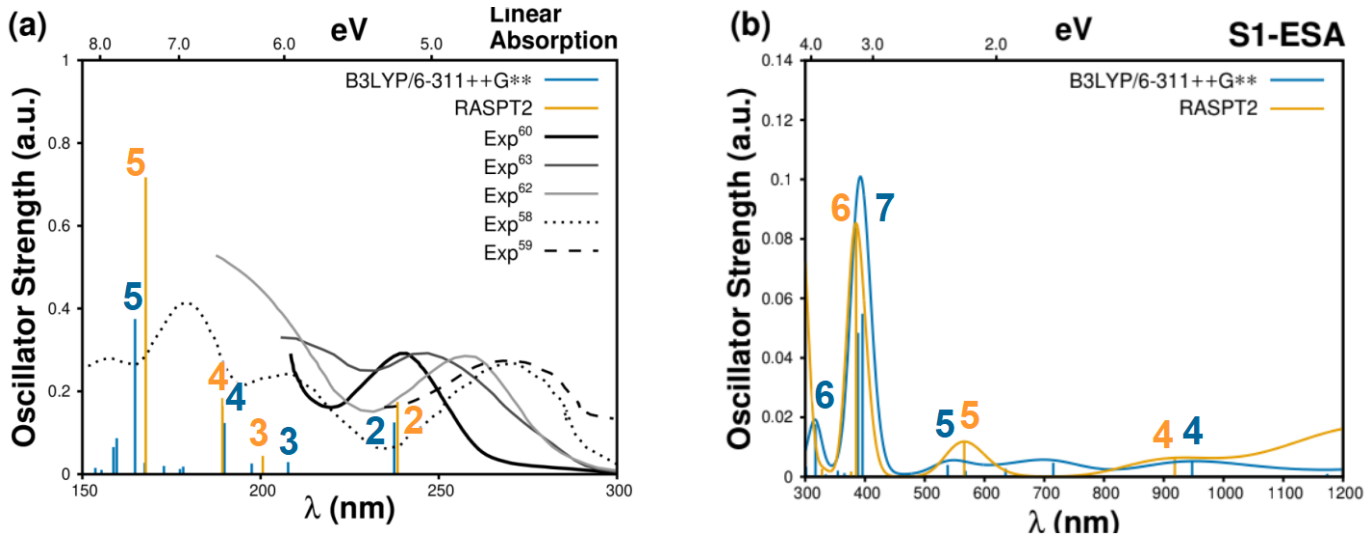

Figure 4. (a) Experimental absorption spectrum of uracil in vapor phase (solid black line from Clark et al.[60], solid dark-gray line from Nowak et al.[62] and solid light-gray line from Ito et al.[63]), from sublimed films (dotted black line from Yamada et al.,[58] dashed black line from Sinsheimer et al.[59]), and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of uracil in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) for (a) $S_0 \rightarrow S_m$, (b) $S_1 \rightarrow S_m$ transitions and corresponding convoluted spectra (colored lines). Only π π* states are labeled, according to RASPT2 reference (see Tables S7-S8).

In Figure 4(b), ESAs computed at RASPT2 and the corresponding convoluted spectra are compared to those at the B3LYP/6-311++G** level. The vertical $S_1 \rightarrow S_m$ excitations computed at RASPT2 indicate the presence of three S_m states, one for each probing window, as shown in Figure 4(b). In the near-IR and Vis probing windows, the excitations from S_1 associated to states 4 and 5, respectively, feature almost negligible TDMs at the RASPT2 level and both their transition energies and TDMs are properly reproduced at the B3LYP/6-311++G** level. In the UV probing window (at ca. 400 nm), on the other hand, a ESA associated to state 6 with non-negligible TDM is predicted by RASPT2, while TD-DFT computations (with both B3LYP and CAM-B3LYP functionals) indicate the presence of two ESAs, one at the same energy as RASPT2 but associated to state 7 and one blue-shifted, associated to state 6. Such swapping of the excited states ordering and the discrepancy with respect to RASPT2 TDMs is due to the multiconfigurational nature of these states (being two linear combinations of the H-2→L and H-3→L excitations)[49] and to the non-negligible double excitations contributions present in state 6.[49] Thus, while in the absence of specific assignments of the $S_1 \rightarrow S_m$ excitations the ESA signals predicted by TD-DFT might appear matching RASPT2 computations, TD-DFT results can be considered in full agreement with the multiconfigurational method only for ESAs lying in the near-IR and Vis probing windows. Moreover, it is worth noting that at the B3LYP/6-311++G** level, a $S_1 \rightarrow S_m$ excitation into a Rydberg orbital is found overlapping with the bright ESA at around 400nm analogous to what was observed for purines in this probing window, which is not observed when employing smaller basis-sets. Finally, CAM-B3LYP transition energies are blue-shifted with respect to RASPT2, independently on the basis set employed (see Figure S7).

