Please do not adjust margins

Journal Name

Two-dimensional electronic spectroscopy as a tool for tracking molecular conformations in DNA/RNA aggregates

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Javier Segarra-Martí,[a,*] Vishal K. Jaiswal,[b] Ana Julieta Pepino,[b] Angelo Giussani,[c] Artur Nenov,[b] Shaul Mukamel,[d] Marco Garavelli[b,*] and Ivan Rivalta[a,*]

A computational strategy to simulate two-dimensional electronic spectra (2DES) is introduced, which allows us to analyse ground state dynamics and to sample and measure different conformations attained by flexible molecular systems in solution. An explicit mixed quantum mechanics/molecular mechanics (QM/MM) approach is employed for the evaluation of the necessary electronic excited state energies and transition dipole moments. The method is applied towards the study of the highly flexible water-solvated adenine-adenine monophosphate (ApA), a system featuring two interacting adenine moieties that display variours intermolecular arrangements, known to deeply affect their photochemical outcome. Molecular dynamics simulations and cluster analysis have been used to select molecular conformations, reducing the complexity of the flexible ApA conformational space. By using our Sum-over-states (SOS) approach to obtain 2DES spectra for each of these selected conformations, we can discern spectral changes and relate them to specific nuclear arrangements: close lying π-stacked bases display a splitting of their respective ^1L_a signal traces; T-stacked ones show the appearance of charge transfer states in the low-energy Vis probing window while displaying no ^1L_a splitting, being particularly favoured when promoting amino to 5-ring interactions; unstacked and distant adenine moieites feature signals similar to those of adenine monomer, as expected for non-interacting nucleobases. 2DES maps thus reveal spectral fingerprints associated with specific molecular conformations, a promising option to enable their quantitative spectroscopic detection beyond standard 1D pump-probe techniques. This is expected to aid the understanding of how nucleobase aggregation controls and modulates the photostability and photo-damage of extended DNA/RNA systems.

Introduction

The conformations attained by molecular systems in condensed (liquid or solid) phases are often surrounded by controversy given the difficulties faced by theorists and experimentalists in locating and assigning specific spectroscopic fingerprints pertaining to a given molecular motif. Various molecular arrangements, which often display completely different photophysical properties due to their featured intermolecular interactions, are thus key to properly assessing and rationalising certain bulk and interfacial phenomena of paramount interest, such as the ultrafast and efficient electronic energy transfer processes registered in light harvesting complexes or the specific charge/transport mechanisms behind state-of-the-art solar cells.[1] This is especially true in biomolecular systems, where a specific conformation may impair or even completely alter a concrete cellular function. One particularly important case highlighting this effect would be the photostability of the genetic material,[2, 3] and how the photophysics of DNA/RNA sequences are severely affected in dimeric and multimeric species by the presence of additional neighbouring stacked nucleobases within a given strand in the double helix. Nuclebase stacking might indeed increase their excited state lifetime and potentially make them more prone to photochemical mutations,[4-7] in contrast to the efficient and ultrafast decay registered in isolated nucleobases,[8, 9] which is expected to strongly promote its photo-protection mechanisms.[10] This effect, believed to strongly modulate the complexity of DNA photophysics,[11, 12] is particularly hard to measure within standard pump-probe set-ups, the experimental workhorse in photochemistry. In standard 1D pump-probe spectroscopy, different signals from diverse molecular motifs overlap strongly in space and time, leading to highly congested spectral regions whose complex interpretation remains elusive.

A recently developed technique that can overcome these aforementioned limitations is two-dimensional (2D) electronic spectroscopy (2DES), which able to disentangle signals coming from different intermolecular interactions that remain hidden in 1D pump-probe spectra. In these multi-pulse techniques, amply used in the IR[13, 14] and being recently extended to the electronic regime,[15-17] the system is excited by a sequence of three pulses with controlled delays t_1 (coherence time) and t_2 (population/waiting time), the

[a] Laboratoire de Chimie UMR 5182, Univ Lyon, ENS de Lyon, CNRS, Université Lyon 1, 46 Allée d'Italie, FR-69342 Lyon, France. E-mail: Javier.segarra-marti@ens-lyon.fr; ivan.rivalta@ens-lyon.fr.
[b] Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Viale del Risorgimento, IT-40136, Italy. E-mail: marco.garavelli@unibo.it.
[c] Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.
[d] Department of Chemistry, University of California, Irvine, California 92697-2025, USA.

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Please do not adjust margins

third pulse at time t_3 emitting the third-order nonlinear signal $S(t_1,t_2,t_3)$.[18-21] Fourier transform with respect to t_1 and t_3, one can extract a sequence of 2D maps $S(\Omega_1,t_2,\Omega_3)$ for different waiting times t_2. These 2D maps contain a vast number of signals that can potentially be employed for tracking all sorts of localised, delocalised, energy and charge transfer photoinduced phenomena with unprecedented detail. However, their interpretation proves to be challenging. It is in this context that theoretical chemistry comes in handy, by simulating the spectra we can separate the different peaks for their contributing specific electronic transitions, providing a route map to recognise, understand and interpret the main fingerprints observed.[22, 23]

In this contribution we introduce a computational strategy that combines ground-state molecular dynamics with QM/MM approaches in order to sample the conformational space of dimeric and multimeric flexible bio-chromophores and separate their different nuclear arrangements to be characterised spectroscopically by 2DES. The methodology is applied to ApA, a model system that contains both unstacked (non-interacting) and π- or T-stacked (interacting) adenine moieties, and is shown to display distinct fingerprints for these different conformations, proving to be a promising tool to separate them and understand precisely how they impact the underlying photophysical and photochemical pathways in such complex and extended DNA/RNA systems.

Computational Details

Molecular dynamics simulations were carried out by using the AMBER 11 software package.[24, 25] A 12x12x12 Å cubic box was employed making use of parmbsc0[26] and TIP3P[27] force fields for the ApA moiety and water molecules, respectively. A 100 ns trajectory was carried out, extracting one snapshot every 2.5 ps, totalling 40000 frames. A root mean square deviation (RMSD) cluster analysis was performed on top of those 40000 structures and set within 1.5 Å, extracting 89 different clusters. From these, 17 were retained for further electronic structure theory evaluations, accounting for up to ~75% of the conformational space sampled by the highly flexible ApA system. The selected snapshots, which the closest structures to each centroid, are subsequently optimised employing a complete active space self-consistent field (CASSCF) comprising 8 electrons in 8 orbitals, 4 π occupied and 4 π* unoccupied, constraining phosphate and sugar groups and allowing relaxations only on the adenine moieties to attain more accurate bond length distances than those given by the force field. This has been done due to the strong dependence of the high-energy excited state manifold on the bond length distances displayed by adenine, as it has been shown elsewhere.[28]

All electronic structure computations were carried out by employing the MOLCAS 8 package.[29] Restricted active space self-consistent field (RASSCF)[30] and its second-order perturbation theory extension (RASPT2)[31, 32] were used on top of the aforementioned selected structures. Cholesky decomposition was used to speed-up the two-electron integrals[33-35] and in conjunction with Atomic Natural Orbital (ANO)[36] large-type basis sets with valence double-ζ polarised contractions. Only the adenine nucleobases were included in the quantum mechanical region, with atom links connecting the QM (adenine moieties) and MM (phosphate and water molecules) layers.[37] An active space of 24 electrons in 18 orbitals was employed, comprising the whole π occupied valence space of both adenine moieties and all except the highest π* virtual

