

Probing the molecular frame of uracil and thymine with high-harmonic generation spectroscopy[†]

Eleonora Luppi,^a and Emanuele Coccia^{b,*}Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

In this work we present computed high-harmonic generation (HHG) spectra of uracil and thymine molecules, by means of the real-time time-dependent formulation of Gaussian-based configuration interaction with single excitations (RT-TD-CIS). According to the experimental work [Hutchinson *et al.*, *Phys. Chem. Chem. Phys. Comparison of high-order harmonic generation in uracil and thymine ablation plumes*, 2013, 15, 12308] a pulse wavelength of 780 nm has been used, together with an intensity of 10^{14} W/cm² and a pulse duration of 23 optical cycles. In order to examine the effect of pulse polarisation, rotationally-averaged (to mimic the gas-phase sample) and single-polarisation have been computed for both molecules. Our results show that the HHG signal for both molecules possibly originates from different ionisation channels, involving HOMO, HOMO-1, HOMO-2 and HOMO-3 orbitals, which lie within 4 eV. We characterize the HHG spectrum of thymine, supporting the idea that the absence of thymine signal in the original work does not depend on the single-molecule behaviour. Present results for uracil are consistent with the experimental data. Moreover, we have observed that states below and above the chosen ionisation threshold provide different contributions to the HHG spectrum in averaged and single-polarisation calculations.

1 Introduction

A growing interest in attosecond atomic and molecular processes is nowadays triggered by the advent of ultrafast laser technology^{1–8}. Novel time-resolved spectroscopies are emerging together with the opportunity to study electron dynamics with unprecedented time resolution⁹. Attosecond light pulses are indeed used to investigate ultrafast electron dynamics in atomic and molecular systems.⁵

The research in the attosecond domain can answer challenging questions such as: which is the role of electron correlation in nonlinear optical process? Which is the electron interference in a molecular system ionised from different channels? Which is the ultrafast redistribution of electrons in molecular charge migration that could be then responsible for bond breaking? As a consequence, attosecond resolution became the fundamental requirement to design original experimental analysis to understand a large variety of physical and chemical processes.

The attosecond laser resolution relies on the nonlinear optical processes, as high-harmonic generation (HHG).^{10–13} The HHG spectra, made of the emitted harmonics of the infrared pulse frequency, encode information about the electronic structure and the dynamical effects of the systems from which the HHG signal is emitted.^{14–30} For instance, from HHG spectra one can extract

information on the electronic wave function for molecular orbital tomography.^{15,31,32} HHG can also detect interference phenomena between ionisation channels in molecules^{20,21,28,33,34}, ultrafast charge migration³⁵, vibrational signatures³⁶, and quantum coherence in atoms³⁷.

HHG spectroscopy is therefore considered a powerful tool to investigate the ultrafast processes occurring in atoms, molecules and nanostructures³⁸, and recently it has also been applied to study organic and biological molecules³⁹.

However, HHG experiments on organic and biological molecules can present technical difficulties. In fact, the main practical issue is due to the fact that molecules at room temperature can be in solid or liquid phase, and producing a high-density gas-phase sample of unfragmented molecules is not trivial.³⁹ Moreover, these molecules have typically low ionisation energies and show a large variety of different conformers.

Uracil and thymine have been widely characterised spectroscopically, from experimental and theoretical works, in terms of valence^{40–44} and core excitations^{45,46}. But only one experimental attempt to measure HHG in uracil and thymine is present in literature, by Hutchinson *et al.*⁴⁷ Authors used a laser ablation technique to produce weakly ionized plasma plumes from solid samples of uracil and thymine. Only the uracil HHG signal has been observed. The difference between the HHG spectra of uracil and thymine was attributed to differences in the ion composition of the ablation plumes of the two molecules, determining conditions of phase mismatching in thymine.

That work certainly opened the route to other experimental investigations of HHG in these biomolecules, as many questions remains still open: why the HHG spectrum of thymine was not measured? Which is the role of fragments in the measured HHG

^a Sorbonne Universités, UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005 Paris, France. CNRS, UMR 7616, Laboratoire de Chimie Théorique, F-75005 Paris, France

^b Dipartimento di Scienze Chimiche e Farmaceutiche, via Giorgieri 1, 34127 Trieste, Italy; E-mail: ecoccia@units.it

[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/cXCP00000x/

spectra? Do electron correlation and interference effects play any role in the strong-field dynamics?

Theoretical calculations can definitely help to answer these questions. Simulations on these molecules are numerically challenging because of their size and complexity. To calculate HHG spectra in uracil and thymine we used the all-electron real-time dependent configuration interaction with single excitations (RT-TD-CIS),^{37,48–59} using the computational strategy we developed in the recent years.^{37,53,55,57,58} Standard Dunning’s basis sets⁶⁰ were augmented with diffuse and continuum-optimal⁶¹ Gaussian functions. This determines a proper description of Rydberg and low-energy continuum states. This approach represents a reasonable compromise between accuracy and computational cost, which permits an affordable treatment of many-electron systems interacting with strong fields⁵⁸. Moreover, this approach allows one to control and systematically improve Gaussian basis sets.

We computed HHG spectra for different linear polarisations of the laser field in the molecular plane of uracil and thymine, in the three-dimensional (3D) space around the molecules. In order to simulate randomly oriented molecules we averaged the time-dependent dipoles projected on the different polarisations in the molecular plane or in the 3D space. We also computed the HHG spectra with a pulse polarisation perpendicular to the molecular plane or along the direction of the ground-state permanent dipole. Moreover, we analysed the contribution to the HHG spectrum due to the states below and above the selected ionisation energy. Results presented here allow us to give a quantum, microscopic description of HHG in uracil and thymine and pose open questions inherent the strong field-electron dynamics in these biomolecules.

The article is organised as follows: in Section 2 the theoretical framework for computing HHG spectra is reviewed, computational details of our simulations are given in Section 3, results are presented and discussed in Section 4, while in Conclusions we summarize the main achievements of the present work and give perspectives for future developments.

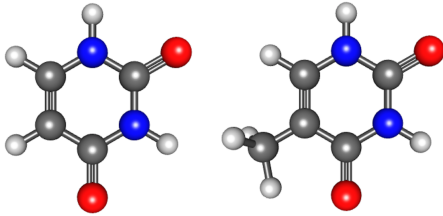


Fig. 1 Uracil (left) and thymine (right) molecules.