4. Thymine

4.1 Ground state absorptions

The experimental absorption spectra of thymine in gas-phase have been reported only for sublimed films,[58-59] and using electron impact methods.[68] The spectra on sublimed films feature absorption maxima that are red-shifted with respect to computed RASPT2 and B3LYP/6-311++G** vertical transition energies (see Figure 5a), in line with what we observed for uracil and purines. Also for thymine nucleobase, computations and experiments feature several similarities in terms of both relative positions and intensities of the absorption bands. In particular, RASPT2 computation predicts the first transition to state 2 in deep-UV region at ca. 247 nm (5.0 eV) in line with the first transition at 4.95(±0.8) eV in EEEL spectra.[68] At higher energies, two transitions involving states 3 and 4 are found around 190 nm at RASPT2 at 6.20 eV and 6.51 eV, while a band at 6.20 eV (±0.8) has been reported in EEEL spectra. Both these transitions are obtained at B3LYP/6-311++G** with transition energies and oscillator strengths in agreement with RASPT2 computations. In the vacuum-UV window, RASPT2 computations predict a transition to state 5 with high TDM at ca. 169 nm (7.32 eV), which is also observed in EEEL spectra at 7.4 eV, and another transition to state 6 at ca. 154 nm, with smaller TDM. These states are significantly blue-shifted at B3LYP/6-311++G** level compared to RASPT2. Moreover, the oscillator strength of state 5 is underestimated at all TD-DFT methods (Figure S8). The effect on linear absorption of basis-set and functionals are similar to those described above for uracil, with blue-shifts in energies when employing the smaller 6-31G** basis-set and the CAM-B3LYP functional (Figure S8).

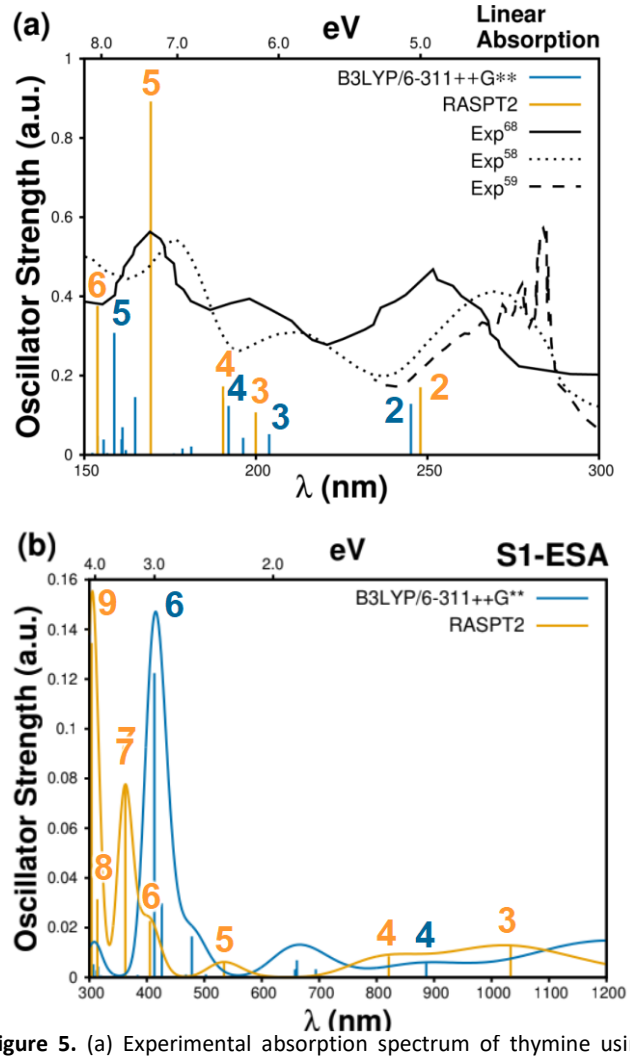

Figure 5. (a) Experimental absorption spectrum of thymine using electron impact energy loss EEEL (solid black line from Abouaf el al.[68]), from sublimed films (dotted black line from Yamada et al.[58], dashed black line from Sinsheimer et al.[59]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of thymine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) for (a) $S_0 \rightarrow S_m$, (b) $S_1 \rightarrow S_m$ transitions and the corresponding convoluted spectra (colored lines). Only $\pi\ \pi^*$ states are labeled, according to RASPT2 reference (Table S9-S10).