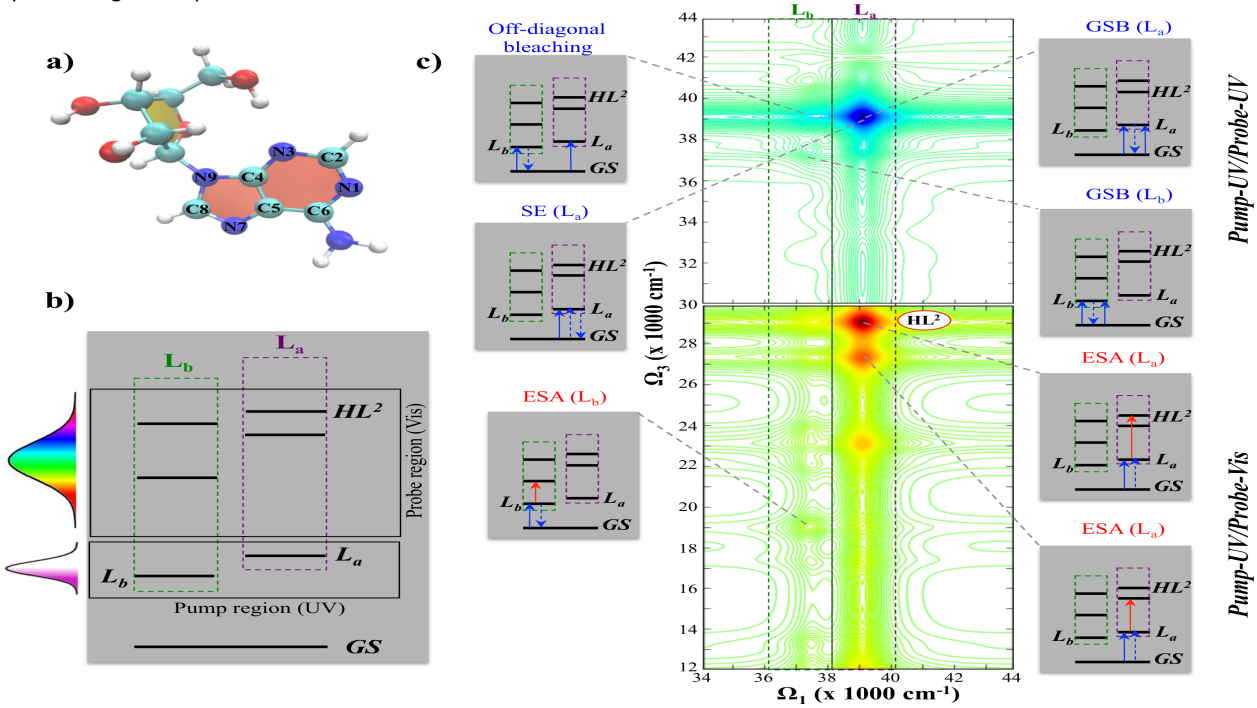


Figure 1. 2DES spectrum of adenosine monophosphate, where a) displays the molecular system and atom labelling, b) provides a scheme of the energy levels accessible with different laser pulses, and c) shows the 2DES in the Vis (lower panel) and UV (upper panel) probe windows, together with energy level schemes depicting the different signals featured in terms of the laser pulse sequence. Red colour denotes positive (ESA) signals. Blue colour denotes negative (GSB, off-diagonal bleach and SE) and signals. The SE signal associated to the ^1L_b state was not included in the scheme for simplicity, being placed on top of GSB(L_b). The signal intensity in each spectrum is normalised according to the absolute value of the most intense signal.

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Please do not adjust margins

orbital of each moiety. Up to four holes/electrons were included in the RAS1/RAS3 subspaces, respectively, leading to RASPT2(4,12|0,0|4,6), and are averaged with equal weights over 100 states. A RASPT2(4,6|0,0|4,3) was used for the monomer as previously benchmarked[28] and averaged over 25 states, its active space being duplicated for the dimer. This particular active space was taken considering previous benchmark studies where its adequacy to represent the energies and transition dipole moments with respect to reference RASPT2(0,0|12,10|2,12) computations on the adenine monomer were assessed.[28] An imaginary level shift[38] of 0.6 a.u. was employed for the dimer and 0.5 a.u. for the monomer as previously benchmarked,[28] the IPEA shift being set to zero.[39, 40] Restricted active space state interaction (RASSI) method[41, 42] was employed to obtain the required transition dipole moments among the different electronic ground and excited states.

The resulting energy levels and transition dipole moments were then employed in conjunction with the Sum-Over-States (SOS) approach[43] within the dipole approximation and a QM/MM scheme (SOS//QM/MM)[18] to simulate the 2DES as implemented in the Spectron 2.7 program.[44] Further details on the working equations have been given elsewhere.[28] Spectral lineshapes are simulated assuming that the dephasing is caused by pure dephasing in the Markovian approximation. Effectively, the spectra of the individual snapshots are subject to a constant line broadening of 1000 cm^{-1}, which has been set so that the broadenings along Ω_1 for the total ApA spectrum mimic the experimental broadening recently measured in 2DUV experiments for adenine.[17] Although several computational approaches have been devised recently to consider spectral diffusion and non-adiabatic effects on the simulated 2DES,[45-49] the size of the current system prevents the usage of more sophisticated approaches. Quasi-absorptive signals were simulated for the condition of a collinear pump pulse pair (thus recording simultaneously rephrasing and non-rephasing contributions). All signals reported use the all-parallel *xxxx* pulse polarisation configuration, are plotted on a linear scale and consider white light (infinite broadband) pulses. This allows us to explore the variety of signals that may be present in the 2D spectra. Both ground state bleaching (GSB) and stimulated emission (SE) contributions appear as negative (blue) peaks, whereas excited-state absorptions (ESAs) appear as positive (red) peaks in the 2D spectra.

Results

The 2DES of water-solvated adenosine is considered first (see Figure 1) and then used as a reference to interpret the signals arising in the ApA dimer. As shown in Figure 1a, adenosine features a sugar group connected to the nucleobase through the N9 position analogously to the adenine moieties contained in ApA and is expected to feature mostly the same spectral signals as those of adenine while not displaying the known mix of tautomers (7H- and 9H-adenine) in the bare nucleobase embodying different excited state lifetimes[50, 51] that complicates its analysis. As can be seen in Figure 1b, by pumping (photo-exciting) in the absorbing UV region, two main excited states of ππ* character are expected to be populated, a more stable and less intense ^1L_b state and a more intense and higher in energy ^1L_a state, even though their relative energetic state ordering is still under debate.[28, 52, 53] These two neighbouring excited states accessible from the ground state give rise to two traces along Ω_1 in the 2DES spectra, featuring ESAs in the Vis probing window and the GSB and off-diagonal bleaching signals in the UV. This strong GSB contribution mostly obscures all signals from ESAs in the UV window. Figure 1c shows the spectra together with a simple scheme depicting the pulse sequence leading to each of the signals, arrows representing electronic transitions produced by the multi-pulse laser sequence. GSB arises from an excitation from the ground state to a given excited state, de-excitation from that state back to the ground state and again an excitation to the very same excited state. Off-diagonal bleaching follows the first two excitations as GSB, with the third pulse promoting excitation to another excited state. SE and ESA signals are formed from a two-pulse coherent excitation from the ground state to a given excited state, which leads to the excited state's population, from which the third pulse either promotes excitation to higher lying states (ESA) or emission back to the ground state (SE).