2 Theory

The time-dependent Schrödinger equation for a molecular system under the influence of an external time-dependent electric field is (atomic units are used) :

$$i \frac{\partial |\Psi(t)\rangle}{\partial t} = (\hat{H}_0 + \hat{V}(t)) |\Psi(t)\rangle, \quad (1)$$

where \hat{H}_0 is the time-independent field-free electronic Hamiltonian and $\hat{V}(t)$ is time-dependent potential that we calculated in

the length gauge. In this gauge $\hat{V}(t) = -\hat{\boldsymbol{\mu}} \cdot \mathbf{E}(t)$ where $\hat{\boldsymbol{\mu}}$ is the molecular dipole and $\mathbf{E}(t)$ is the time-dependent electric field.⁵¹ The electric field is linearly polarised $\mathbf{E}(t) = E_0 \hat{\mathbf{n}} \sin(\omega_0 t + \phi) f(t)$ along the direction $\hat{\mathbf{n}}$, E_0 is the maximum field strength, ω_0 is the carrier pulse frequency, ϕ is the field phase, and $f(t)$ is a \cos^2 envelope function.⁵⁵

We solved Eq. (1) in the framework of time-dependent configuration interaction with simple excitations (RT-TD-CIS).^{48,49,54,55,62–65} The time-dependent wave function $|\Psi(t)\rangle$ is expanded in a discrete basis of the eigenstates of the field-free Hamiltonian \hat{H}_0 composed of the Hartree-Fock ground state ($k = 0$) and all the CIS excited states⁶⁶ ($k > 0$)

$$|\Psi(t)\rangle = \sum_{k \geq 0} c_k(t) |\Psi_k\rangle, \quad (2)$$

where $c_k(t)$ are time-dependent coefficients.

By inserting Eq. (2) into Eq. (1) and projecting on $\langle \Psi_l |$, the time-dependent equation for the coefficients is:

$$i \frac{\partial \mathbf{c}(t)}{\partial t} = (\mathbf{H}_0 + \mathbf{V}(t)) \mathbf{c}(t), \quad (3)$$

where $\mathbf{c}(t)$ is the vector of the coefficients $c_k(t)$, \mathbf{H}_0 is the diagonal matrix representation of \hat{H}_0 , with elements $\mathbf{H}_{0,lk} = \langle \Psi_l | \hat{H}_0 | \Psi_k \rangle = E_k \delta_{lk}$ (E_k is the eigenenergy of the k -th state), and $\mathbf{V}(t)$ is the non-diagonal matrix representation of $\hat{V}(t)$, with elements $\mathbf{V}_{lk}(t) = \langle \Psi_l | \hat{V}(t) | \Psi_k \rangle = -\mathbf{E}(t) \langle \Psi_l | \hat{\boldsymbol{\mu}} | \Psi_k \rangle$. The initial wave function at $t = t_i = 0$ is the field-free ground state, i.e. $c_k(t_i) = \delta_{k0}$.

The Eq. (3) is solved by discretizing the time (Δt is the time step) and using the split-propagator scheme⁵⁵:

$$\mathbf{c}(t + \Delta t) \approx e^{-i\mathbf{V}(t)\Delta t} e^{-i\mathbf{H}_0\Delta t} \mathbf{c}(t). \quad (4)$$

The matrix \mathbf{H}_0 is diagonal, as a consequence $e^{-i\mathbf{H}_0\Delta t}$ is also a diagonal matrix of elements $e^{-iE_k\Delta t} \delta_{lk}$. The exponential of the non-diagonal matrix $\mathbf{V}(t)$ is calculated by the following transformation

$$e^{-i\mathbf{V}(t)\Delta t} = \mathbf{U}^\dagger e^{-i\mathbf{V}_d(t)\Delta t} \mathbf{U}, \quad (5)$$

where \mathbf{U} is the unitary matrix describing the change of basis between the original eigenstates of \hat{H}_0 and a basis in which $\hat{V}(t)$ is diagonal, i.e. $\mathbf{V}_d(t) = \mathbf{U}\mathbf{V}(t)\mathbf{U}^\dagger$.^{51,55}

From the knowledge of the time-dependent wavefunction $|\Psi(t)\rangle$, the time-dependent dipole $\boldsymbol{\mu}(t)$ is then calculated as

$$\boldsymbol{\mu}(t) = \sum_{lk} c_l^*(t) c_k(t) \langle \Psi_l | \hat{\boldsymbol{\mu}} | \Psi_k \rangle, \quad (6)$$

and by taking its Fourier transform the HHG spectrum is obtained as

$$P(\omega) = \left| \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \boldsymbol{\mu}(t) \cdot \hat{\mathbf{n}} e^{-i\omega t} dt \right|^2, \quad (7)$$

with t_f is the final propagation time. In Figure 2 we schematically show how we simulated an HHG for a laser linearly polarised in the direction perpendicular to the molecular plane (panel a)) and along the direction of the permanent molecular dipole (panel b)). Next, we considered the molecules randomly oriented in the plane and in the 3D space. The randomly oriented molecules were simulated averaging over many different polarisation direc-

tions of the laser. In the case of the plane we used a normalised polarisation vector $\hat{\mathbf{n}}_s = (\hat{n}_{s,x}, \hat{n}_{s,y}, 0)$ while in case of the three-dimensional space we used $\hat{\mathbf{n}}_s = (\hat{n}_{s,x}, \hat{n}_{s,y}, \hat{n}_{s,z})$, where $\hat{n}_{s,x}$ and $\hat{n}_{s,y}$ (and also $\hat{n}_{s,z}$ for the 3D case) are uniform random numbers in [-1:1] interval. Then, we calculated an averaged time-dependent dipole $\bar{\boldsymbol{\mu}}(t)$ defined as

$$\bar{\boldsymbol{\mu}}(t) = \frac{1}{S} \sum_s \boldsymbol{\mu}_s(t) \cdot \hat{\mathbf{n}}_s, \quad (8)$$

where S is the number of simulations made with a randomly chosen electric field polarisation $\hat{\mathbf{n}}_s$ and $\boldsymbol{\mu}_s(t)$ is the corresponding time-dependent dipole. The averaged HHG spectrum is then calculated as

$$\bar{P}(\omega) = \left| \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \bar{\boldsymbol{\mu}}(t) e^{-i\omega t} dt \right|^2. \quad (9)$$

In Figure 2 we schematically show how we simulated an HHG rotationally averaged in the plane (panel c)) and in 3D (panel d)). Rationalizing the generation mechanism of high-order harmonics in thymine and uracil implies the inclusion of several factors. Tunneling ionisation and recombination can involve different channels for the electron removal (as discussed below) and centers of recombinations. By defining a threshold energy (E_T), which corresponds to the ionisation energy of a particular channel, we are able to disentangle the role played by electronic states above and below E_T , and, as a consequence, to more deeply investigate their possible contribution to HHG and mutual interactions.⁵² The time-dependent dipole moments, $\boldsymbol{\mu}(t)$ (Eq. 6) or $\bar{\boldsymbol{\mu}}(t)$ (Eq. 8), are hence rewritten as a (partial) sum of terms depending only on the states below E_T , labeled as B, or states above E_T , labeled as A. The ground state is indicated as G . Following this decomposition, the time dependent dipole moments become:

$$\boldsymbol{\mu}_B(t) = \boldsymbol{\mu}_{GG}(t) + \sum_{i \in B} \boldsymbol{\mu}_{Gi}(t) + \sum_{ij \in B} \boldsymbol{\mu}_{ij}(t) \quad (10)$$

$$\bar{\boldsymbol{\mu}}_B(t) = \frac{1}{S} \sum_s \boldsymbol{\mu}_{B,s}(t) \cdot \hat{\mathbf{n}}_s \quad (11)$$

and

$$\boldsymbol{\mu}_A(t) = \boldsymbol{\mu}_{GG}(t) + \sum_{i \in A} \boldsymbol{\mu}_{Gi}(t) + \sum_{ij \in A} \boldsymbol{\mu}_{ij}(t) \quad (12)$$

$$\bar{\boldsymbol{\mu}}_A(t) = \frac{1}{S} \sum_s \boldsymbol{\mu}_{A,s}(t) \cdot \hat{\mathbf{n}}_s. \quad (13)$$