4.2 Excited state absorptions

As for the uracil case, we limited the analysis of ESAs of thymine to the $S_1 \rightarrow S_m$ transitions. RASPT2 computations predict two transitions in the near-IR probing window, involving states 3 and 4 (Figure 5b). These transitions are properly reproduced by TD-DFT methods with similar oscillator strengths and slightly red- and blue-shifted transition energies at B3LYP and CAM-B3LYP, respectively (Figure 5b and S9). The ESA signal for state 3 shifts from 1033 nm (i.e. 1.20 eV) at RASPT2 to 1209 nm (i.e. 1.02 eV) at B3LYP/6-311++G**, lying just outside the spectral window plotted in Figure 5b. At higher energies, in the Vis to near-UV probing region until 300 nm, RASPT2 computations feature a large number of ESA transitions from S_1 involving states 5, 6, 7 and 9, with an increasing trend in oscillator strengths. In contrast, TD-DFT computations only feature the transition to state 6, with a markedly increased oscillator strength compared to RASPT2. The pronounced disagreement between TD-DFT and RASPT2 in this region, which is reflected in the qualitative comparison of the convoluted spectra (Figure 5b), is due to the substantial double-excited contributions to excited states (mainly states 7 and 9) at RASPT2, which are not adequately described at TD-DFT levels. We thus conclude that the TD-DFT methods benchmarked here are only reliable in the near-IR to Vis windows when probing the S_1-S_n ESAs.

5. Cytosine

5.1 Ground state absorptions

Figure 6a reports the vertical electronic transitions ($S_0 \rightarrow S_n$) of cytosine in the gas-phase, computed at the RASPT2 and B3LYP/6-311++G** levels of theory along with experimental spectra in argon-matrix,[67] from sublimed films[58-59] and EEEL spectra.[69] The comparison of RASPT2 computations and experiments on cytosine is analogous to that of the other nucleobases described above. In the deep-UV region (above 200 nm) two states involving transitions to states 2 (4.66 eV) and 3 (5.59 eV) are found at RASPT2 with relatively small oscillator strengths (ca. 0.1). These data are in line with the absorption maxima at ca. 267 nm (4.64 eV) observed in all available experiments and with the rising shoulder appearing at ca. 220 nm (5.63 eV) in matrix-isolated spectrum along with those reported in EEEL spectra at 4.65 eV and 5.50 eV (±0.1 eV).

Moving higher up in energy towards the vacuum-UV, three transitions (to states 4 (6.46 eV), 5 (6.90 eV) and 6 (8.01 eV)) are found at the RASPT2 level, with the transition associated to state 4 (Figure 6a) featuring the largest oscillator strength, which can be associated to the most intense experimental absorption band found at around 200 nm in sublimed films and matrix-isolated spectra. Analogous trend of transitions has been reported in the EEL spectra, featuring bands at 6.2 eV, 6.7 eV and 8.0 eV (±0.1 eV). All $S_0 \rightarrow S_n$ RASPT2 transitions of cytosine are quite well reproduced at the B3LYP/6-311++G** level. However, the use of the larger 6-311++G** basis-set in this energy window also locates weak transitions to states of mixed nσ*/ππ* nature after state 4. The employment of smaller basis-set induces a slight blue-shift of all transitions, while a more pronounced shift is observed when employing the CAM-B3LYP functional (see Figure S10).

5.2 Excited state absorptions

As for other pyrimidines, only $S_1 \rightarrow S_m$ transitions have been considered for the ESAs of cytosine. In Figure 6(b), ESAs of cytosine computed at RASPT2 are compared to those at the B3LYP/6-311++G** level along with the corresponding convoluted spectra. In the Vis probing window, two main transitions involving states 4 and 5 (at 687 nm and 553 nm, respectively) are found at the RASPT2, with B3LYP/6-311++G** computations showing a good agreement for the transition energies while the oscillator strength for state 5 is significantly underestimated. As remarked in the previous section, transitions of mixed nσ*/ππ* nature above state 4 are found when employing the bigger 6-311++G** basis-set and the $S_1 \rightarrow S_m$ transitions to some of these states are also found in this probing window at TD-DFT with relatively small oscillator strength.