Our results for water-solvated adenosine are compared against previous reference computations on gas-phase adenine performed with larger active spaces and basis sets *in vacuo*,[28] displaying notable differences along Ω_1 (linear absorption of the photo-activated ^1L_a and ^1L_b states of adenine) by being red-shifted by 0.2-0.3 eV, and which are partially expected due to the addition of a polar protic solvent like water that produces an analogous red-shifting effect in DNA/RNA nucleobases.[37, 54-56] It is worth noting that more accurate approaches could be used to compute the spectrum of the monomer as shown in previous benchmarks[28] yet those cannot be extended to dimeric species, justifying our use of a cost-effective RASPT2(4,6|0,0|4,3) scheme. Beyond the red-shift along Ω_1, the other main difference observed with respect to the reference is the relative intensity of the ^1L_b trace with respect to the ^1L_a, which appears to be obscured due to the usage of infinitely wide pulses that enhance the ^1L_a trace, as opposed to the tailored finite pulses employed in the reference computations to highlight this relatively dark state.[28] The absence of signals in the low-energy window as opposed to the reference computations[28] can be attributed to the use of infinitely broad pulses, which obscures them as it has also been reported for the ^1L_b trace. The usage of such pulses allows us to focus on the overall most intense signals without considering particular pump and probe set-ups that could be potentially use with finite pulses to enhance specific peaks and that is for the moment out of scope of the present manuscript. Our reduced RASPT2(4,6|0,0|4,3) model is able to quantitatively reproduce the most important signal, that is, the one related to the transition between the ^1L_a and the doubly excited HOMO to LUMO state placed ~29000 cm^{-1} along Ω_3 (HL^2), which is very intense and appears in different spectral regions for the different nucleobases[28, 57] being the main monomer-based fingerprint to potentially follow excited state decays on separate monomers simultaneously. An additional signal right below HL^2 can be observed in the spectrum (see Figure 1c), being related to a ππ*<-^1L_a transition that appears to be less intense and that, given additional homogeneous broadenings were included, would most likely merge with that of HL^2, its in-depth analysis being considered not as important. In the

Please do not adjust margins

Please do not adjust margins

higher energy UV-probe window all signals are obscured in solution as opposed to what was observed in the gas phase reference computations,[28] where ESAs along both ^1L_b and ^1L_a traces could still be observed albeit with low intensity compared to the GSB signals.

Table 1. Clustering parameters coming from the 100 ns molecular dynamics simulation of ApA. Relative weights are normalised and fit to (‰) for the 2DES simulations. C5-C5 distance ranges are given (in Å) for each cluster.

Cluster	nº snapshots contained	% over total (40000 snapshots)	Normalised % (over 29421 snapshots)	Weight (‰)	C5-C5 distance (Å)
A	4063	10.16	13.81	138	> 6.0
B	3542	8.86	12.04	120	4.0 – 4.5
C	3470	8.68	11.79	118	< 4.0
D	1916	4.79	6.51	65	< 4.0
E	1744	4.36	5.93	59	4.5 – 5.0
F	1713	4.28	5.82	58	4.5 – 5.0
G	1689	4.22	5.74	57	4.0 – 4.5
H	1681	4.20	5.71	57	5.0 – 6.0
I	1661	4.15	5.65	57	4.5 – 5.0
J	1498	3.75	5.09	51	> 6.0
K	1496	3.74	5.08	51	< 4.0
L	1245	3.11	4.23	42	4.5 – 5.0
M	917	2.29	3.12	31	> 6.0
N	821	2.05	2.79	28	< 4.0
O	793	1.98	2.70	27	5.0 – 6.0
P	601	1.50	2.04	20	4.0 – 4.5
Q	571	1.43	1.94	19	5.0 – 6.0
TOTAL	29421	73.55%	99.99%	998	

As indicated previously, in order to obtain a 2DES spectrum representing the whole conformational space, a reductionist approach was employed where a cluster analysis was performed and the most populated clusters along the dynamics are used to reduce the number of QM/MM evaluations, the actual bottleneck of the computational procedure. In this way, by considering the first 17 clusters (from the most to the least populated), which have been labelled in alphabetical (from A to Q) order, we can account for up to ~75% of the sampled conformational space of ApA. All details are given in Table 1, which contains the concrete number of structures contained in each cluster, their contribution to the overall conformational space, a normalised estimate (assuming sampling of the full conformational space) and the corresponding weights (in ‰) employed in the simulation of the total 2DES. As shown in Table 1, populations drop quickly in weight, the first 8 clusters accounting for ~50% of the population while the next 9 represent ~25% of the conformational space, cluster A accounting for 7 times the contributions of cluster Q. The ~75% conformational space considered is assumed to be made into the full space for obtaining the concrete weighting factors to be used for each of the clusters, thus accounting for their different contributions to the overall 2DES of ApA. At this point it is important to note that a more accurate clustering method than the RMSD adopted here would be desirable in order to obtain less clusters yet more representative, thus further minimising the number of QM/MM evaluations required and provide better weightings for the simulation of the full 2DES spectrum. Accurate weightings leading to direct comparison with future experiments are beyond the scope of this work, which focuses instead on setting a viable strategy to simulate 2DES spectra and the inhomogeneous broadenings due to different molecular conformations. We have opted, indeed, for a qualitative treatment where the length of the ground state molecular dynamics simulation might not be enough to sample the full conformational space of ApA. Moreover, the force field employed was parameterised for double helix DNA chains (and not strictly for flexible systems like ApA), which may introduce an additional source of error. However, employing the so-called "stacking coordinate" introduced by Florián and co-workers[58] we obtain a ~45% population within a stacking coordinate of 4.2, which, assuming is representative of stacked instances, seems to be in agreement with reported circular dichorism ApA estimates and attributed to these interacting arrangements.[7] Considering all the caveats, we envision that this computational strategy could be systematically improved in various ways by employing for instance longer simulations, more suitable force fields, more accurate clustering techniques and *ab initio* simulations as well as more sophisticated treatments for the spectral line shapes.

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Please do not adjust margins

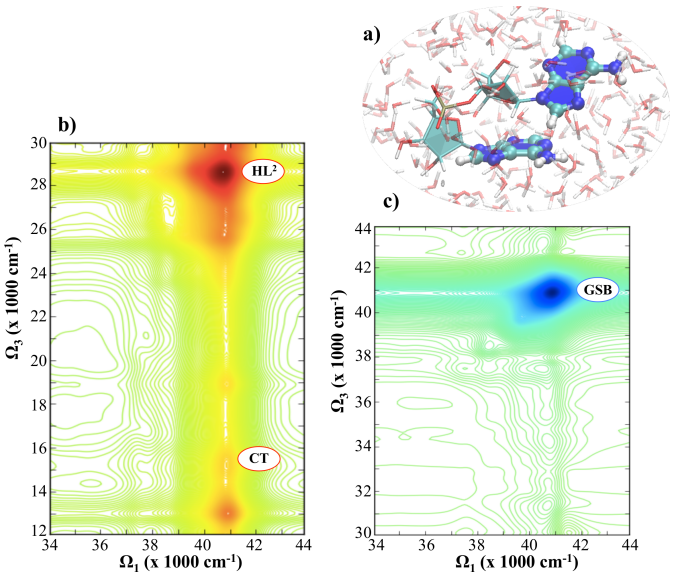


Figure 2. 2DES of water-solvated ApA considering the full conformational space (see Table 1), and showing a) the molecular structure, b) the pump-UV/probe-Vis spectrum, and c) the pump-UV/probe-UV spectrum. Colour scheme for the 2DES spectra are set according to Figure 1.