The partial contributions to time-dependent dipole moments can therefore be identified as: $\boldsymbol{\mu}_{GG}(t) = |c_0(t)|^2 \langle \Psi_0 | \hat{\boldsymbol{\mu}} | \Psi_0 \rangle$ is the ground-ground contribution, $\boldsymbol{\mu}_{Gi}(t) = c_i^*(t) c_0(t) \langle \Psi_i | \hat{\boldsymbol{\mu}} | \Psi_0 \rangle$ is the ground-excited contribution with the index i running over the states below E_T (B) or above E_T (A), and $\boldsymbol{\mu}_{ij}(t) = c_j^*(t) c_i(t) \langle \Psi_j | \hat{\boldsymbol{\mu}} | \Psi_i \rangle$ is the excited-excited contribution with the index i and j running over the states below E_T (B) or above E_T (A). $\boldsymbol{\mu}_{B,s}(t)$ and $\boldsymbol{\mu}_{A,s}(t)$ are the s -th realization of the partial $\boldsymbol{\mu}_B(t)$ and $\boldsymbol{\mu}_A(t)$ dipoles. The HHG spectrum from each partial contribution is then computed as in Eqs 7 or 9.

States below and above E_T are also named in this work as bound

and continuum states, respectively.

3 Computational details

The geometrical structures of uracil and thymine are shown in Figure 1. These structures have been optimised in the ground-state at DFT/B3LYP level using the 6-31G(d,p) basis set with the software Q-Chem.⁶⁷

We performed CIS calculations⁶⁷ to obtain a number of electronic excited-state energies and transition dipole moments of the field-free Hamiltonian \hat{H}_0 , which are then used to propagate the time-dependent wavepacket in RT-TD-CIS, as shown in Eq. 3.

The RT-TD-CIS simulations were performed by means of the homemade code Light⁵¹⁻⁵⁵, that propagates the electronic wavepacket under the influence of a time-dependent strong laser field. We used the same pulse wavelength of the experiment of Ref. 47, i. e. 780 nm (equal to $\omega_0 = 1.59$ eV). The pulse intensity is defined as $I = \frac{1}{2} |E_0|^2$. We used a pulse intensity $I = 10^{14}$ W/cm² and pulse duration of 23 optical cycles. The phase ϕ was set to zero. Nuclei were kept frozen at their equilibrium position during the time propagation.

For RT-TD-CIS we used the cc-pVDZ Dunning's basis set⁶⁰ for the H atoms of uracil and thymine. Instead, for all the other atoms (C, O and N) we augmented the cc-pVDZ basis set in order to increase the number of bound and continuum states. We added to the cc-pVDZ kernel three sets of diffuse functions and three sets of optimised Gaussian functions for continuum (K) for each angular momentum. We thus obtained the 3aug-cc-pVDZ+3K basis set. The K functions⁶¹ determine a progression of continuum states, i.e. above the ionisation energy E_T , which mimics the true manifold of molecular continuum states.⁵⁴⁻⁵⁶

Ionisation in the simulated strong-field electron dynamics was treated by means of the heuristic lifetime model, originally reported in Ref 50 and modified by us in Refs. 55,58. In this work, we used as electron escape lengths $d_0 = \frac{E_0}{\omega_0^2} = 15.67 a_0$ for low energy continuum and $d_1 = 0.1 a_0$ for high energy continuum. Within such double- d version of the heuristic lifetime model the value of d_0 corresponds to the semi-classical maximum electron excursion predicted by the three-step model^{58,68,69}. For more details see discussion in Ref. 58.

The experimental first, i.e. lowest, ionisation energy of uracil is in the 9.45-9.68 eV range⁷⁰⁻⁷⁴, while for thymine is in the range 9.02-9.20 eV⁷¹⁻⁷³. Our theoretical estimates, computed at Hartree-Fock level of theory with the 3aug-cc-pVDZ+3K basis set, are reported in Table 1. Our calculations overestimate by around 0.5 eV the lowest ionisation energy, corresponding to minus HOMO energy according to Koopmans' theorem. The order of magnitude of the discrepancy between experimental and theoretical ionisation energy values is the same found for CO₂, with a similar theoretical approach²⁸.

In order to also include in the propagation contributions from higher ionisation channels (namely, a larger amount of energy is needed to remove the electron from inner orbitals) we considered that electrons can also be extracted from HOMO-1, HOMO-2 and HOMO-3. This choice corresponds to use minus HOMO-3 energy as E_T . With this choice the 3aug-cc-pVDZ+3K basis set shows a reasonable balance between energy states below and above E_T .

Table 1 Molecular-orbital energies (in eV) for uracil and thymine computed at the Hartree-Fock/3aug-cc-pVDZ+3K level of theory.

	HOMO	HOMO-1	HOMO-2	HOMO-3
uracil	-10.04	-11.89	-12.25	-14.29
thymine	-9.58	-11.81	-12.19	-13.03

For uracil we have 355 states below E_T and 146 states above E_T . For thymine we have 399 states below E_T and 102 states above E_T .

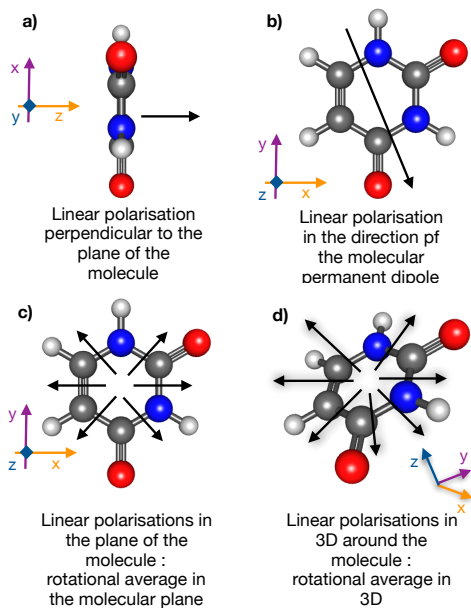


Fig. 2 Sketch representation of the different types of simulations presented in this work. Uracil was chosen as model for showing the computational strategy.

4 Results and Discussion

In the experimental work of Hutchison *et al.*⁴⁷, harmonics were generated from the uracil plasma plume using driving laser pulses at wavelengths of 780 and 1300 nm. The largest harmonic was found to be H33 and H39 in different conditions (Figure 3 and 4 in Ref. 47) with 780 nm. On the other hand, no thymine signal was experimentally recorded.