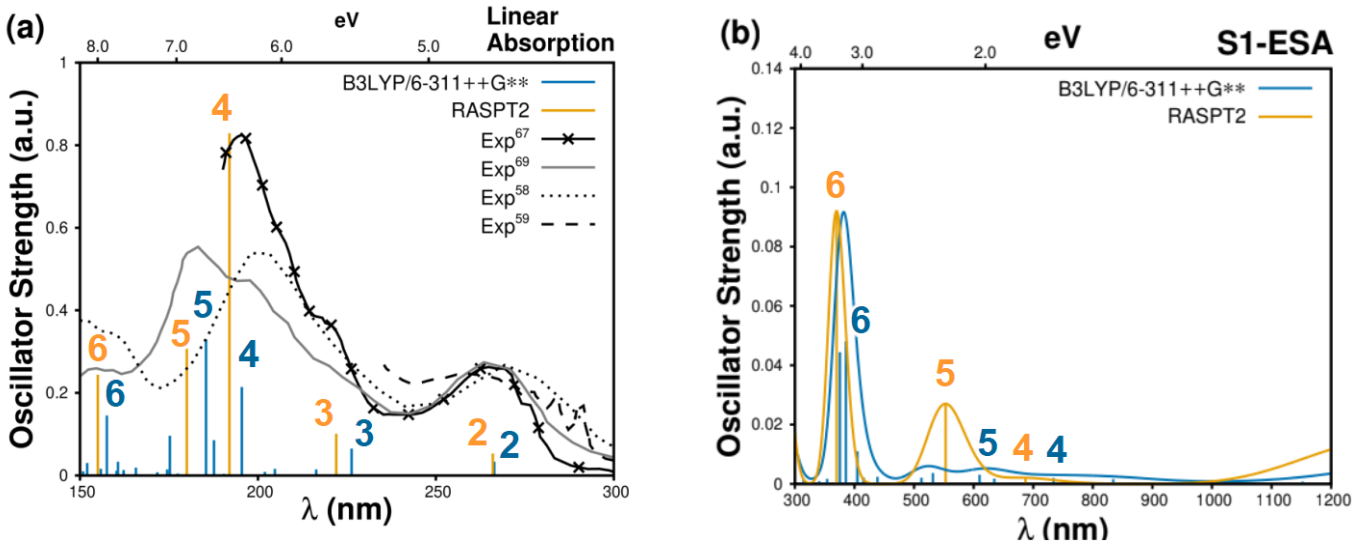

Figure 6. (a) Experimental absorption spectrum of cytosine in argon-matrix (crossed black line from Bazsó et al.[67]), from EEEL spectra (gray black line from Abouaf et al.[69]), from sublimed films (dotted black line from Yamada et al.[58], dashed black line from Sinsheimer et al.[59]) and computed vertical $S_0 \rightarrow S_n$ excitations (colored sticks) of cytosine in gas-phase. Reference theoretical values at RASPT2 (yellow sticks) are compared with B3LYP/6-311++G** (blue sticks) for (a) $S_0 \rightarrow S_m$, (b) $S_1 \rightarrow S_m$ transitions and the corresponding convoluted spectra (colored lines). Only $\pi\pi^*$ states are labeled, according to RASPT2 reference (Table S11-S12).

At higher energies, in the near-UV probing window, a comparatively more intense $S_1 \rightarrow S_m$ transition at 370 nm to state 6 is found at RASPT2 level, which is properly reproduced at the B3LYP/6-311++G** level. However, at this TD-DFT level, a state involving excitations to Rydberg-like orbitals is found to be almost degenerate with the $\pi\pi^*$ state 6, in analogy to what observed for uracil and purines in the UV probing window. As for the other cases, this Rydberg-like excitation is not observed when employing smaller basis-sets (see Figure S11). CAM-B3LYP functional provides results similar to B3LYP ones, showing blue-shifts similar to other pyrimidines (see Figure S11).

Computational details

All computations have been carried out in the gas phase to isolate the specific differences among the diverse theoretical methods considered, neglecting the influence of the solvent to the vertical excitation energies and oscillator strengths (*f*). It is well-known that solvation can strongly modulate the energetic state ordering of the DNA/RNA nucleobases and even affect their excited state lifetimes.[6] However, in the present study we focus on comparing our linear absorption estimates with experimental evidence mainly recorded in gas-phase, allowing us to benchmark the different theoretical approaches under the same conditions.

The ground state (GS or S_0) geometry of the different nucleobases were optimized at the Complete Active Space Self-Consistent Field (CASSCF) level, employing 8 electrons in 8 orbitals (4 π bonding and 4 π^* anti-bonding) for the purines and 6 electrons in 6 orbitals (3 π bonding and 3 π^* anti-bonding) for pyrimidines. On top of those equilibrium geometries, vertical transitions from the S_0 state ($S_0 \rightarrow S_n$, with n>0) and from the S_n excited states ($S_n \rightarrow S_m$, with n=1-2 and m>n) have been computed at different levels of theory, making use of the C_s symmetry to limit our calculations to the more important and initially accessed $\pi\pi^*$ transitions, only. In this work, the excited states are labeled in adiabatic notation according to RASPT2 root numbers (reordered in terms of relative energy, e.g. S_1 excited state is label as "2", S_2 as "3", and so on), except for the first two excited states of purines for which the commonly used Platt's notation[76] (L_b and L_a) has been preferred, as reported in Tables S1-S12. Such vertical transitions are representative of the excitations accessible in the Franck-Condon (FC) region as we have shown elsewhere,[41, 77] and may be compared to linear ($S_0 \rightarrow S_n$) and time-resolved non-linear ($S_n \rightarrow S_m$) spectra, the latter referring to pump-probe data collected at zero (or very short) delay times. It is worth noting that more correlated methods could in principle be employed to obtain the equilibrium geometries, their specific role having been assessed previously for adenine[51] and being out of the scope of the present manuscript that focuses on intrinsic differences due to the electronic structure theory employed for the excited states. For a qualitative comparison between TD-DFT and RASPT2 computations of ESAs, electronic transitions have been convoluted using Gaussian functions with full width half maximum of 0.3 eV.