Figure 2 shows the molecular structure of water-solvated ApA, and both pump-UV/probe-Vis and pump-UV/probe-UV 2DES windows. The key fingerprint in the Vis probing window is the HL^2 transition described previously, and that features as the main fingerprint of each of the adenine monomers, displaying a less localised monomer transition right below HL^2 at ~25k cm^{-1} and that is expected to form part of the same peak. The main difference from the monomer is the large broadening attained by considering several structures, which is partially induced due to the interaction between the ^1L_a states of the different adenine moieties leading to a splitting in their energy estimates along Ω_1, as it will be analysed in detail below. Further considerations would have to be made to properly account for the signal line shape, which is here accounted for through homogeneous broadening due to pure dephasing and is thus symmetric in the single snapshots, asymmetries in the total spectrum arising through the averaging of the peaks over many different conformational arrangements. The other difference easily observed upon comparison with the monomer is the appearance of a signal in the low-energy Vis window (~15k cm^{-1} along Ω_3) and that is attributed mainly to contributions arising from CT states, as contrasted by looking at their specific wave function contributions in our model. This signal is covered in some cases by localised monomer excitations that can also be observed at ~13000 and ~19000 cm^{-1} and obscure the CT signals making their experimental characterisation challenging. These signals are not further discussed since they are associated with localised monomeric fingerprints. Moving onto the pump-UV/probe-UV window (Figure 2c), all that can be seen is the inhomogeneous broadening of the GSB signal along the diagonal and anti-diagonal widths, which arises mainly due to the different nuclear arrangements considered and that weakly displays also the contributions of the ^1L_b state diagonal GSB located at $\Omega_1=\Omega_3\approx$ 38000 cm^{-1}. It is worth noting that the anti-diagonal width of the signal has been reported to hold information regarding the formation of aggregates,[59, 60] species formed upon light absorption due to exciton delocalisation over several monomers that can be monitored through this particular parameter. The diverse water arrangements around the different clusters will also contribute to this broadening, its concrete role being analysed in detail on a monomer for simplicity in an upcoming work and out of scope of the present manuscript.

The analysis of the 2DES spectrum of highly flexible systems like ApA is rather cumbersome as many of the fingerprints are specific to certain conformations. Our computational strategy allows us to dissect the full spectrum into the different clusters or representative nuclear arrangements giving rise to these signals, enabling us to extract the specific interactions behind a given contribution. In order to do so, we have separated the different clusters depending on their C5-C5 distance (see Figure 1 for atom

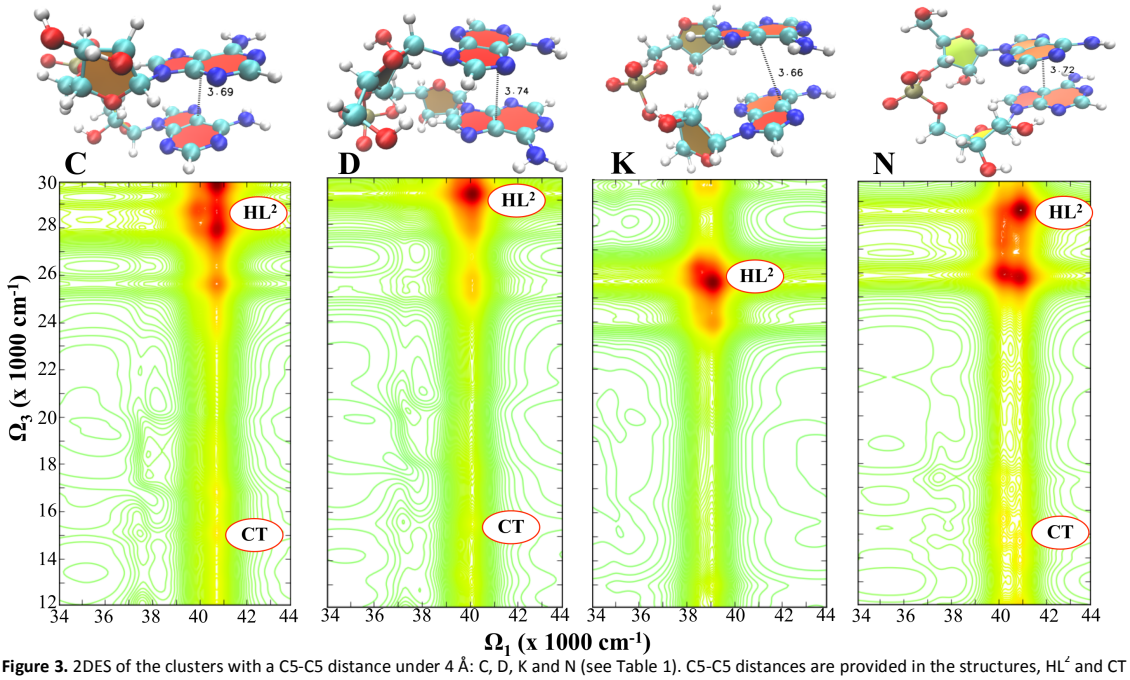


Figure 3. 2DES of the clusters with a C5-C5 distance under 4 Å: C, D, K and N (see Table 1). C5-C5 distances are provided in the structures, HL^2 and CT labels are given to aid in the assignment. Colour scheme for the 2DES spectra are set according to Figure 1.

Please do not adjust margins

Please do not adjust margins

labelling), which is close to the centre of mass of the molecule and thus a decent descriptor to group the different structures. Nevertheless, we are aware that not only the distance but the specific orientation would play a role, and that may also be said from the particular orientation of the adenine moieties with respect to the sugar groups to which they are attached. Thus, we use the simplest descriptor (C5-C5 distance) to partition the clusters into groups, while discussing specific nucleobase arrangements for each group of clusters.

The first group analysed is composed of clusters with C5-C5 distances up to 4 Å, which are close and thus correspond to different π-stacked arrangements with both near face-to-face and various degrees of rotated stacked instances. Due to the lack of fingerprints in the pump-UV/probe-UV window beyond the GSB and off-diagonal bleaching, in the following we will only consider the signal-rich pump-UV/probe-Vis region. Figure 3 displays clusters C, D, K and N (see Table 1), which show different rotated and π-stacked adenine moieties. These type of structures account for ~19% of the conformational space and, importantly, do not show fully (or perfectly) stacked moieties similarly to how the bases appear rotated when embedded in a double helix DNA chain.[61, 62] As can be seen in Figure 3, these bases show strong contributions of the doubly excited HL^2 signal in the 26-30k cm^{-1} range, oscillating along Ω_1 and also along Ω_3. This is more obvious in cluster K, where the ^1L_a traces (consisting of two close-lying traces due to the ^1L_a states of the two adenine moieties) along Ω_1 are slightly split but strongly red-shifted with respect to the other clusters due to the presence of close-lying water molecules pointing to the C4-C5 bond, not featured for the other conformations, which seems to strongly red-shift these excitations. Despite treating the neighbouring water molecules at the MM level, their inclusion at this lower level has been recently shown to reproduce the qualitative features for related systems of the low-lying electronic excitations,[63] such as ^1L_a, and is expected to qualitatively reproduce the solvatochromic shift for the states on the high-energy window like HL^2.