The analysis of the ablation plumes with mass spectrometry shows that the plumes contain a large number of fragments of the parent molecular ion, with various shape and electronic structure. A higher degree of atomization in the thymine plume was found for thymine, with a corresponding higher density of plasma free electrons. Authors argue that abundance of free electrons in the thymine plasma can induce a strong phase mismatch of the HHG signal, making the recording of it not possible. Furthermore, the observed uracil HHG signal is probably the result of a mixed contribution from the entire molecule and its fragments.

Our goal is the characterization of the HHG spectrum of the two molecules by an ab initio all-electron dynamics, as explained in Section 2. It is worth comparing the information encoded in the experimental spectra of Ref. 47 with what is computed by means of our theoretical approach. The outcomes of our simulations can

be compared with the experimental results, but it is necessary to discuss the advantages, drawbacks and limits of the theoretical comparison we propose.

The experimental sample is a plume containing the entire molecules and a number of atomic and molecular fragments, as reported above, while we compute the HHG spectrum of unfragmented uracil and thymine molecules: analysis of the fragmentation pattern⁷⁵⁻⁷⁷ will allow us to also simulate, as a next step in future works, the HHG signal from fragments, and to possibly dissect fragment contribution in the final HHG spectrum.

For all the simulations we have employed an intensity of 10^{14} W/cm², which is a typical value for getting HHG. In Ref. 47 authors report that the infrared probe pulse acquires an intensity up to 10^{15} W/cm² in the experiments: even when the energy of probe pulse is reported, extracting the precise intensity value corresponding to the recorded HHG spectrum is not easy, making a quantitative comparison with our spectra impracticable. Moreover, in our calculations phase matching conditions can not be simulated, but signatures of electronic structure and dynamics encoded in the HHG spectrum are mainly due to single-molecule behaviour.³⁹

With these (current) limitations in mind, we are however able to provide a detailed description of the main features of HHG spectra of uracil and thymine. The present work should be therefore seen as a first and original proposal for the application of a full quantum-dynamics approach to the study of complex molecules of biological interest.

4.1 Uracil

In Figure 3 HHG spectra of uracil have been reported. The calculations were carried out following the simulation schemes presented in Figure 2 and described in Section 2.

The two averaged (plane and 3D) HHG spectra show the same general shape for the low-energy and plateau regions and a substantial difference around the cutoff. Harmonics up to around H39-H41 are seen for the 3D rotational average, whereas the largest harmonic for the molecular-plan average is H29-H31. The three-step model^{68,69} estimation of the cutoff is approximately H22. This comparison suggests that contributions to HHG from out-of-plane polarisations are important for higher harmonics. Moreover, the HHG spectra averaged in the molecular plane show higher intensity harmonics in the H13-H31 region. The background of the molecular-plane spectrum is much higher than that of the 3D-averaged spectrum beyond H17.

The HHG spectrum with polarisation perpendicular to the molecular plane strongly differs from the averaged spectra starting from H9. From H9 to H11 the harmonics are much higher than in the two averaged HHG spectra. Moreover, the spectrum presents a second plateau from H21 to H35, but the intensity of these harmonics is much lower than those of the first part of the spectrum, up to H19. For the second plateau, the cutoff is around H45.

The HHG spectrum with polarisation in the direction of the permanent dipole has similar behaviour as the perpendicular-polarisation HHG spectrum. Low-intensity harmonics in the high energy region of the spectrum are still present. In this case

even harmonics are clearly visible, due to the inversion-symmetry breaking. For perpendicular polarisation inversion symmetry is given at much less extent, resulting in nearly negligible even peaks.

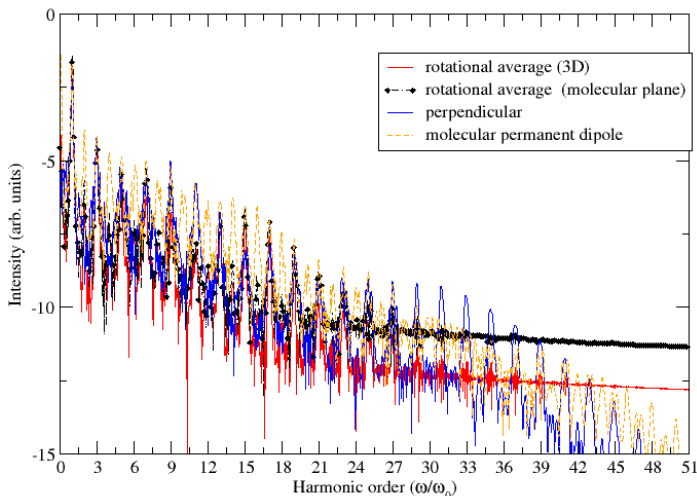


Fig. 3 HHG spectra of uracil with pulse wavelength and intensity of 780 nm and 10^{14} W/cm².

The difference between averaged and perpendicular-polarisation spectra can be rationalised by examining the possible ionisation channels and their symmetry. For this reason, a selection of Hartree-Fock uracil molecular orbitals is reported in Figure 4. In our calculations we included possible ionisation from HOMO, HOMO-1, HOMO-2 and HOMO-3 (see Table 1) which lie within 4 eV. With the energy provided by the laser pulse, the four ionisation channels can be activated.

The net effect of the perpendicular pulse polarisation on the HHG spectrum is to produce high-energy harmonics, beyond H25. These harmonics are much more intense than those of the molecular-plane average spectrum, when present. They are also higher than the corresponding ones in the HHG spectrum along the permanent dipole. One could formulate the hypothesis that this feature is exclusively determined by the ionisation from orbitals with the proper π symmetry, say HOMO and HOMO-1, which are more sensitive to the perpendicular polarisation than HOMO-2 and HOMO-3. The interaction between HOMO (and/or HOMO-1) and a perpendicular pulse enhances the HHG signal at high energies. In other words, one could argue that ionization from HOMO and/or HOMO-1 is responsible for the appearance of the second plateau in the perpendicular-polarisation HHG spectrum. A confirmation is indirectly given by the presence of harmonics up to H39 in the 3D-average spectrum, though softened by the average along the perpendicular out-of-plane component $\hat{n}_{s,z}$. These harmonics are instead missing in the molecular-plane HHG spectrum.

Looking at the molecular-plane spectrum, one can ask: since

the out-of-plane component of the pulse is exactly zero, do the harmonics of the spectrum (cutoff at H29-H31) derive only from σ orbitals as HOMO-2 and/or HOMO-3, which have the good symmetry, or, anyway, HOMO and/or HOMO-1 still contribute to the harmonic generation, despite the "unfavorable" symmetry? We can not provide a complete answer using the present computational protocol: a detailed analysis of the role of the various ionisation channels during the strong-field dynamics is mandatory to dissect the possible contributions and interference effects.