Multiconfigurational wavefunction methods

Restricted active space self-consistent field (RASSCF) and its second order perturbation theory extension (RASPT2) was used throughout as implemented in the MOLCAS 8 package.[78] Reference RASPT2 computations were obtained by adding an additional set of diffuse and uncontracted basis functions to the center of charge of the molecule and subsequently removing systematically π^* Rydberg-like orbitals. This procedure was initially employed by Roos and co-workers and allows minimizing the overstabilization reported in multiconfigurational perturbation theory due to the presence of

Rydberg-like orbitals in the secondary space.[79] These orbitals are not properly represented within the employed Atomic Natural Orbital (ANO)-type triple-ζ basis-set used, preventing also the appearance of quasi-Rydberg states that, however, will be strongly blue-shifted in realistic models including solvation effects. It is also important mentioning that attempts to include Rydberg orbitals in the RAS spaces in order to properly account for Rydberg states would be computationally unfeasible, especially considering the large state average schemes employed here. We have previously carried out an extensive analysis on the high-lying excited state manifold of DNA/RNA pyrimidine (uracil, thymine and cytosine)[49] and purine (adenine[51] and guanine[50]) derivatives following the aforementioned protocol. This has led to set as reference calculations the RASPT2 computations based on a RAS(0,0|10,8|2,12) active space for pyrimidines and RAS(0,0|12,10|2,12) for adenine, while RAS(0,0|14,11|2,8) was the highest level affordable for guanine. These active spaces include all π and π* valence orbitals in a complete active space (i.e. with full configuration interaction) plus additional π* orbitals (12 for all bases except guanine that features 8) included with up to 2 electron permutations allowed in order to further increase the dynamic correlation towards convergence. In the multiconfigurational state-average treatment we have included 30 singlet excited states in order to account for enough states to monitor all ESA signals of interest. These computations provide the best available theoretical values of excitation energies and transition dipole moments for the first 30 excited states of DNA/RNA canonical nucleobases and represent reference computations that hereafter are referred to simply as RASPT2.

TD-DFT calculations

Time Dependent-Density Functional Theory (TD-DFT) calculations were performed in order to test their reliability compared to multiconfigurational perturbation theory methods and available experimental data. Linear and quadratic response calculations are required to simulate both linear and non-linear spectra and were performed with the Dalton 2016 code,[80] using the Cs point group as molecular symmetry for all studied structures. The lowest lying 50 singlet states, were considered and the excitation energies were calculated using the Becke three-parameter hybrid functional B3LYP[52, 53] and the long-range corrected functional CAM-B3LYP.[54] Concerning the basis-sets, both Pople's polarized double-ζ 6-31G**[55, 56, 81, 82] and augmented triple-ζ 6-311++G**[55-57] were used, in order to check the effect of the inclusion of polarization and diffuse functions on the excited states, providing information on the role of Rydberg states and Rydberg-valence coupling in the transition energies and dipole moments.

The excited state nature of the vertical transitions was characterized analyzing the electronic density reorganization through Natural Transition Orbitals (NTOs),[83] obtained using the Nancy_EX code.[84, 85] TD-DFT state assignment was done to match that of the RASPT2 reference[49-51] by choosing those transitions whose character (orbital transition and relative weight) are more similar.

Since the aim is a qualitative comparison between computations and experiments we have compared our simulated data with the experimental absorption spectra, extracted using WebPlotDigitizer.[86] Note that we have normalized all experimental spectra to the intensity of the highest energy absorption band in case of guanine and the lowest absorption bands for all other nucleobases.

Conclusions

In the present work we have thoroughly assessed the performance of TD-DFT methodologies in the investigation of the singlet excited state manifold of canonical nucleobases by comparing it with high-quality reference wavefunction-based RASPT2 linear and non-linear (excited-state) absorptions. B3LYP functional with a triple-ζ quality basis-set provided the closest agreement with the reference RASPT2 computations and the available experimental evidence, outperforming CAM-B3LYP in computing vertical excitation energies of TD-DFT systems *in vacuo*. There are some common trends exhibited by TD-DFT computations in all nucleobases: i) the 6-31G** basis-set produces blue-shifted signals compared to the larger (more diffuse) 6-311++G** basis-set; ii) the long-range corrected CAM-B3LYP functional also leads to blue-shifted signals compared to B3LYP hybrid functional; iii) in addition to $\pi \rightarrow \pi$* excitations, TD-DFT methods also locate symmetry-allowed n→σ* excitations from lone-pair orbitals localized on Nitrogen/Oxygen atoms to virtual orbitals, which appear in the vacuum-UV spectral window and have usually low oscillator strengths (but gain intensity when they mix significantly with $\pi \rightarrow \pi$* states, as it occurs with the 6-311++G** basis-set); iv) $\pi \rightarrow \pi$*(Rydberg) excitations appear in the vacuum-UV spectral window due to the stabilization of Rydberg-type orbitals when using large triple-ζ basis-set.

The outcome suggests that for pyrimidines, the employed TD-DFT methods are able to locate all the important signals in the UV linear absorption spectra up to the vacuum-UV spectral window. Such performance of TD-DFT methods is also observed in the excited-state absorptions (arising from the first singlet excited state) in the near IR/Visible spectral probing window. For purines, on the other hand, the higher complexity of the excited state manifold leads to a good agreement for ground state absorptions only in the near-UV energy window. The substantial contributions by doubly excited contributions in the vacuum-UV energy window cannot be captured by TD-DFT, which leads to comparatively worse agreement with the RASPT2 computations. Consequently, ESA signals from the L_a state in purines can be reproduced faithfully by TD-DFT only in the near-IR region and worsens towards the Vis probing window.