The splitting of the two ^1L_a traces appears to be the main fingerprint of these nuclear arrangements, where the bases are close and strongly interac leading to a 500-1500 cm^{-1} splitting between these two analogous states centred at the different adenine moieties, one being red-shifted and one blue-shifted. Moreover, this fingerprint occurs along Ω_1, which is not accessible with standard pump-probe set-ups and therefore highlights the potential of 2DES for disentangling conformations even on difficult cases such as homo-dimers. Besides the HL^2 signal, other ππ* states are accessed within 2000-3000 cm^{-1} of the latter, which are expected to merge and broaden that signal further. Additional signals appear in the low-energy window, some of them being related to localised ESAs that were obscured in the monomer but appeared in our previous gas-phase reference computations,[28] and some (being highlighted in Figure 3) corresponding to CT states that arise due to the proximity between the adenine moieties. However, these CT states appear to be quite weak and hard to discern from the localised ESAs on the different monomers (~13000 cm^{-1} along Ω_3), the more intense contributions observed in C, D and N (Figure 3) at ~15000 cm^{-1}.

Figure 4 displays the clusters with C5-C5 distances between 4 and 4.5 Å, namely clusters B, G and P (see Table 1). These account for ~17% of the overall conformational space sampled, and display a range of π-stacked and tilted geometries, which resemble those previously shown in Figure 3. Again, the main fingerprints observed are associated to the very intense HL^2 ESA, which is present between 25-28k cm^{-1} along Ω_3, red-shifted as compared to the clusters displaying shorter C5-C5 distances, and that is strongly modulated by the position of the ^1L_a traces along Ω_1, which oscillates as previously discussed. A splitting of the ^1L_a traces can also be observed, particularly in cluster G, but also observed to a minor extent in the others. Additional ππ* ESAs either slightly above or below the HL^2 signal are again displayed, this peak being found (and labelled as signal 11) in reference computations[28] and that refers to a HOMO-2 to LUMO+2 transition accessed through the ^1L_a state. Less intense CT and localised ESAs appear in the low-energy region (~15000 and ~13000 cm^{-1}, respectively) as compared to those observed at shorter C5-C5 distances (see Figure 3).

At longer C5-C5 distances, between 4.5 and 5 Å, four clusters E, F, I and L are featured, which contribute to ~16% to the full conformational space of ApA. As can be seen in Figure 5, upon increase of the C5-C5 distance the traces of the ^1L_a states of the different adenine moieties overlap with one another, giving rise to a narrower signal along Ω_1 analogous to that of the monomer and producing the main intense HL^2 ESA signal at the same energies (~29000 cm^{-1}) as that of the monomer. An analogous set of signals is produced by both adenine moieties, which overlap displaying the other relatively intense ππ* ESAs at slightly different energies and thus giving rise to what appear to be more signals, as can be seen in cluster F. Due to the larger distance, the interaction among ^1L_a traces is not observed and the additional broadening along Ω_1 lost, thus pointing out their non-interactive (unstacked) nature, which also show no intense contributions in the low-energy Vis window (12-15k cm^{-1}) as was observed in the monomer. However, it is worth noting that slightly intense ESAs of CT character may still be reached in conformations displaying these long C5-C5 distances, as can be seen from clusters I and L in Figure 5. The former presents a

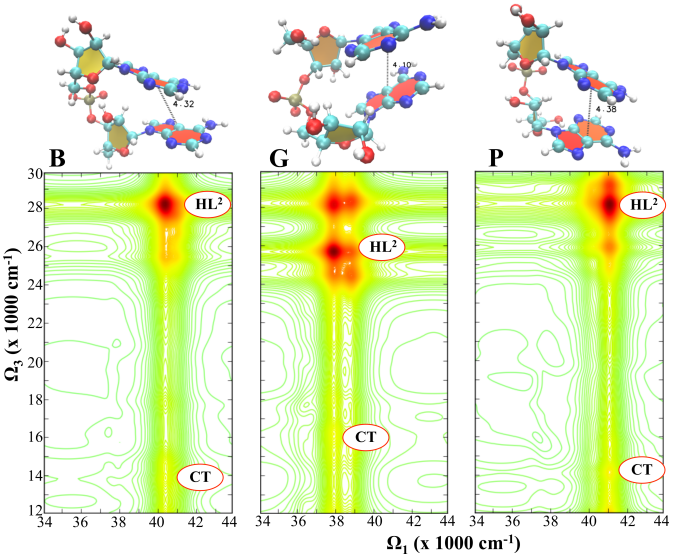


Figure 4. 2DES of the clusters with a C5-C5 distance between 4 and 4.5 Å: B, G, and P (see Table 1). C5-C5 distances are provided in the structures, HL^2 and CT labels are given to aid in the assignment. Colour scheme for the 2DES spectra are set according to Figure 1.

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Please do not adjust margins

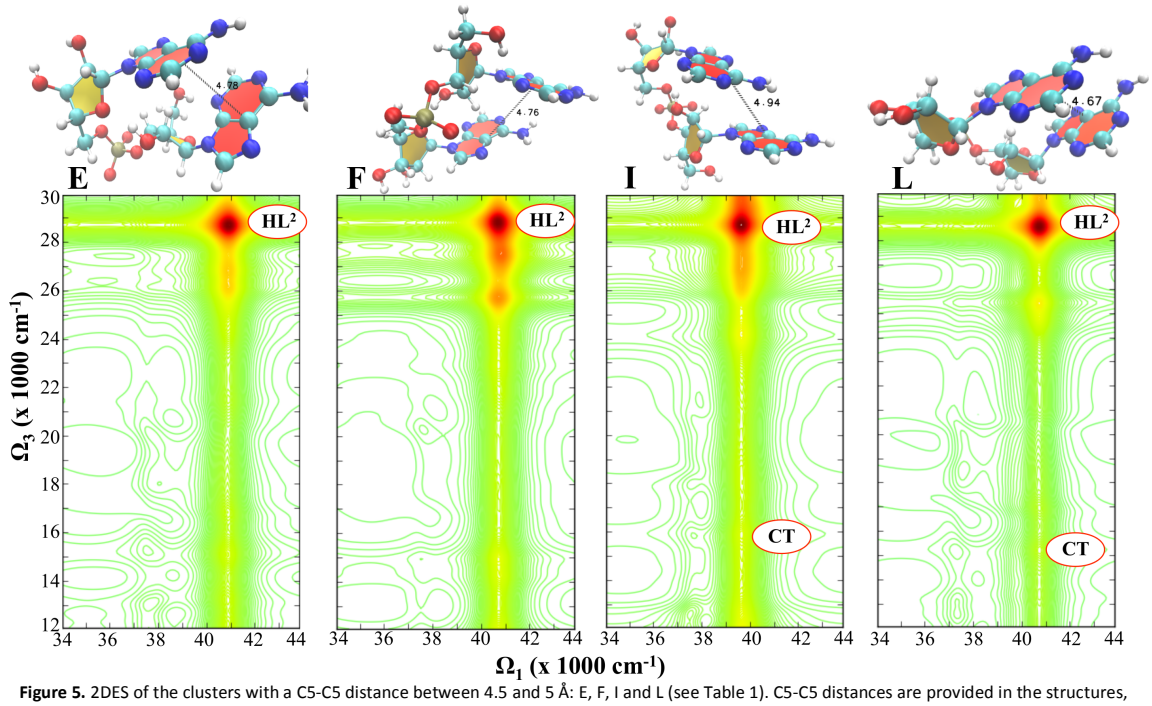


Figure 5. 2DES of the clusters with a C5-C5 distance between 4.5 and 5 Å: E, F, I and L (see Table 1). C5-C5 distances are provided in the structures, HL^2 and CT labels are given to aid in the assignment. Colour scheme for the 2DES spectra are set according to Figure 1.

tilted base pointing towards the other with its amino group, which appears to enhance the appearance of CT states, whereas the latter, despite the long distances displayed, with a fully face-to-face stacked conformation that is expected to enhance the π-stacking interaction facilitating the appearance of these barely bright CT states in the low-energy Vis window as reported above.