By summarizing, the HHG spectrum of uracil seems not to originate from HOMO ionisation only. Multiple ionisation channels, close in energy, seem to contribute to the HHG spectra.

The cutoff of the 3D-averaged spectrum in Figure 3, H39-H41, is consistent with the experimental findings of H33 and H39.

In Figure 5 we disentangle the contribution of energy states below and above E_T . We remind that E_T value used in our simulations corresponds to minus HOMO-3 energy. In the left panel, we show the HHG spectrum averaged in the plane of the molecule (Equation 8), the same type of HHG spectrum limited to states below E_T (Equation 11), and to states above E_T (Equation 13). Below- and above- E_T states contribute similarly to the full HHG spectrum. Above- E_T states are seen to give slight higher H17-H21 harmonics, and generally provide a higher background of the below- E_T . In the right panel of Figure 5 the same analysis is shown for the HHG spectra with perpendicular polarisation. Spectra were computed from the dipoles given in Equations 10 and 12, obtained by below- E_T and above- E_T states, respectively. In this case the states above E_T play a role only for the harmonics belonging to the first plateau of the HHG spectrum, while the energy states below E_T contribute everywhere and have a central role for the second plateau between H20 and H50. Differences between the above- E_T and below- E_T spectra are in this case dramatic, at variance with what occurs with the rotational average in the plane. The same behaviour is observed for the HHG spectrum polarised in the direction of the molecular permanent (in-plane) dipole (not shown). This could suggest that the behaviour observed for the HHG spectra in the left panel of Figure 5 is mainly due to the spatial average, which reduces differences between above- E_T and below- E_T states.

The HHG spectrum averaged in 3D contains by construction the behaviour of both the spectrum averaged in the plane together and of the spectrum with perpendicular polarisation (not shown).

4.2 Thymine

Results for thymine show the same general behaviour observed for uracil. This finding is not unexpected, since the electronic structure of uracil and thymine are rather similar.

The HHG spectrum of thymine with a perpendicular pulse polarisation shows much higher harmonics in the high-energy region with respect to the averaged ones, as collected in Figure 6. As for uracil, a second plateau is observed at around H19-H33. When compared with uracil results, the difference for thymine is even more pronounced. By averaging the dipole moment only with respect to the molecule plane, one gets a HHG spectrum

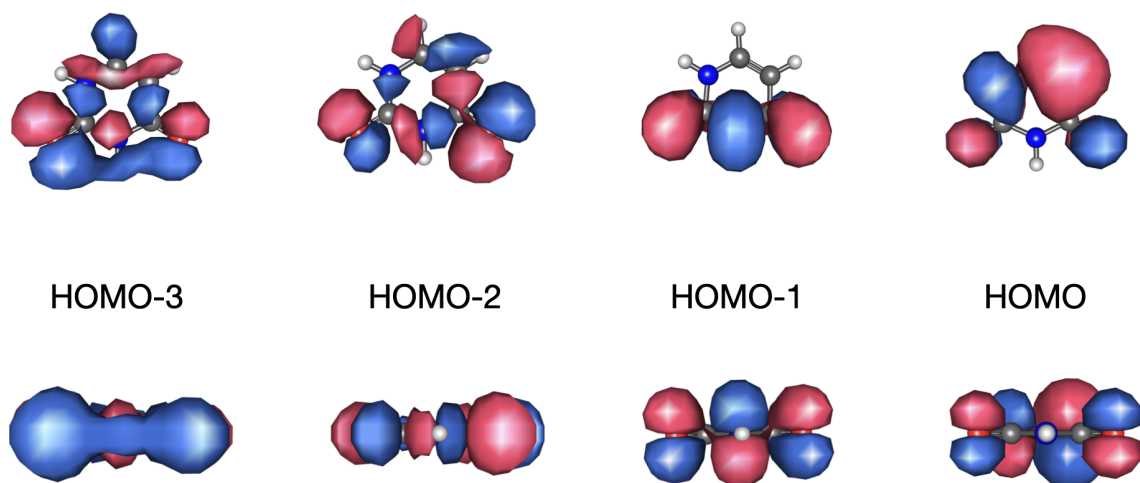


Fig. 4 Top (upper) and oxygen side (lower) view of HOMO-3, HOMO-2, HOMO-1 and HOMO orbitals of uracil.

which is very similar to that obtained by a full rotational average (Figure 6) up to H19. Cutoff values for the average molecular-plane and 3D spectra are H23 and H39-H41, respectively. The comments applied to uracil can be automatically done for thymine: Hartree-Fock molecular orbitals of thymine (Figure 7) have the same symmetry of the uracil ones. HOMO, HOMO-1, HOMO-2 and HOMO-3 are even closer in energy than the uracil orbitals. Hence, also for thymine different ionisation channels are supposed to play a significant role in the strong-field electron dynamics, with the same caveats reported for uracil.

By using our all-electron quantum approach the HHG spectrum of thymine naturally arises, as expected. This result points out that the absence of HHG signal in the experimental work⁴⁷ does not originate from intrinsic features of the molecule, confirming that the issue has an extra-molecule explanation⁴⁷. However, comparison between the HHG spectra in Figures 3 and 6 shows the thymine 3D-averaged high-energy signal is lower than that of thymine at high energies. Though all the ionisation energies (minus orbital energies, see Table 1) for thymine are smaller than those of uracil, one observes a larger cutoff for thymine, when the 3D average is considered: this finding could suggest that uracil and thymine are characterised by different couplings of the various ionisation channels, making the cutoff value a complicated function of individual ionisation energies. Contribution analysis for thymine is reported in Figure 8. The trend observed for uracil also characterizes the thymine spectra, with few major exceptions: i) HHG spectra from above- and below- E_T are nearly identical for the average case; ii) contribution from below- E_T states for the perpendicular pulse polarisation is considerably small in the middle-energy range. For both molecules, the smaller above- E_T contribution to the HHG spectrum can be quantitatively explained by the smaller number of above- E_T states described by

the 3aug-ccpVDZ+3K basis set, with respect of that of below- E_T states. It is worth mentioning that a simplified description of the electronic structure, as provided by only the ground and continuum states, is not accurate enough to account for the complex richness of the HHG signal of many-electron systems as uracil and thymine: accounting for bound, below- E_T states is mandatory to successfully represent their HHG spectrum. However, a detailed investigation on the interplay between the chosen basis set and the ionisation energy could give a deeper insight on the general quality of the propagated wavepacket.

5 Conclusions

In this work we presented HHG spectra of uracil and thymine. HHG spectra have been computed using the all-electron RT-TD-CIS computational protocol based on the real-time propagation of the electron wavepacket, obtained using CIS and Gaussian basis sets adapted for the continuum.

We simulated randomly oriented molecules by averaging the time-dependent dipoles projected on the different polarisations in the molecular plane or in the 3D space. We also computed the HHG spectra with a pulse polarisation perpendicular to the molecular plane or along the direction of the ground-state permanent dipole. This theoretical work contributed to the understanding and characterisation of the experiments by Hutchinson *et al.*⁴⁷.