Overall, our results validate the use of TD-DFT as a cost-effective approach for the computations of excitation energies and oscillator strengths that are associated to time-resolved non-linear optical spectra of DNA/RNA-based systems but only for low-energy probing windows in the near-IR and for pyrimidines also in the Vis. This paves the way for the concerted theoretical and experimental study of larger DNA sequences of greater biological significance where these benchmarked low-energy windows might be used to monitor

the different photo-processes triggered in DNA complex systems upon UV-light absorption.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank the Agence Nationale de la Recherche project FEMTO-2DNA (ANR-15-CE-29-0010). I.R. acknowledges the use of HPC resources of the "Pôle Scientifique de Modélisation Numérique" (PSMN) at the École Normale Supérieure de Lyon, France. Part of the calculations have been performed on the LPCT computing system. J.S.-M. acknowledges support from the European Commission through the Marie Curie actions (AttoDNA, FP8-MSCA-IF, grant nº 747662), and the Generalitat Valenciana GenT Fellowship programme (CDEIGENT/2019/021).

Notes and references

1. C. E. Crespo-Hernandez, B. Cohen, P. M. Hare and B. Kohler, *Chem. Rev.*, 2004, **104**, 1977-2019.
2. W. J. Schreier, P. Gilch and W. Zinth, *Annu. Rev. Phys. Chem.*, 2015, **66**, 497-519.
3. C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217-239.
4. A. Giussani, J. Segarra-Martí, D. Roca-Sanjuán and M. Merchán, in *Photoinduced Phenomena in Nucleic Acids I: Nucleobases in the Gas Phase and in Solvents*, eds. M. Barbatti, A. C. Borin and S. Ullrich, Springer International Publishing, Cham, 2015, DOI: 10.1007/128_2013_501, pp. 57-97.
5. R. Improta and V. Barone, in *Top. Curr. Chem.*, Springer Berlin Heidelberg, 2014, DOI: 10.1007/128_2013_524, ch. 524, pp. 1-29.
6. R. Improta, F. Santoro and L. Blancafort, *Chem. Rev.*, 2016, **116**, 3540-3593.
7. A. A. Beckstead, Y. Zhang, M. S. de Vries and B. Kohler, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24228-24238.
8. J. Chen, Y. Zhang and B. Kohler, in *Photoinduced Phenomena in Nucleic Acids II*, eds. M. Barbatti, A. C. Borin and S. Ullrich, Springer International Publishing, 2015, vol. 356, ch. 570, pp. 39-87.
9. D. Markovitsi, T. Gustavsson and I. Vayá, *J. Phys. Chem. Lett.*, 2010, **1**, 3271-3276.
10. T. Gustavsson, R. Improta and D. Markovitsi, *J. Phys. Chem. Lett.*, 2010, **1**, 2025-2030.
11. I. Vaya and T. Gustavsson, 2010, 11834-11835.
12. A. Francés-Monerris, H. Gattuso, D. Roca-Sanjuán, I. Tuñón, M. Marazzi, E. Dumont and A. Monari, *Chemical Science*, 2018, **9**, 7902-7911.
13. V. J. Borràs, A. Francés-Monerris and D. Roca-Sanjuán, *ChemPhotoChem*, 2019, **3**, 889-896.
14. K. E. Szkaradek, P. Stadlbauer, J. Šponer, R. W. Góra and R. Szabla, *Chem. Commun.*, 2020, **56**, 201-204.
15. D. B. Bucher, B. M. Pilles, T. Carell and W. Zinth, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 4369-4374.
16. D. B. Bucher, C. L. Kufner, A. Schlueter, T. Carell and W. Zinth, *J. Am. Chem. Soc.*, 2016, **138**, 186-190.