Figure 6 shows C5-C5 between 5 and 6 Å, containing clusters H, O and Q that amount to ~9% of the total conformational space, showing different arrangements of distant adenine moieties that do not interact strongly with one another. As can be seen, and upon comparison of the monomer spectrum given in Figure 1, the fingerprints displayed essentially relate to those register from their monomeric contributions, with the main HL^2 ESA in the 26-28k cm^{-1} range, except for cluster O that experiences a strong red-shift along both Ω_1 and Ω_3 due to the presence of a close-lying water molecule coordinating directly to the C4 position. No interaction is observed for clusters H and Q, as indicated by the absence of the ^1L_a splitting, featured by the closer and more interacting clusters, despite slightly displaced HL^2 signals along Ω_3 can be observed. Cluster O, on the other hand, displays the shortest distance within this cluster group (5.09 Å, see Figure 6) and shows a slight splitting in the trace along Ω_1, while showing no strong contributions in the low-energy Vis window. Despite displaying larger C5-C5 distances, clusters H and O show CT ESA signals in the low energy (~15k cm^{-1}) Vis probe window, which appear to be mediated mainly through the interaction of one of the amino groups with an aromatic group of the opposite adenine. While displaying large C5-C5 distances, adenine moieties in H and O still remain at much closer amino-to-base distances, which are the ones facilitating the interaction and thus the reason behind the weak CT signals in the low-energy window. Therefore, the proximity between the amino groups to the aromatic moieties on the opposing base, also noted previously for cluster I in Figure 5, appear to be an important aspect to consider behind the appearance of CT states among adenine dimers like ApA.

The last group of clusters consists on the ones with the longest distances, beyond 6 Å, and enclose clusters A, J and M (see Table 1), which amount to 16% of the sampled conformational space, and contain the most repeated conformation (A) that accounts on its own for 10% of the overall population. Figure 7 shows the different spectra, showcasing three apparently different behaviours: the most populated cluster A features joint ^1L_a traces along Ω_1, not displaying the previously reported splitting due to strongly interacting π-stacked systems, while showing a strong localised ESA at ~13000 and a CT signal at ~15-16k cm^{-1}, the more intense HL^2 being placed in this case beyond the Vis probe window. The relative intensity of the CT state is attributed to the proximity of one of the

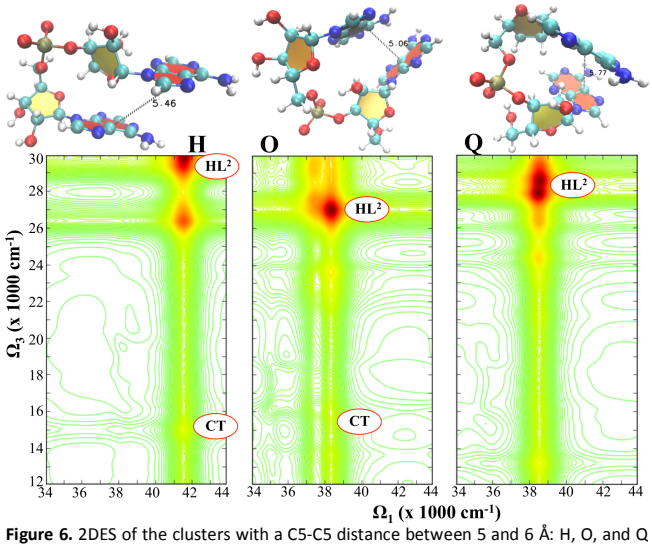


Figure 6. 2DES of the clusters with a C5-C5 distance between 5 and 6 Å: H, O, and Q (see Table 1). C5-C5 distances are provided in the structures, HL^2 and CT labels are given to aid in the assignment. Colour scheme for the 2DES spectra are set according to Figure 1.

Please do not adjust margins

Please do not adjust margins

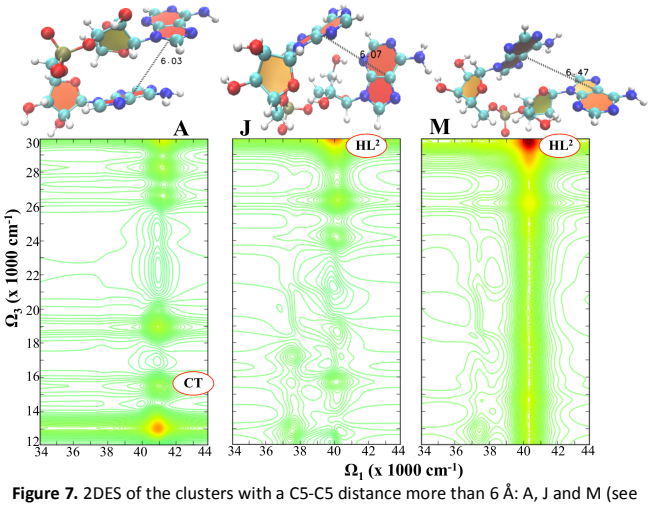


Figure 7. 2DES of the clusters with a C5-C5 distance more than 6 Å: A, J and M (see Table 1). C5-C5 distances are provided in the structures, HL^2 and CT labels are given to aid in the assignment. Colour scheme for the 2DES spectra are set according to Figure 1.

adenine's amino group to the 5-ring of the opposing moiety. This interaction, which appears to stabilise CT transitions, is behind most T-stacked interactions, particularly those mediated by the interaction of the amino group with the opposing nucleobase, and thus expected to be a fingerprint for these particular conformations, coupled with the lack of broadenings along Ω_1 characteristic of these distant nuclear arrangements. Cluster J shows a qualitatively different spectrum due to the partial appearance of the HL^2 state in the high-energy window, which obscures the rest of the localised signals appearing in the lower energy window and displaying no CT contributions. Cluster M, on the other hand, displays monomer-like ESAs, displaying a weak signal at ~14000 cm^{-1} attributed to a localised ESA. Overall, more distant bases lack the additional broadening along Ω_1, originated due to the splitting of the two ^1L_a states of the different strongly interacting adenine moieties in ApA, while showing a degree of charge transfer character in T-stacked-like conformations. This is particularly the case on those structures where the amino group points towards the opposing 5-ring moiety, being fingerprints for these particular conformations coupled with narrow bands along Ω_1. It is worth noting that both π- and T-stacked CT signals appear in the same energy window, the different expected broadenings along Ω_1 being again the only way to disentangle them.