We discussed the differences between the sample in the original work (unfragmented molecule and fragments) and in our calculations (unfragmented molecule). In the case of the 3D-averaged HHG spectrum our results can be directly compared with those in Ref. 47, and the computed uracil spectrum is consistent with experimental data.

Differences between rotational-averaged and single-orientation

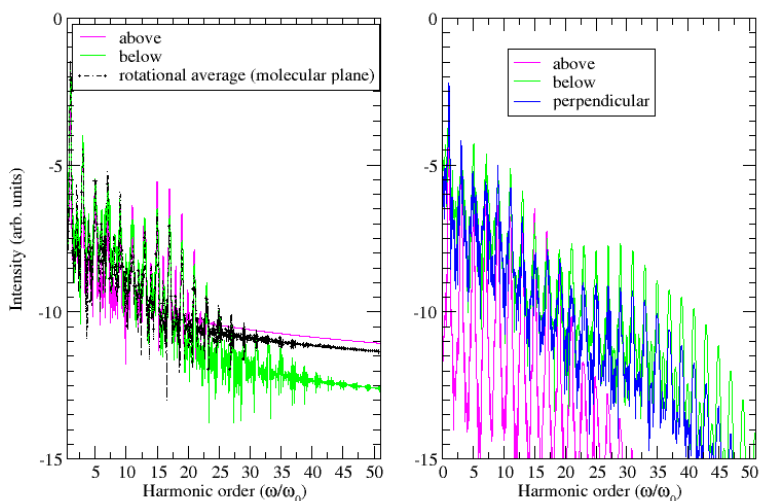


Fig. 5 Contribution to the HHG spectrum of uracil from below- and above-threshold states for a rotational average in the molecular plane (left panel) and a perpendicular polarisation of the pulse (right panel). Pulse wavelength and intensity of 780 nm and 10^{14} W/cm².

sis given in Ref. 47 according to which the absence of the thymine HHG spectrum in the experiment is not caused by intrinsic features of the molecular system.

Furthermore, for both systems the contributions from energy states below or above the ionisation from the HOMO-3 is generally more important for spectra from single polarisations. In conclusion, our approach has been seen to be accurate and reliable in describing HHG of biomolecules, and opens the way to predictive and interpretative studies of strong-field processes occurring in large molecules.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

EC acknowledges funding from the University of Trieste under the grant "Microgrants 2020". Authors acknowledge the Computation Centre of the University of Trieste for computational support.

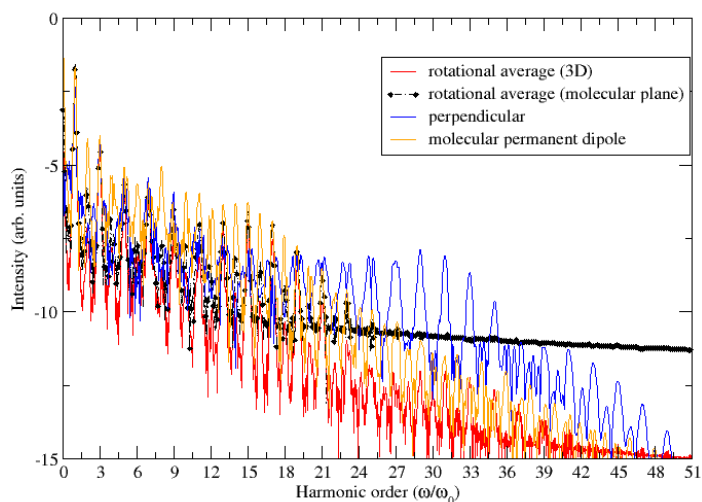


Fig. 6 HHG spectra of thymine with pulse wavelength and intensity of 780 nm and 10^{14} W/cm².

HHG spectra indicate that the HHG signal can originate from more than one ionisation channel. We discussed the possible role of π and σ orbitals in modulating the shape of the HHG spectrum.

Moreover, we obtain a clear HHG spectrum of thymine for all the different calculations schemes. Though less intense than the uracil HHG spectra at high energy, the thymine HHG spectra are certainly obtained. For this reason our work enforces the hypothe-

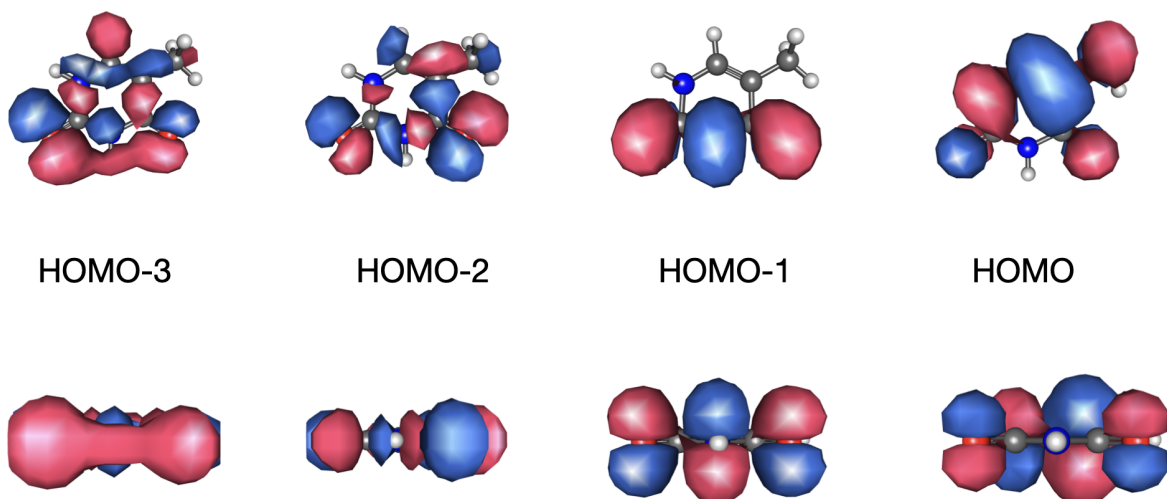


Fig. 7 Top (upper) and oxygen side (lower) view of HOMO-3, HOMO-2, HOMO-1 and HOMO orbitals of thymine.

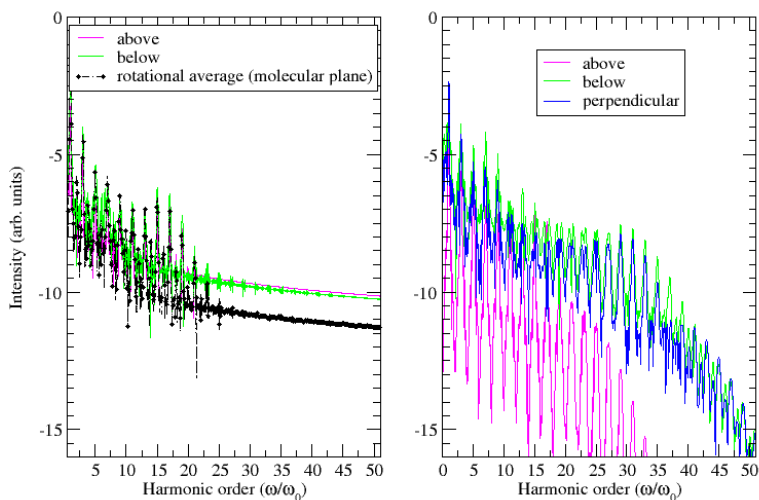


Fig. 8 Contribution to the HHG spectrum of thymine from below- and above-threshold states for a rotational average in the molecular plane (left panel) and a perpendicular polarisation of the pulse (right panel). Pulse wavelength and intensity of 780 nm and 10^{14} W/cm².