17. D. B. Bucher, A. Schlueter, T. Carell and W. Zinth, *Angew. Chem. Int. Ed.*, 2014, **53**, 11366-11369.
18. F. Buchner, A. Nakayama, S. Yamazaki, H.-H. Ritze and A. Lübcke, *J. Am. Chem. Soc.*, 2015, **137**, 2931-2938.
19. V. I. Prokhorenko, A. Picchiotti, M. Pola, A. G. Dijkstra and R. J. D. Miller, *J. Phys. Chem. Lett.*, 2016, **7**, 4445-4450.
20. W.-M. Kwok, C. Ma and D. L. Phillips, *J. Phys. Chem. B*, 2009, **113**, 11527-11534.
21. C. Ma, C. C.-W. Cheng, C. T.-L. Chan, R. C.-T. Chan and W.-M. Kwok, *Phys. Chem. Chem. Phys.*, 2015, **17**, 19045-19057.
22. G. W. Doorley, M. Wojdyla, G. W. Watson, M. Towrie, A. W. Parker, J. M. Kelly and S. J. Quinn, *J. Phys. Chem. Lett.*, 2013, **4**, 2739.
23. I. Vayá, T. Gustavsson, T. Douki, Y. Berlin and D. Markovitsi, *J. Am. Chem. Soc.*, 2012, **134**, 11366-11368.
24. B. A. West, J. M. Womick and A. M. Moran, *J. Phys. Chem. A*, 2011, **115**, 8630-8637.
25. B. K. McFarland, J. P. Farrell, S. Miyabe, F. Tarantelli, A. Aguilar, N. Berrah, C. Bostedt, J. D. Bozek, P. H. Bucksbaum, J. C. Castagna, R. N. Coffee, J. P. Cryan, L. Fang, R. Feifel, K. J. Gaffney, J. M. Glownia, T. J. Martinez, M. Mucke, B. Murphy, A. Natan, T. Osipov, V. S. Petrović, S. Schorb, T. Schultz, L. S. Spector, M. Swiggers, I. Tenney, S. Wang, J. L. White, W. White and M. Gühr, *Nat Commun*, 2014, **5**.
26. M. Merchán, R. González-Luque, T. Climent, L. Serrano-Andrés, E. Rodriguez, M. Reguero and D. Pelaez, *J. Phys. Chem. B*, 2006, **110**, 26471-26476.
27. L. Serrano-Andres, M. Merchan and A. C. Borin, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 8691-8696.
28. S. Matsika, *J. Phys. Chem. A*, 2004, **108**, 7584-7590.
29. S. Matsika, *J. Phys. Chem. A*, 2005, **109**, 7538-7545.
30. A. Sobolewski and W. Domcke, *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics*, 2002, **20**, 369-374.
31. A. L. Sobolewski and W. Domcke, *Europhysics News*, 2006, **37**, 20-23.
32. F. Plasser, R. Crespo-Otero, M. Pederzoli, J. Pittner, H. Lischka and M. Barbatti, *J. Chem. Theory Comput.*, 2014, **10**, 1395-1405.
33. J. Segarra-Martí, A. Francés-Monerris, D. Roca-Sanjuán and M. Merchán, *Molecules*, 2016, **21**, 1666.
34. J. Segarra-Martí, M. Garavelli and F. Aquilante, *J. Chem. Theory Comput.*, 2015, **11**, 3772-3784.
35. M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Nachtigallova, P. Hobza and H. Lischka, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 21453-21458.
36. D. Nachtigallová, T. Zelený, M. Ruckenbauer, T. Müller, M. Barbatti, P. Hobza and H. Lischka, *J. Am. Chem. Soc.*, 2010, **132**, 8261-8263.
37. S. Mai, M. Richter, P. Marquetand and L. González, in *Top. Curr. Chem.*, Springer Berlin Heidelberg, 2014, DOI: 10.1007/128_2014_549, ch. 549, pp. 1-55.
38. A. J. Pepino, J. Segarra-Martí, A. Nenov, R. Improta and M. Garavelli, *J. Phys. Chem. Lett.*, 2017, DOI: 10.1021/acs.jpclett.7b00316, 1777-1783.
39. L. Martínez-Fernández, A. J. Pepino, J. Segarra-Martí, J. Jovaišaitė, I. Vaya, A. Nenov, D. Markovitsi, T. Gustavsson, A. Banyasz, M. Garavelli and R. Improta, *J. Am. Chem. Soc.*, 2017, **139**, 7780-7791.