Unfortunately, no 2DES experimental data is currently available for ApA to confront our present findings, the first time-resolved 2D spectrum for water-solvated adenine in the UV- pump and probe windows being recently reported in the literature,[17] together with a theoretical model rationalising the different signals displayed and their monitoring.[64] This high-energy window employed has been shown to be capable of registering the excited state dynamics of adenine also in 1D pump-probe spectroscopy by means of monitoring the GSB recovery.[50] While capable of registering the dynamics of one single nucleobase, this window is expected to not be as efficient for monitoring dimeric[7] and multimeric[65] species featuring a range of different conformations with very distinct associated decay lifetimes. In this sense, the present work proposes the UV-pump/Vis-probe range as a more suited spectral window to monitor the different conformations attained in multimeric species, displaying distinct fingerprints that may be associated to particular nuclear arrangements, albeit their experimental characterisation may prove more challenging. Upcoming experimental efforts, coupled with carefully built theoretical models, will be required in the near future to ascertain the role played by different conformations in the photophysics and photochemistry of dimeric and multimeric species like ApA.

Conclusions

A computational protocol is introduced that allows simulating 2DES in highly flexible solvated monophosphate dinucleosides and disentangling the spectroscopic fingerprints of various stacked/unstacked configurations of the ApA dimer case. 2DES spectra are expected to reveal information about molecular arrangements in dimeric and multimeric systems, going beyond the limitations of standard 1D pump-probe techniques for the study of photo-initiated processes. By employing ground state molecular dynamics and clustering techniques, 17 structures of the water-solvated ApA dimer have been selected in order to represent ~75% of the sampled conformational space. Accurate QM/MM RASPT2 computations have been performed on top of each of the selected structures to obtain the electronic energy levels and transition dipole moments to simulate the 2DES signals within the SOS//QM/MM approach. By weighting in the contributions of each cluster to the signal according to their relative population, the overall 2DES spectrum is obtained, showing fingerprints related to both unstacked (non-interacting) and either π- or T-stacked (interacting) ApA conformations. The extracted clusters can be analysed one-by-one, enhancing our understanding of the nuclear arrangements behind specific spectral signals and allowing to characterise particular fingerprints that arise for each different possible intermolecular interaction. We show how for π-stacked arrangements, a splitting of the two ^1L_a traces of both adenine moieties leads to a marked broadening of the ^1L_a trace along Ω_1, which can potentially mark the appearance of strongly interacting stacked species and being a specific fingerprint of 2DES that cannot be observed in standard pump-probe set-ups. T-stacked conformations, on the other hand, are shown to display relatively intense ESA CT signals in the low-energy Vis window through the interaction between the amino group of one adenine monomer and the opposing base, and being particularly intense for those cases promoting amino to 5-ring interactions. Unstacked ApA conformations are shown to display very similar spectra to those of the monomer, as expected from these non-interacting species. Overall, 2DES is shown to be a promising alternative to standard pump-probe experiments for elucidating and studying the concrete photo-initiated events in DNA multimeric systems, being potentially able to separate contributions arising from a wide range of conformations and thus providing unique insights into how photostability and photo-damage is modulated due to the presence of extended and interacting nucleobase sequences in our genomic material.

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins

Please do not adjust margins

Acknowledgements

M.G. acknowledges support by the European Research Council Advanced Grant STRATUS (ERC-2011-AdG No. 291198). I.R. gratefully acknowledges support by the Agence National de la Recherche (FEMTO-2DNA, ANR-15-CE29-0010) and the ENS-Lyon (grants 900/S81/BS81-FR14 and MI-LOURD-FR15). J.S.-M. and I.R. acknowledge the use of HPC resources of the "Pôle Scientifique de Modélisation Numérique" (PSMN) at the ENS-Lyon, France. S.M. gratefully acknowledges the support of the National Science Foundation (grant CHE-1663822), and the Chemical Sciences, Geosciences, and Biosciences division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy, award DE-FG02-04ER15571.

Notes and references

1. G. D. Scholes, G. R. Fleming, L. X. Chen, A. Aspuru-Guzik, A. Buchleitner, D. F. Coker, G. S. Engel, R. van Grondelle, A. Ishizaki, D. M. Jonas, J. S. Lundeen, J. K. McCusker, S. Mukamel, J. P. Ogilvie, A. Olaya-Castro, M. A. Ratner, F. C. Spano, K. B. Whaley and X. Zhu, *Nature*, 2017, **543**, 647-656.
2. C. E. Crespo-Hernandez, B. Cohen, P. M. Hare and B. Kohler, *Chemical Reviews*, 2004, **104**, 1977-2019.
3. C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernandez and B. Kohler, in *Annual Review of Physical Chemistry*, Annual Reviews, Palo Alto, 2009, vol. 60, pp. 217-239.
4. D. Markovitsi, T. Gustavsson and I. Vayá, *The Journal of Physical Chemistry Letters*, 2010, **1**, 3271-3276.
5. W. J. Schreier, P. Gilch and W. Zinth, *Annual Review of Physical Chemistry*, 2015, **66**, 497-519.
6. W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernández, V. N. Swaminathan, T. Carell, W. Zinth and B. Kohler, *Science*, 2007, **315**, 625-629.
7. T. Takaya, C. Su, K. d. L. Harpe, C. E. Crespo-Hernández and B. Kohler, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 10285-10290.
8. A. Giussani, J. Segarra-Martí, D. Roca-Sanjuán and M. Merchán, in *Photoinduced Phenomena in Nucleic Acids I*, eds. M. Barbatti, A. C. Borin and S. Ullrich, Springer International Publishing, 2015, vol. 355, ch. 501, pp. 57-97.
9. M. Barbatti, A. J. A. Aquino, J. J. Szymczak, D. Nachtigallova, P. Hobza and H. Lischka, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 21453-21458.
10. L. Serrano-Andrés and M. Merchán, *J. Photochem. Photobiol. C-Photochem. Rev.*, 2009, **10**, 21-32.
11. D. B. Bucher, B. M. Pilles, T. Carell and W. Zinth, *Proceedings of the National Academy of Sciences*, 2014, **111**, 4369-4374.
12. D. B. Bucher, C. L. Kufner, A. Schlueter, T. Carell and W. Zinth, *Journal of the American Chemical Society*, 2016, **138**, 186-190.
13. L. E. Buchanan, J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. de Pablo, J. L. Skinner and M. T. Zanni, *Proceedings of the National Academy of Sciences*, 2014, **111**, 5796-5801.
14. A. Ghosh, J.-J. Ho, A. L. Serrano, D. R. Skoff, T. Zhang and M. T. Zanni, *Faraday Discussions*, 2015, DOI: 10.1039/C4FD00173G.
15. B. A. West, J. M. Womick and A. M. Moran, 2011, 8630-8637.
16. B. A. West and A. M. Moran, 2012.
17. V. I. Prokhorenko, A. Picchiotti, M. Pola, A. G. Dijkstra and R. J. D. Miller, *The Journal of Physical Chemistry Letters*, 2016, **7**, 4445-4450.
18. I. Rivalta, A. Nenov, G. Cerullo, S. Mukamel and M. Garavelli, *International Journal of Quantum Chemistry*, 2014, **114**, 85-93.
19. S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*, O.U.P., New York, 1995.
20. S. Mukamel, *Annual Review of Physical Chemistry*, 2000, **51**, 691-729.
21. D. Polli, I. Rivalta, A. Nenov, O. Weingart, M. Garavelli and G. Cerullo, *Photochemical & Photobiological Sciences*, 2015, DOI: 10.1039/C4PP00370E.
22. Q. Li, A. Giussani, J. Segarra-Martí, A. Nenov, I. Rivalta, A. A. Voityuk, S. Mukamel, D. Roca-Sanjuán, M. Garavelli and L. Blancafort, *Chemistry – A European Journal*, 2016, **22**, 7497-7507.
23. A. Nenov, S. a Beccara, I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli, *ChemPhysChem*, 2014, **15**, 3282-3290.
24. D. Case, T. A. Darden, T. E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. Kollman, *Amber 11*, University of California, San Francisco, 2011.
25. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *Journal of Computational Chemistry*, 2005, **26**, 1668-1688.
26. A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, III, C. A. Laughton and M. Orozco, *Biophysical Journal*, **92**, 3817-3829.
27. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
28. A. Nenov, A. Giussani, J. Segarra-Martí, V. K. Jaiswal, I. Rivalta, G. Cerullo, S. Mukamel and M. Garavelli, *The Journal of Chemical Physics*, 2015, **142**, 212443.
29. 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, *Journal of Computational Chemistry*, 2016, **37**, 506-541.
30. P. A. Malmqvist, A. Rendell and B. O. Roos, *The Journal of Physical Chemistry*, 1990, **94**, 5477-5482.