Notes and references

- 1 P. B. Corkum and F. Krausz, *Nat. Phys.*, 2007, **3**, 381.
- 2 F. Krausz and M. Ivanov, *Rev. Mod. Phys.*, 2009, **81**, 163–234.
- 3 S. Haessler, J. Caillat, W. Boutu, C. Giovanetti-Teixeira, T. Ruchon, T. Auguste, Z. Diveki, P. Breger, A. Maquet, B. Carré, R. Taieb and P. Salières, *Nature Phys.*, 2010, **6**, 200–206.

- 4 S. Baker, I. A. Walmsley, J. W. G. Tisch and J. P. Marangos, *Nat. Phys.*, 2011, **5**, 664–671.
- 5 M. Nisoli, P. Decleva, F. Calegari, A. Palacios and F. Martín, *Chem. Rev.*, 2017, **117**, 10760.
- 6 M. Ossiander, F. Siegrist, V. Shirvanyan, R. Pazourek, A. Sommer, T. Latka, A. Guggenmos, S. Nagele, J. Feist, J. Burgdörfer, R. Kienberger and M. Schultze, *Nat. Phys.*, 2017, **13**, 280–285.
- 7 P. Peng, C. Marceau and D. M. Villeneuve, *Nat. Rev. Phys.*, 2019, **1**, 144–155.
- 8 G. Sansone, *Nat. Photon.*, 2020, **14**, 131–133.
- 9 M. Hentschel, R. Kienberger, C. Spielmann, G. A. Reider, N. Milosevic, T. Brabec, P. Corkum, U. Heinzmann, M. Drescher and F. Krausz, *Nature*, 2001, **414**, 509.
- 10 P. M. Kraus and H. J. Wörner, *Angew. Chem. Int. Ed.*, 2018, **57**, 5228.
- 11 P. M. Kraus, M. Zürch, S. K. Cushing, D. M. Neumark and S. R. Leone, *Nat. Rev. Chem.*, 2018, **2**, 82.
- 12 T. T. Gorman, T. D. Scarborough, P. M. Abanador, F. Mauger, D. Kieseewetter, P. Sándor, S. Khatri, K. Lopata, K. J. Schafer, P. Agostini, M. B. Gaarde and L. F. DiMauro, *J. Chem. Phys.*, 2019, **150**, 184308.
- 13 A. J. Uzan, H. Soifer, O. Pedatzur, A. Clergerie, S. Larroque, B. D. Bruner, B. Pons, M. Ivanov, O. Smirnova and N. Dudovich, *Nat. Photon.*, 2020, **14**, 188–194.
- 14 X. Chu and S.-I. Chu, *Phys. Rev. A*, 2001, **63**, 023411.
- 15 J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Ppin, J. C. Kieffer, P. B. Corkum and D. M. Villeneuve, *Nature*, 2004, **432**, 867–871.
- 16 L. Cui, J. Zhao, Y. J. Hu, Y. Y. Teng, X. H. Zeng and B. Gu, *Appl. Phys. Lett.*, 2006, **89**, 211103.