40. A. J. Pepino, J. Segarra-Martí, A. Nenov, I. Rivalta, R. Improta and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2018, **20**, 6877-6890.
41. J. Segarra-Martí, E. Zvereva, M. Marazzi, J. Brazard, E. Dumont, X. Assfeld, S. Haacke, M. Garavelli, A. Monari and J. Léonard, *J. Chem. Theory Comput.*, 2018, **14**, 2570-2585.
42. O. Weingart, A. Nenov, P. Altoè, I. Rivalta, J. Segarra-Martí, I. Dokukina and M. Garavelli, *J. Mol. Model.*, 2018, **24**, 271.
43. M. K. Shukla and J. Leszczynski, *J. Comput. Chem.*, 2004, **25**, 768-778.
44. D. Varsano, R. Di Felice, M. A. Marques and A. Rubio, *J. Phys. Chem. B*, 2006, **110**, 7129-7138.
45. A. Tsolakidis and E. Kaxiras, *J. Phys. Chem. A*, 2005, **109**, 2373-2380.
46. R. Improta and V. Barone, *J. Am. Chem. Soc.*, 2004, **126**, 14320-14321.
47. F. Santoro, V. Barone, A. Lami and R. Improta, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4934-4948.
48. C. M. Isborn, N. Luehr, I. S. Ufimtsev and T. J. Martínez, *J. Chem. Theory Comput.*, 2011, **7**, 1814-1823.
49. A. Giussani, J. Segarra-Martí, A. Nenov, I. Rivalta, A. Tolomelli, S. Mukamel and M. Garavelli, *Theor. Chem. Acc.*, 2016, **135**, 1-18.
50. J. Segarra-Martí, A. J. Pepino, A. Nenov, S. Mukamel, M. Garavelli and I. Rivalta, *Theor. Chem. Acc.*, 2018, **137**, 47.
51. A. Nenov, A. Giussani, J. Segarra-Martí, V. K. Jaiswal, I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli, *J. Chem. Phys.*, 2015, **142**, 212443.
52. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652.
53. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
54. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51-57.
55. W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257-2261.
56. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213-222.
57. T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294-301.
58. T. Yamada and H. Fukutome, *Biopolymers*, 1968, **6**, 43-54.
59. R. L. Sinsheimer, J. F. Scott, J. R. Loofbourow, R. H. With the technical assistance of B. Best, E. Vollmer, and M. Westergaard, *J. Biol. Chem.*, 1950, **187**, 313-324.
60. L. B. Clark, G. G. Peschel and I. Tinoco, *The Journal of Physical Chemistry*, 1965, **69**, 3615-3618.
61. L. Li and D. M. Lubman, *Anal. Chem.*, 1987, **59**, 2538-2541.
62. M. J. Nowak, K. Szczepaniak, A. Barski and D. Shugar, *Zeitschrift für Naturforschung C*, 1978, **33**, 876-883.
63. Y. Tsuchiya, T. Tamura, M. Fujii and M. Ito, *The Journal of Physical Chemistry*, 1988, **92**, 1760-1765.
64. E. F. Gomez, V. Venkatraman, J. G. Grote and A. J. Steckl, *Adv. Mater.*, 2015, **27**, 7552-7562.
65. K. Polewski, D. Zinger, J. Trunk, D. C. Monteleone and J. C. Sutherland, *J. Photochem. Photobiol. B: Biol.*, 1994, **24**, 169-177.
66. K. Polewski, D. Zinger, J. Trunk and J. C. Sutherland, *Radiat. Phys. Chem.*, 2011, **80**, 1092-1098.
67. G. Bazsó, G. Tarczay, G. Fogarasi and P. G. Szalay, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6799-6807.
68. R. Abouaf, J. Pommier and H. Dunet, *Chem. Phys. Lett.*, 2003, **381**, 486-494.
69. R. Abouaf, J. Pommier, H. Dunet, P. Quan, P.-C. Nam and M. T. Nguyen, *J. Chem. Phys.*, 2004, **121**, 11668-11674.
70. J. Segarra-Martí, F. Segatta, T. A. Mackenzie, A. Nenov, I. Rivalta, M. Bearpark and M. Garavelli, *Faraday Discuss.*, 2019.
71. A. Nenov, A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel and M. Garavelli, *Phys. Chem. Chem. Phys.*, 2015, **17**, 30925-30936.
72. S. K. Khani, R. Faber, F. Santoro, C. Hättig and S. Coriani, *J. Chem. Theory Comput.*, 2019, **15**, 1242-1254.
73. F. Santoro, R. Improta, T. Fahleson, J. Kauczor, P. Norman and S. Coriani, *J. Phys. Chem. Lett.*, 2014, **5**, 1806-1811.
74. Z. Benda and P. G. Szalay, *J. Phys. Chem. A*, 2014, **118**, 6197-6207.
75. R. Borrego-Varillas, G. Cerullo and D. Markovitsi, *J. Phys. Chem. Lett.*, 2019, **10**, 1639-1643.
76. J. R. Platt, *J. Chem. Phys.*, 1949, **17**, 484-495.
77. J. Segarra-Martí, S. Mukamel, M. Garavelli, A. Nenov and I. Rivalta, in *Multidimensional Time-Resolved Spectroscopy*, Springer, 2019, pp. 63-112.
78. F. Aquilante, J. Autschbach, R. Carlson, L. Chibotaru, M. G. Delcey, L. De Vico, I. Fernández Galvan, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. A. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. Vysotskiy, O. Weingart, F. Zapata and R. Lindh, *J. Comput. Chem.*, 2016, **37**, 506-541.
79. J. Lorentzon, P.-Å. Malmqvist, M. Fülscher and B. Roos, *Theor. Chim. Acta*, 1995, **91**, 91-108.
80. K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 269-284.
81. J. D. Dill and J. A. Pople, *J. Chem. Phys.*, 1975, **62**, 2921-2923.
82. M. Francl, *J. Chem. Phys*, 1982, **77**, 3654.
83. R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775-4777.
84. T. Etienne, X. Assfeld and A. Monari, *J. Chem. Theory Comput.*, 2014, **10**, 3906-3914.
85. T. Etienne, X. Assfeld and A. Monari, *J. Chem. Theory Comput.*, 2014, **10**, 3896-3905.
86. A. Rohatgi, WebPlotDigitizer, (https://automeris.io/WebPlotDigitizer).