Please do not adjust margins

Please do not adjust margins

31. P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer and L. Gagliardi, *The Journal of Chemical Physics*, 2008, **128**, 204109.
32. V. Saurí, L. Serrano-Andrés, A. R. M. Shahi, L. Gagliardi, S. Vancoillie and K. Pierloot, *Journal of Chemical Theory and Computation*, 2011, **7**, 153-168.
33. F. Aquilante, R. Lindh and T. Bondo Pedersen, *The Journal of Chemical Physics*, 2007, **127**, 114107.
34. F. Aquilante, T. B. Pedersen and R. Lindh, *The Journal of Chemical Physics*, 2007, **126**, 194106.
35. T. Pedersen, F. Aquilante and R. Lindh, *Theoretical Chemistry Accounts*, 2009, **124**, 1-10.
36. P. O. Widmark, B. Joakim, Persson and B. O. Roos, *Theor. Chim. Acta*, 1991, **79**, 419-432.
37. S. F. Altavilla, J. Segarra-Martí, A. Nenov, I. Conti, I. Rivalta and M. Garavelli, *Frontiers in Chemistry*, 2015, **3**.
38. N. Forsberg and P. A. Malmqvist, *Chemical Physics Letters*, 1997, **274**, 196-204.
39. J. P. Zobel, J. J. Nogueira and L. Gonzalez, *Chemical Science*, 2017, DOI: 10.1039/C6SC03759C.
40. G. Ghigo, B. O. Roos and P.-Å. Malmqvist, *Chemical Physics Letters*, 2004, **396**, 142-149.
41. P.-Å. Malmqvist and B. O. Roos, *Chemical Physics Letters*, 1989, **155**, 189-194.
42. P. Å. Malmqvist, B. O. Roos and B. Schimmelpfennig, *Chemical Physics Letters*, 2002, **357**, 230-240.
43. S. Mukamel, R. Oszwaldowski and D. Abramavicius, *Phys. Rev. B*, 2007, **75**, 245305.
44. D. Abramavicius, B. Palmieri, D. V. Voronine, F. Šanda and S. Mukamel, *Chemical Reviews*, 2009, **109**, 2350-2408.
45. M. Richter and B. P. Fingerhut, *Journal of Chemical Theory and Computation*, 2016, **12**, 3284-3294.
46. A. Nenov, A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel and M. Garavelli, *Physical Chemistry Chemical Physics*, 2015, **17**, 30925-30936.
47. R. Tempelaar, C. P. van der Vegte, J. Knoester and T. L. C. Jansen, *The Journal of Chemical Physics*, 2013, **138**, 164106.
48. C. P. van der Vegte, A. G. Dijkstra, J. Knoester and T. L. C. Jansen, *The Journal of Physical Chemistry A*, 2013, **117**, 5970-5980.
49. M. Kowalewski, B. P. Fingerhut, K. E. Dorfman, K. Bennett and S. Mukamel, *Chemical Reviews*, 2017, DOI: 10.1021/acs.chemrev.7b00081.
50. B. Cohen, P. M. Hare and B. Kohler, *Journal of the American Chemical Society*, 2003, **125**, 13594-13601.
51. T. Gustavsson, A. Sharonov, D. Onidas and D. Markovitsi, *Chemical Physics Letters*, 2002, **356**, 49-54.
52. F. Santoro, R. Improta, T. Fahleson, J. Kauczor, P. Norman and S. Coriani, *The Journal of Physical Chemistry Letters*, 2014, **5**, 1806-1811.
53. P. G. Szalay, T. Watson, A. Perera, V. F. Lotrich and R. J. Bartlett, *Journal of Physical Chemistry A*, 2012, **116**, 6702-6710.
54. R. Improta and V. Barone, Springer Berlin Heidelberg, 2014, DOI: 10.1007/128_2013_524, ch. 524, pp. 1-29.
55. A. J. Pepino, J. Segarra-Martí, A. Nenov, R. Improta and M. Garavelli, *The Journal of Physical Chemistry Letters*, 2017, **8**, 1777-1783.
56. L. Martinez-Fernandez, A. J. Pepino, J. Segarra-Martí, J. Jovaisaite, I. Vayá, A. Nenov, D. Markovitsi, T. Gustavsson, A. Banyasz, M. Garavelli and R. Improta, *Journal of the American Chemical Society*, 2017, **139**, 7780-7791.
57. A. Giussani, J. Segarra-Martí, A. Nenov, I. Rivalta, A. Tolomelli, S. Mukamel and M. Garavelli, *Theoretical Chemistry Accounts*, 2016, **135**, 1-18.
58. S. Jafilan, L. Klein, C. Hyun and J. Florián, *The Journal of Physical Chemistry B*, 2012, **116**, 3613-3618.
59. I. Stiopkin, T. Brixner, M. Yang and G. R. Fleming, *The Journal of Physical Chemistry B*, 2006, **110**, 20032-20037.
60. N. S. Ginsberg, Y.-C. Cheng and G. R. Fleming, *Accounts of Chemical Research*, 2009, **42**, 1352-1363.
61. I. Conti, A. Nenov, S. Ho, S. F. Altavilla, I. Rivalta, E. Dumont, G. Orlandi and M. Garavelli, *Physical Chemistry Chemical Physics*, 2015, **17**, 7291-7302.
62. A. Nenov, J. Segarra-Martí, A. Giussani, I. Conti, I. Rivalta, E. Dumont, V. K. Jaiswal, S. F. Altavilla, S. Mukamel and M. Garavelli, *Faraday Discussions*, 2015, **177**, 345-362.
63. L. Martínez-Fernández, A. J. Pepino, J. Segarra-Martí, A. Banyasz, M. Garavelli and R. Improta, *Journal of Chemical Theory and Computation*, 2016, **12**, 4430-4439.
64. A. G. Dijkstra and V. I. Prokhorenko, *The Journal of Chemical Physics*, 2017, **147**, 064102.
65. C. Su, C. T. Middleton and B. Kohler, *Journal of Physical Chemistry B*, 2012, **116**, 10266-10274.

1

This journal is © The Royal Society of Chemistry 20xx

Please do not adjust margins