- 17 S. Patchkovskii, Z. Zhao, T. Brabec and D. M. Villeneuve, *Phys. Rev. Lett.*, 2006, **97**, 123003.
- 18 N. Kajumba, R. Torres, J. G. Underwood, J. S. Robinson, S. Baker, J. W. G. Tisch, R. de Nalda, W. A. Bryan, R. Velotta, C. Altucci, I. Procino, I. C. E. Turcu and J. P. Marangos, *New J. of Phys.*, 2008, **10**, 25008.
- 19 D. A. Telnov and S.-I. Chu, *Phys. Rev. A*, 2009, **80**, 043412.
- 20 O. Smirnova, Y. Mairesse, S. Patchkovskii, N. Dudovich, D. Villeneuve, P. Corkum and M. Y. Ivanov, *Nature*, 2009, **460**, 972.
- 21 H. J. Wörner, J. Bertrand, P. Hockett, P. Corkum and D. Villeneuve, *Phys. Rev. Lett.*, 2010, **104**, 233904.
- 22 S. Sukiasyan, S. Patchkovskii, O. Smirnova, T. Brabec and M. Y. Ivanov, *Phys. Rev. A*, 2010, **82**, 043414.
- 23 M. C. H. Wong, J.-P. Brichta, M. Spanner, S. Patchkovskii and V. R. Bhardwaj, *Phys. Rev. A*, 2011, **84**, 051403(R).
- 24 J. Heslar, D. Telnov and S.-I. Chu, *Phys. Rev. A*, 2011, **83**, 043414.
- 25 X. Chu and G. C. Groenenboom, *Phys. Rev. A*, 2012, **85**, 53402.
- 26 A. Wardlow and D. Dundas, *Phys. Rev. A*, 2016, **93**, 023428.
- 27 P. Mulholland and D. Dundas, *Phys. Rev. A*, 2017, **97**, 43428.
- 28 M. Ruberti, P. Decleva and V. Averbukh, *Phys. Chem. Chem. Phys.*, 2018, **20**, 8311.
- 29 C.-T. Le, D.-D. Vu, C. Ngo and V.-H. Le, *Phys. Rev. A*, 2019, **100**, 053418.
- 30 B. Wang, Y. He, X. Zhao, L. He, P. Lan, P. Lu and C. D. Lin, *Phys. Rev. A*, 2020, **101**, 63417.
- 31 C. Vozzi, M. Negro, F. Calegari, G. Sansone, M. Nisoli, S. D. Silvestri and S. Stagira, *Nat. Photon.*, 2011, **7**, 822.
- 32 P. Salières, A. Maquet, S. Haessler, J. Caillat, and R. Taïeb, *Rep. Prog. Phys.*, 2012, **75**, 062401.
- 33 H. J. Wörner, H. Niikura, J. B. Bertrand, P. B. Corkum and D. M. Villeneuve, *Phys. Rev. Lett.*, 2009, **102**, 103901.
- 34 B. D. Bruner, Z. Masin, M. Negro, F. Morales, D. Brambila, M. Devetta, D. Faccialá, A. G. Harvey, M. Ivanov, Y. Mairesse, S. Patchkovskii, V. Serbinenko, H. Soifer, S. Stagira, C. Vozzi, N. Dudovich and O. Smirnova, *Faraday Discuss.*, 2016, **194**, 364.
- 35 D. R. Tuthill, F. Mauger, T. D. Scarborough, R. R. Jones, M. B. Gaarde, K. Lopata, K. J. Schafer and D. M. L. F., *J. Mol. Spectrosc.*, 2020, **372**, 111354.
- 36 W. Li, X. Zhou, R. Lock, S. Patchkovskii, A. Stolow, H. C. Kapteyn and M. M. Murnane, *Science*, 2008, **322**, 1207.
- 37 E. Coccia, *Mol. Phys.*, 2020, <https://doi.org/10.1080/00268976.2020.1769871>.
- 38 S. Kim, J. Jin, Y.-J. Kim, I.-Y. Park, Y. Kim and S.-W. Kim, *Nature*, 2008, **453**, 757.
- 39 J. P. Marangos, *J. Phys. B*, 2016, **49**, 132001.
- 40 L. B. Clark, G. G. Peschel and I. Tinoco, *J. Phys. Chem.*, 1965, **69**, 3615.
- 41 L. Li and D. M. Lubman, *Anal. Chem.*, 1987, **59**, 2538.
- 42 H. Du, R. C. A. Fuh, J. Z. Li, L. A. Corkan and J. S. Lindsey, *Photochem. Photobiol.*, 1998, **68**, 141.
- 43 M. Barbatti, A. J. A. Aquino and H. Lischka, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4959.
- 44 V. A. Ovchinnikov and D. Sundholm, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6931.
- 45 T. J. A. Wolf, R. Myhre, J. P. Cryan, S. Coriani, R. J. Squibb, A. Battistoni, N. Berrah, C. Bostedt, P. Bucksbaum, G. Coslovich, R. Feifel, K. J. Gaffney, J. Grilj, T. J. Martinez, S. Miyabe, S. P. Moeller, M. Mucke, A. Natan, R. Obaid, T. Osipov, O. Plekan, S. Wang, H. Koch and M. Gühr, *Nat. Commun.*, 2017, **8**, 29.
- 46 C. Ehlert, M. Gühr and P. Saalfrank, *J. Chem. Phys.*, 2018, **149**, 144112.
- 47 C. Hutchison, R. Ganeev, M. Castillejo, I. Lopez-Quintas, A. Zair, S. Weber, F. McGrath, Z. Abdelrahman, M. Oppermann, M. Martin, D. Lei, S. Maier, J. Tisch and J. Marangos, *Phys. Chem. Chem. Phys.*, 2013, **15**, 12308.
- 48 P. Krause, T. Klamroth and P. Saalfrank, *J. Chem. Phys.*, 2005, **123**, 074105.
- 49 P. Krause, T. Klamroth and P. Saalfrank, *J. Chem. Phys.*, 2007, **127**, 034107.
- 50 S. Klinkusch, P. Saalfrank and T. Klamroth, *J. Chem. Phys.*, 2009, **131**, 114304.
- 51 E. Luppi and M. Head-Gordon, *Mol. Phys.*, 2012, **110**, 909.
- 52 E. Luppi and M. Head-Gordon, *J. Chem. Phys.*, 2013, **139**, 164121.
- 53 A. White, C. J. Heide, P. Saalfrank, M. Head-Gordon and E. Luppi, *Mol. Phys.*, 2016, **114**, 947.
- 54 E. Coccia and E. Luppi, *Theor. Chem. Acc.*, 2016, **135**, 43.
- 55 E. Coccia, B. Mussard, M. Labeye, J. Caillat, R. Taïeb, J. Toulouse and E. Luppi, *Int. J. Quant. Chem.*, 2016, **116**, 1120.
- 56 E. Coccia, R. Assaraf, E. Luppi and J. Toulouse, *J. Chem. Phys.*, 2017, **147**, 014106.
- 57 M. Labeye, F. Zapata, E. Coccia, V. Veniard, J. Toulouse, J. Caillat, R. Taïeb and E. Luppi, *J. Chem. Theory Comput.*, 2018, **14**, 5846.
- 58 E. Coccia and E. Luppi, *Theor. Chem. Acc.*, 2019, **138**, 96.
- 59 F. Bedurke, T. Klamroth, P. Krause and P. Saalfrank, *J. Chem. Phys.*, 2019, **150**, 234114.
- 60 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
- 61 K. Kaufmann, W. Baumeister and M. Jungen, *J. Phys. B: At. Mol. Opt. Phys.*, 1989, **22**, 2223.
- 62 H. B. Schlegel, S. M. Smith and X. Li, *J. Chem. Phys.*, 2007, **126**, 244110.
- 63 P. Krause and H. Schlegel, *J. Chem. Phys.*, 2014, **141**, 174104.
- 64 J. A. Sonk, M. Caricato and H. B. Schlegel, *J. Phys. Chem. A*, 2011, **115**, 4678.
- 65 P. Saalfrank, F. Bedurke, C. Heide, T. Klamroth, S. Klinkusch, P. Krause, M. Nest and J. C. Tremblay, *Adv. Quant. Chem.*, 2020, **81**, 15.
- 66 A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009.
- 67 Y. Shao, L. F. Molnar, Y. Jung, J. r. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr, R. C. Lochan, T. Wang, G. J. O.

- Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. Simmonett, J. E. Subotnik, H. Lee Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172.
- 68 P. B. Corkum, *Phys. Rev. Lett.*, 1993, **71**, 1994.
- 69 M. Lewenstein, P. Balcou, M. Y. Ivanov, A. L'Huillier and P. B. Corkum, *Phys. Rev. A*, 1994, **49**, 2117.
- 70 A. Padva, P. R. LeBreton, R. J. Dinerstein and J. N. A. Ridyard, *Biochem. Biophys. Res. Commun.*, 1974, **60**, 1262.
- 71 N. S. Hush and A. S. Cheung, *Chem. Phys. Lett.*, 1975, **34**, 11.
- 72 G. Lauer, W. Schafer and A. Schweig, *Tetrahedron Lett.*, 1975, **45**, 3939.
- 73 D. Dougherty, K. Wittel, J. Meeks and S. P. McGlynn, *J. Am. Chem. Soc.*, 1976, **98**, 3815.
- 74 M. Palmer, I. Simpson and R. Platenkamp, *J. Mol. Struct.*, 1980, **66**, 243.
- 75 E. Itälä, D. T. Ha, K. Kooser, E. Rachlew, M. A. Huels and E. Kukk, *J. Chem. Phys.*, 2010, **133**, 154316.
- 76 I. Lopez-Quintas, M. Oujja, M. Sanz, A. Benitez-Canete, C. Hutchison, R. de Nalda, M. Martin, R. Ganeev, J. P. Marangos and M. Castillejo, *Appl. Surf. Sci.*, 2014, **302**, 299.
- 77 A. Semmeq, S. Ouaskit, A. Monari and M. Badawi, *Phys. Chem. Chem. Phys.*, 2019, **21**, 4810.